Faculty of Mathematics and Physics
Charles University

Radiative effects in gauge extensions of the Standard Model of particle interactions

Michal Malinský

Habilitation thesis

2020

The effort to understand the universe is one of the very few things that lifts human life a little above the level of farce, and gives it some of the grace of tragedy.

Steven Weinberg
The first three minutes

Dedication

I dedicate this thesis to the people that make my life meaningful – to my wife whose patience I often stretch to limits with no effect on her unflattering support, to my children to whom I owe a lot, to my parents who gave me all they could, and to all friends whose enthusiasm keeps me going.

Professionally, I would like to thank to three men that, over the years, formed my way of thinking about physics – Jiří Hořejší for the attitude, Stefano Bertolini for the method and Goran Senjanović for the passion.

In Prague, February 2020

Contents

Chapter 1

Introduction

The quest for a simple and accurate description of the elementary constituents of matter and their interactions is arguably one of the biggest intellectual adventures of all times. Based on the revolutionary insights of Maxwell, Einstein, Planck and many other giants of modern physics, our understanding of electromagnetic, strong and weak nuclear forces governing the quantum world matured over the last century into what is nowadays claimed to be the most accurate physics theory ever, the Standard Model (SM) of particles and their interactions. Fifty years of an enormous experimental scrutiny, crowned recently by the discovery of its last missing ingredient – the Higgs boson – makes us believe that the SM, indeed, provides a perfectly valid description of the microcosmos at energies stretching up to, at least, the TeV scale.

Yet, there are now clear signals that the Standard Model in its original Glashow-Salam-Weinberg (GSW) formulation [1–3] does not encompass all observable particle physics phenomena – the discovery of neutrino flavour oscillations [4–6] strongly indicates that neutrinos, the most elusive of all SM particles, are not massless as assumed in the GSW construction. Moreover, their unprecedented lightness, as implied either directly by the beta-decay data (see, e.g., [7]) or indirectly by various cosmological limits [8], suggests that masses of neutrinos may have a rather different dynamical origin than those of all other SM matter fields.

Indeed, the most plausible explanation of their sub-eV mass scale has to do with the interesting option that neutrinos – as the only electrically neutral matter fermions – may be in fact Majorana spinors whose mass could be associated (in the SM picture) to the unique $d = 5$ lepton-number-violating operator $LLHH/\Lambda$ [9]; in that case, the smallness of the neutrino masses would be attributed to the largeness of the so-called seesaw scale Λ typically assumed to be in the 10^{12-14} GeV ballpark. By definition, such a Majorana mass term violates all additive quantum numbers the neutrinos carry; in particular, it breaks the global rephasing invariance of the SM Lagrangian associated to the lepton number (L). Although, for a long time, L has been considered a perfect empirical symmetry of the low-energy world, there is actually nothing sacred about the SM lepton number conservation – it is indeed violated by non-perturbative effects associated to the so-called triangle anomalies [10, 11] of the corresponding quantum currents. On the other hand, at low temperature these effects are so small that there is essentially no way to look for them in laboratory-based searches; nevertheless, they might have played a crucial role in the hot early Universe [12]. In this respect, the seesaw realisation of neutrino masses can be seen as a first indication of a perturbative lepton number violation (LNV) in beyond-Standard-Model (BSM) physics; as such, it is subject to enormous experimental efforts stretching form the neutrinoless double-beta decay activities at the intensity frontier (see, e.g., [13, 14]) to the LNV collider searches (cf. [15]) on the high-energy side.

Another piece of observation the SM fails to account for is the value of the baryon-to-photon number density ratio in the early Universe, a crucial parameter governing the abundances of the light nuclei created during Weinberg's "first three minutes" of its thermal history. Indeed, the measurements of the intensity of the Deuterium absorption lines in the light of high-redshift quasars yield values which are some 8 orders of magnitude above the 10^{-18} SM estimate based on the assumption of an exact baryon-number (B) symmetry of the initial conditions.

Since the SM baryon-number current features the same anomaly structure as the leptonic one and, as such, it may be viewed as just the other facet of essentially the same coin, it is very natural to ask whether there are any indications that it may also be perturbatively violated in the BSM physics. In the effective SM picture this is analogous to asking at which scale the $d = 6$ baryon-number-violating (BNV) operators [9, 16, 17] mediating "classically" forbidden processes like, e.g., proton decay, BNV neutron decays etc., are generated and whether there is any chance to get a firm grip on them.

From that perspective, the profound idea of grand unification of strong and electroweak interactions [18] as a possible theory of perturbative baryon number violation comes about as a perfectly natural and logical continuation of this line of thoughts[1]. Indeed, if the physics at some high energy scale was governed by a gauge theory based on a simple gauge group G [such as, e.g., $SU(5)$ or $SO(10)$] containing the SM gauge symmetry $G_{\rm SM} = [SU(3) \otimes SU(2) \otimes U(1)]/Z_6$ as a subgroup, the quarks and leptons would occupy common irreducible representations of G; consequently, the gauge bosons associated to the coset $G/G_{\rm SM}$ (and the dynamical remnants of the Higgs multiplets triggering the relevant spontaneous symmetry breakdown) will mediate B and L vio-

[1]Needless to say, this is not the original reasoning of Georgi and Glashow who were rather motivated by the uniqueness of the SM gauge group embedding into the $SU(5)$, c.f. [18].

lating quark-lepton transitions. The $d = 6$ (and higher-order) BNV operators are then obtained by integrating these heavy degrees of freedom out of the low-energy effective theory and, in any specific scheme, their structure will closely reflect the details of the underlying dynamics. Besides that, such a unified-gauge-theory realisation of the $d = 6$ SM BNV operators has one great virtue: the relevant energy scale M (defined essentially as the mass of the mediators of the BNV transitions) is fixed by the requirement that the running gauge couplings in the effective theory converge to a common value close to M. Remarkably enough, this is exactly what happens in the SM at around 10^{15-16} GeV (focusing on the non-abelian sector and assuming, for simplicity, no new physics between M and the electroweak scale). Hence, unlike for the seesaw scale Λ governing the $d = 5$ neutrino mass operator (which can be anywhere below about 10^{15} GeV as long as the coefficient of the Weinberg operator is suppressed accordingly), in grand unified theories (GUTs) the scale of perturbative $d = 6$ BNV is subject to a strong constraint.

Although, obviously, the extreme remoteness of M makes any direct collider scrutiny of the GUT paradigm inconceivable, there are several reasons to believe that the basic gauge unification idea can be to a high degree testable even with the current technology. Besides proton decay as a hallmark of BNV, the spontaneous breakdown of a unified gauge-group G to the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ of the Standard Model can give rise to massive topological defects (monopoles, cosmic strings, domain walls) corresponding to stable extended gauge/Higgs field configurations with very peculiar features. These, among other things, include the so-called Callan-Rubakov effect [19,20] corresponding to, e.g., a strong enhancement of the BNV rates in the core of such objects. Although the relevant monopole fluxes are nowadays likely to be exponentially suppressed by the early cosmological inflation, the discovery of monopoles would provide such a robust evidence for grand unification that they are still subject to dedicated experimental efforts (cf. [21]) typically along with the "classical" proton decay searches. In this respect, Nature seems to be (as often) rather generous because the raw estimate of M in the 10^{15-16} GeV ballpark yields proton lifetime in the range of about 10^{30-36} years, right at the verge of feasibility; let us note that these numbers correspond roughly to one decay in about a ton (10^{30} years) to a megaton (10^{36} years) of material in about a year.

This, obviously, represents an enormous experimental challenge and, thus, the progress has been relatively slow (amounting to roughly one order of magnitude improvement per decade). To name just few of the most important early searches, let us mention the NUSEX [22], FREJUS [23], SOUDAN [24] and Kamiokande [25] experiments from 1980's and the 3rd phase of IMB [26] at the beginning of the 1990's which probed the proton lifetime up to about 10^{33} years. The current best limits from Super-Kamiokande (SK) [27] reach up to 2×10^{34} years in the $p^+ \to e^+\pi^0$ channel and to about 10^{33} years for $p^+ \to K^+\overline{\nu}$. On the other hand, these numbers suggest that we may be just on the brink of really observing the first experimental signal of this kind. In any case, the close complementarity between the methodology of the proton decay searches and that of the very lively and extensive neutrino-physics programme (which is central to most of the upcoming large-volume facilities like DUNE or Hyper-K) ensures a bright future for the experimental searches for the baryon number violation for at least the next three decades. Hence, we should cross our fingers and stay tuned.

In this thesis we shall review the crucial points along the lines of reasoning sketched above and provide a brief account of the candidate's contribution to the evolution of this thrilling field of the high-energy physics research. In doing so we shall necessarily be rather selective as to which of its many aspects shall be entertained and which – with all due respect to their pioneers – shall be suppressed (or even entirely omitted). Let us also note that the selected publications enclosed in Chapter 5 represent just a small fragment of candidate's achievements; their complete list is available as a part of his professional CV enclosed in the file.

Chapter 2

Perturbative baryon and lepton number violation in gauge extensions of the Standard Model

2.1 Neutrino masses in simple extensions of the Standard Model

The overwhelming evidence that at least two out of the three known neutrino states are massive clearly calls for a generalisation of the original GSW formulation of the Standard Model (SM) where they were, by construction, two-components Weyl fermions.

At first glance, this exercise has a trivial solution: postulating the existence of new spin-$\frac{1}{2}$ quantum fields N_R transforming as full singlets under the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ SM gauge symmetry does not seem to be a big deal, the more that the presence of such fields (if provided in three copies) nicely restores the symmetry among the numbers of the left- and right-handed (LH, RH) fermions of the theory.

2.1.1 RH singlet fermions and Dirac neutrino masses

Indeed, with the extra RH leptons at hand one can speculate about a leptonic analogue of the up-type Yukawa coupling structure $\overline{Q}_L Y_u u_R \tilde{H}$ in the form

$$\mathcal{L} \ni \overline{L}_L Y_\nu N_R \tilde{H} + h.c. \,, \tag{2.1}$$

producing, in the broken phase with $\langle \tilde{H} \rangle = \frac{1}{\sqrt{2}}(v, 0)^T$, a mass term of the same (Dirac) type as those encountered in the SM for the charged fermions. Let us note that, assuming the standard hypercharge assignment of all the SM charged matter fields[1]

$$Y_{Q_L} = +\tfrac{1}{6}\,,\; Y_{u_R} = +\tfrac{2}{3}\,,\; Y_{d_R} = -\tfrac{1}{3}\,,\; Y_{L_L} = -\tfrac{1}{2}\,,\; Y_{e_R} = -1\,, \tag{2.2}$$

N_R is enforced to be a complete SM gauge singlet and, as such, it would not feel any of the SM gauge interactions. In this respect, adding such a new field into the play does not upset the basic SM phenomenology in any way. Second, requiring $m_\nu = Y_\nu v/\sqrt{2}$ to be in the sub-eV ballpark (as indicated by beta decay experiments [7] and, independently, also by cosmology [8]) the Yukawa matrix Y_ν is sentenced to be very small $|Y_\nu| \lesssim 10^{-11}$, almost at the verge of negligibility. Note, however, that in principle there is nothing wrong about the relative smallness of Y_ν as compared to the other Yukawa matrices in the SM; indeed, Yukawa couplings are self-renormalized and, hence, a small one remains small even at the loop level.

Let us turn to the phenomenology though. As simple as the SM extension with (2.1) looks, the physical consequences of the presence of even a single extra term along with the traditional SM interactions are paramount:

1. *Lepton flavour violation.* With a "complete" set of RH leptons in the game the flavour structure of the leptonic charged-current (CC) interaction Lagrangian fully resembles the situation in the quark sector. In the mass basis, a 3×3 unitary matrix (usually called Pontecorvo-Maki-Nakagawa-Sakata, PMNS) encompasses all the flavour-changing charged-current (CC) interactions involving leptons:

$$\mathcal{L}^{\rm CC}_{\rm lept.} = \frac{g}{\sqrt{2}}\overline{\ell^\alpha_L}\gamma^\mu U^{\rm PMNS}_{\alpha i}\nu^i_L W^-_\mu + h.c. \tag{2.3}$$

   Besides neutrino oscillations, loop-induced flavour-changing neutral current processes (FCNC) such as $\mu\to e\gamma$, $\tau\to 3e$ will be allowed, albeit with immesurably small rates, for instance[2] [29,30]

$$\Gamma(\mu\to e\gamma) \sim \frac{G_F^2 m_\mu^5}{192\pi^3}\frac{3\alpha}{128\pi}\sin^2 2\theta\left(\frac{m_2^2-m_1^2}{M_W^2}\right)^2\,, \tag{2.4}$$

   where a convenient two-flavour approximation (reflected by the presence of a single leptonic mixing angle and only two masses) has been employed.

2. *Leptonic CP violation.* Another phenomenon intimately connected to the non-trivial mixing in the lepton sector CC's (2.3) is the presence of a new source of

[1] At first glance this is the most natural thing to do; however, this assumption may not be fully justified in extensions of the Standard Model featuring new chiral fermions, cf. Sect. 2.1.2.

[2] Numerically, formula (2.4) yields BR($\mu\to e\gamma$) in the 10^{-40} ballpark. Note that there is no chance to see a signal like that as the sensitivity of the current experiments (such as MEG [28]) does not exceed about 4×10^{-13}.

CP violation related to the (thus far unknown) complex structure of the PMNS matrix. Besides its potentially striking effects in the neutrino oscillation phenomena it would also contribute to other leptonic CP-odd observables such as the electric dipole moment (EDM) of the electron d_e. To this end, let us note that $d_e \sim 10^{-38}$e.cm obtained in the SM [31] (to be compared with the experimental limit [32] $d_e^{\rm exp} \lesssim 10^{-29}$e.cm) receives its first contribution only at the four-loop level due to the need to transfer the CP phase from the quark sector Cabibbo-Kobayashi-Maskawa (CMK) matrix into the leptons; with the extra CP violation in the PMNS matrix the same observable can be generated already at two loops [33] (though hardly enhanced; the purely leptonic contribution to d_e would not exceed $d_e \lesssim 10^{-43}$e.cm either).

2.1.2 Dirac neutrinos and the SM hypercharge de-quantization

So far, the introduction of the RH singlet(s) into the SM field list and the addition of the relevant Yukawa term (2.1) to the SM Lagrangian was merely beneficial. Unfortunately, at the quantum level, this simple scheme suffers from a serious pathology when it comes to one of the most beautiful features of the original SM, namely, the mechanism of the (hyper)charge quantization by the requirement of a complete cancellation of chiral anomalies.

To this end, let us first recall that unlike for the non-Abelian charges (i.e., the eigenvalues of the T_c^3, T_c^8 and T_L^3 Cartans of the $SU(3)_c \otimes SU(2)_L$ part of the SM gauge group) whose quantization is ensured by the properties of the special unitary groups' representations, Abelian charges like, e.g., the eigenvalues of the hypercharge operator Y (the generator of $U(1)_Y$ of the SM) can be, from the group theory perspective, arbitrary real numbers. From this point of view, the hypercharges of the SM matter fields in Eq. (2.2) may be (at the classical level) viewed as rather phenomenological quantities which, however, happen to be carefully selected to yield the desired pattern of electric charges following the Gell-Mann-Nishijima relation

$$Q = T_L^3 + Y \,. \tag{2.5}$$

The precision with which this game must be played is rather unprecendented; for instance, the current neutron neutrality constraints [34] are such that the hypercharges of the up and down quarks must conspire to the level of at least 21 significant digits!

Cancellation of chiral anomalies in the SM

From this perspective, it is great that the SM provides a clear rationale for this "fine-tuning" (and, in turn, for the electric neutrality of atomic matter) at the quantum level. The argument has to do with the delicacy of the quantum structure of chiral gauge theories that, in general, are prone to a certain pathological behaviour which consists in the possible dependency of some of the physical amplitudes on the selection of the gauge in the corresponding calculations. As such, these so called *chiral anomalies* [10,11] must be absent if a theory under consideration is to play the role of an internally consistent and calculable model of Nature.

The requirement of absence of such anomalies may thus be used as a powerful discriminator among different settings and, in particular, among different sets of charges chiral fermions can attain in specific scenarios. The point is that the core anomaly structure is purely algebraic,

$$\mathcal{A}^{abc} \propto \mathrm{Tr}_{R_L}\left(\left\{T^a_{R_L}, T^b_{R_L}\right\} T^c_{R_L}\right) - \mathrm{Tr}_{R_R}\left(\left\{T^a_{R_R}, T^b_{R_R}\right\} T^c_{R_R}\right) , \tag{2.6}$$

where $T^{a,b,c}$ stand for the generators of all the gauge symmetries at play and the traces (Tr) are taken over all their representations $R_{L,R}$ accommodating left-handed (L) and right-handed (R) Weyl spinors, respectively. As such, demanding $\mathcal{A}^{abc} = 0$ for all a, b and c translates into algebraic equations for the charges involved (especially the Abelian ones).

Remarkably enough, playing this game for the hypercharges[3] of five different types of matter fields encountered in each of the three SM generations, one reveals the traditional hypercharge configuration (2.2) as the only non-trivial one[4] (modulo the overall normalization). As this is one of the central points of the discussion here, let us elaborate on these lines a little bit further. In fact, there is a neat trick which makes the inspection of the entire set of anomalies unnecessary; it can be shown [35,36] that the solution of the general problem is equivalent to considering just the $SU(2)_L \otimes U(1)_Y$ anomalies together with the constraints emerging from the requirement of the gauge invariance of the up and down quark and charged leptons' Yukawa terms in the SM Lagrangian. Working this out, one reveals the following set of equations for the hypo-

[3]Note that the logic of the game here is different from the usual calculations performed in order to *confirm* that the SM with hypercharges selected as in Eq. (2.2) is indeed anomaly free. In the current case, we take all $Y_{Q,u,d,L,e}$ as free parameters and search for all non-trivial (i.e., non-zero) solutions of the relevant system of algebraic equations.

[4]Note for completeness that there is in principle another (almost trivial) solution with $Y_Q = Y_L = Y_e = 0$ and $Y_u = -Y_d$ which, however, is rather bizarre and may be discarded either on phenomenological grounds or by elevating the SM gauge group to a full-fledged left-right symmetry (cf. Sect. 3.1.1); for more information see, e.g., [35].

thetically unknown values of[5] Y_Q, Y_u, Y_d, Y_L and Y_e (and Y_H of the Higgs field):

$$U(1)^3 \text{ anomaly:} \quad 6Y_Q^3 + 2Y_L^3 - 3Y_u^3 - 3Y_d^3 - Y_e^3 = 0\,, \tag{2.7}$$
$$SU(2)^2 \otimes U(1) \text{ anomaly:} \quad 3Y_Q + Y_L = 0\,, \tag{2.8}$$
$$\text{up-type-quark Yukawa:} \quad -Y_Q + Y_u - Y_H = 0\,, \tag{2.9}$$
$$\text{down-type-quark Yukawa:} \quad -Y_Q + Y_d + Y_H = 0\,, \tag{2.10}$$
$$\text{charged-lepton Yukawa:} \quad -Y_L + Y_e + Y_H = 0\,. \tag{2.11}$$

This comprises 5 equations for 6 unknowns; taking Y_H as a parameter the system is readily solved (barring the quasi-trivial option):

$$Y_Q = +\tfrac{1}{3}Y_H\,,\ Y_u = +\tfrac{4}{3}Y_H\,,\ Y_d = -\tfrac{2}{3}Y_H\,,\ Y_L = -Y_H\,,\ Y_e = -2Y_H\,. \tag{2.12}$$

Hence, the anomaly-free hypercharges compatible with the SM data are *homogeneous* in Y_H and, thus, their ratios are fixed to be fractions of small integers!

Cancellation of chiral anomalies in the SM+3 RH $SU(3)_c \otimes SU(2)_L$ singlets

Naïvely, one would expect little change in the game above if a RH neutrino field N_R (i.e., a full singlet with respect to the $SU(3)_c \otimes SU(2)_L$ of the SM) with an a-priori unknown hypercharge Y_N is admitted into the play; indeed, with an extra unknown there is also a new constraint stemming from the existence of the corresponding Yukawa term (2.1) so the situation should not change qualitatively at all. But it does! The point is that the resulting system of 6 equations, namely, (2.8)–(2.11) which remain intact, together with a slightly modified version of Eq. (2.7)

$$6Y_Q^3 + 2Y_L^3 - 3Y_u^3 - 3Y_d^3 - Y_e^3 - Y_N^3 = 0\,, \tag{2.13}$$

and an extra Yukawa-sector condition

$$-Y_L + Y_N - Y_H = 0\,, \tag{2.14}$$

is no longer independent. This may be readily seen by substituting into (2.13) for Y_Q from (2.8) and for Y_u, Y_d, and Y_e from (2.9)–(2.11) which yields

$$(Y_H + Y_L)^3 - Y_N^3 = 0\,. \tag{2.15}$$

This, however, is trivially fulfilled if (2.14) holds. Hence, the general solution here is parametrised by 2 leftover quantities, e.g., Y_H and Y_N, and it turns out to be *inhomogeneous* in either of the two. Normalizing the entire set so that $Y_H = +\frac{1}{2}$ one

[5] For simplicity, from now on we shall drop the chirality labels of the matter fields $f_{L,R}$ in Y_f's.

reveals

$$Y_Q = +\tfrac{1}{6} - \tfrac{1}{3}Y_N\,,\; Y_u = +\tfrac{2}{3} - \tfrac{1}{3}Y_N\,,\; Y_d = -\tfrac{1}{3} - \tfrac{1}{3}Y_N\,,\; Y_L = -\tfrac{1}{2} + Y_N\,,\; Y_e = -1 + Y_N \tag{2.16}$$

with no constraint of Y_N! Thus, there is a one-parametric class of admissible hypercharges out of which only some (those with $Y_N \in Z$) correspond to a quantized set.

Needless to say, this results has profound implications for the understanding of the 1 per 10^{21} level of neutrality of neutrons [34] – in the SM extension with 3 generations of RH neutrinos advocated in Sect. 2.1.1 the rationale for this to be the case is simply gone!

$U(1)_{B-L}$ as a gaugeable symmetry in the 3 RH neutrino context

The peculiar qualitative change in the behaviour of solutions to the SM chiral anomaly cancellation conditions in presence of three RH $SU(3)_c \otimes SU(2)_L$ singlets may be understood rather simply by inspecting the linear dependency of the "dequantized" hypercharges (Y'_f) of Eq. (2.16) on the (apriori unknown) real parameter Y_N. Interestingly, one can write Y'_fs as

$$Y'_f = Y_f + x(B-L)_f \tag{2.17}$$

where Y_f stands for the SM solution (2.12), $x = -Y_N$ is a real number and $(B-L)_f$ is the difference of the baryon and lepton number charges of the f-type fermions. Since, however, $U(1)_{B-L}$ can be promoted to a full-fledged gauge symmetry in the presence of 3 RH neutrinos[6], any linear combination of the non-anomalous Y_f of the SM and another non-anomalous and potentially local $B-L$ charge as in Eq. (2.17) is also a candidate for a non-anomalous $U(1)$ gauge symmetry generator that may play the role of an alternative SM hypercharge. Again, *the beautiful and profound mechanism of the SM (hyper)charge quantization by means of the chiral anomaly cancellation requirements is lost in the SM extension with three RH $SU(3)_c \otimes SU(2)_L$ fermionic singlets.*

2.1.3 Majorana fermions

There are two basic approaches to dealing with this rather unpleasant situation:

1. In principle, one does not need to add three copies of the fermionic $SU(3)_c \otimes SU(2)_L$ RH singlets. Hence, $B-L$ would not be anomaly free as a potentially local charge and the hypercharge redefinition freedom (2.17) would not exist. This

[6] Note that in the SM B as well as L are individually anomalous but $B-L$ is a non-anomalous global symmetry. However, the $(B-L)^3$ current is still anomalous and, hence, $B-L$ can not be promoted to a local symmetry in the SM. The presence of one extra $SU(3)_c \otimes SU(2)_L$ RH gauge singlet with $L=1$ per generation ensures that $A_{(B-L)^3}$ also vanishes.

is the case of, e.g., the two minimal alternatives to the canonical type-I seesaw based on the scalar or fermionic $SU(2)_L$ triplets (see also Sect. 2.2.2) which, by construction, deprive the (local) $B-L$ anomaly cancellation mechanism from a crucial ingredient.

2. Alternatively, one can entirely dismiss any thoughts about "gaugeability" of $B-L$ by writing down a Majorana mass term for the $SU(3)_c \otimes SU(2)_L$ RH fermionic singlets N_R which breaks this symmetry explicitly[7].

The possibility to have a massive spinor with only two dynamical components was first noticed by Ettore Majorana in 1937 [37]. It can be heuristically understood by counting the degrees of freedom: A massive charged spin-$\frac{1}{2}$ particle is described by a four-component object because the associated quantum field (Dirac spinor) must be able to describe the two opposite-charge states of the particles and antiparticles, together with the two helicity degrees of freedom for each. This number can be reduced from four to two by only two means; either giving up half of the helicity states (reducing the original Dirac to a Weyl spinor describing a massless particle) or giving up all charges[8] and, thus, any distinction between a particle and an antiparticle (a Majorana spinor).

Majorana mass term for RH neutrinos

Assuming that the extra RH spinors discussed above are completely $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ neutral the free part of the relevant Majorana Lagrangian can be written as

$$\mathcal{L}_M = i\overline{N_R}\gamma^\mu \partial_\mu N_R - \tfrac{1}{2} M_M N_R^T C N_R + h.c. \tag{2.18}$$

where M_M is, in general, a complex mass parameter (a symmetric complex matrix if more than a single N_R is considered), C stands for the spinorial charge conjugation matrix and the extra factor of $\frac{1}{2}$ in the second (Hermitean) term is there in order to compensate for the double counting due to the omnipresent $+h.c.$ associated with all other mass/Yukawa terms in the SM Lagrangian. A couple of comments may be worth here:

1. It is trivial to see that the asymptotic states created from the vacuum by the rising operators in N_R obey the standard relativistic free-particle dispersion relation of the form $p^2 - M_M^2 = 0$.

[7] Scenarios in which $U(1)_{B-L}$ is broken spontaneously are discussed at length in Chapter 3.

[8] More precisely, the internal symmetry representation under which such an object transforms should be real, admitting, for instance, a triplet of $SU(2)$ or an octet of $SU(3)$.

2. Due to the symmetry of (the matrix version of) M_M and the equality of the associated unitary matrices U acting on it from the left and right sides there is no residual freedom in the phase redefinition of U that would preserve the real diagonal form of M_M. This, in turn, makes it impossible here to absorb all 5 out of 6 complex phases of the "raw form" of $U^{\rm PMNS}$ in (2.3) as in the Dirac case of Sect. 2.1.1 but only 3. Thus, there are in principle three irremovable (i.e., physical) CP-violating phases left in the game (for three generations of N_R) which may play a role in situations/processes related to the Majorana nature of N_R (such as lepton number violation, see below).

Type-I seesaw mechanism and the typical RH neutrino mass scale

In practice, i.e., in case of the simplest SM extensions with three N_R's, the Majorana mass term (2.18) should not be alone in the full Lagrangian otherwise the SM neutrinos will still be massless. This potential issue is readily resolved by the addition of the Yukawa term (2.1); the complete structure of the N_R-Lagrangian then reads

$$\mathcal{L} \ni -Y_\nu \overline{L_L} N_R \tilde{H} - \tfrac{1}{2} M_M N_R^T C N_R + h.c. \tag{2.19}$$

Obviously, the key is the mixing of N_R with the neutral components of the SM L_L provided by the Yukawa term in (2.19). In the broken phase, the non-derivative part of the electrically neutral sector of the structure above can be readily rewritten as

$$\mathcal{L}^{\rm mass} = -\tfrac{1}{2} n_L^T C \mathcal{M} n_L + h.c.\,, \tag{2.20}$$

where $n_L^T = (\nu_L, N_R^c)^T$ and

$$\mathcal{M} \equiv \begin{pmatrix} 0 & m_D \\ m_D^T & M_M \end{pmatrix} \tag{2.21}$$

is a 6×6 complex symmetric matrix[9] which, besides the Majorana part M_M, includes a general complex 3×3 Dirac-type mass matrix defined as

$$m_D \equiv \frac{1}{\sqrt{2}} Y_\nu v\,. \tag{2.22}$$

The asymptotic neutrino states are obtained by a suitable diagonalisation of this matrix which, unlike in the Dirac case[10], yields in general 6 different (in size, not only in sign) eigenvalues describing 6 independent Majorana spinors!

[9]Here we assume three generations of N_R being added to the SM in order to treat all its three generations symmetrically. However, from the phenomenology point of view this is not strictly necessary as one can be well off even with just two copies of N_R.

[10]Note that for Dirac neutrinos $M_M = 0$ and the same diagonalisation procedure yields pairs of two-component eigenstates with eigenvalues of the same size but opposite CP parities which is a situation equivalent to having 3 four-component objects.

Dynamically, these eigenstates can be in most cases organised into two triplets representing a light and a heavy part of the neutrino sector. This is a consequence of the following simple argument:

- Unlike for m_D whose magnitude is tightly connected to the electroweak symmetry breaking, M_M is an explicit singlet mass term that has nothing to do with the electroweak scale v and, as such, it may be significantly larger[11] than v.

- For $|m_D| \ll |M_M|$ the diagonalisation of (2.21) proceeds in two steps, the first of which (a diagonalisation into 3×3 blocks) yields

$$\mathcal{M} \equiv \begin{pmatrix} m_\nu & 0 \\ 0 & M_N \end{pmatrix}, \tag{2.23}$$

with

$$m_\nu \equiv -m_D M_M^{-1} m_D^T + \mathcal{O}(|m_D|^3 |M_M|^{-2}) \tag{2.24}$$

and

$$M_N \equiv M_M + \mathcal{O}(|m_D|), \tag{2.25}$$

respectively. Thus, one ends up with three *light* SM-like neutrino mass eigenstates, dominated by the $SU(2)_L$-doublet components (ν_L) with just a tiny admixture of the singlet (N_R^c) ones, with a mass term in the form

$$\mathcal{L}^{\text{light}} = -\tfrac{1}{2}\nu^T C m_\nu \nu + h.c.\,, \tag{2.26}$$

and three heavy eigenstates with masses driven by

$$\mathcal{L}^{\text{heavy}} = -\tfrac{1}{2}N^T C M_N N + h.c.\,, \tag{2.27}$$

which are singlet-dominated and, thus, practically sterile with respect to the SM gauge interactions[12].

Note that the stipulated hierarchy between the Dirac and the RH Majorana mass terms $|m_D| \ll |M_M|$ makes it possible to attribute the smallness of m_ν to the $M_M^{-1}m_D^T$ factor in (2.24) and, thus, unload the enormous $|Y_\nu| \lesssim 10^{-11}$ suppression imposed on the Yukawa couplings in the Dirac neutrino case of Sect. 2.1.1. In principle, even $\mathcal{O}(1)$ entries in Y_ν are admissible[13] as long as M_M falls into the ballpark of

$$M_M \sim 10^{12-13}\text{GeV}\,. \tag{2.28}$$

[11] To this end, it is usually assumed that M_M is generated by the breaking of a higher gauge symmetry encompassing that of the SM in such a way that N_R transforms non-trivially under its action.

[12] It is worth emphasising here that this does not mean that the heavy Majorana neturinos do not entertain any interactions - one should not forget about the Yukawa ones!

[13] This assumption is actually easy to justify as in most SM extensions discussed in Chapter. 3 the neutrino Yukawa couplings are naturally correlated with those or the up-type quarks by the extended symmetries of the underlying Lagrangians.

Let us also mention that the inverse proportionality of m_ν to M_M earns this scheme the traditional name of "seesaw mechanism" [38]. However, one should bear in mind that, as simple as it is, this is by far not the only dynamical way to devise naturally light massive neutrinos in simple extensions of the SM, cf. Sect. 2.2.

2.1.4 Massive Majorana neutrino phenomenology

Remarkably enough, even with just a tiny mixing among the $SU(2)_L$ doublet and singlet components in the physical neutrino spectrum (driven by the off-diagonal $\mathcal{O}(|m_D|/|M_M|)$ factors in the unitary transformation bringing the defining-basis mass matrix (2.22) into the block-diagonal form (2.23)) the low-energy phenomenology of the seesaw schemes can differ drastically from that of the SM:

- *Perturbative lepton number violation:* The first obvious difference is the perturbative non-conservation of the lepton number owing to the shape[14] of Eq. (2.26) which, in the SM, is a good accidental global symmetry, at least at the perturbative level[15]. This may, in principle, exhibit itself in extremely-low-background processes where even a minuscule effect can be deciphered.

- *Leptonic CP violation:* Another common aspect of the stipulated Majorana nature of the light neutrinos is the presence of new sources of CP violation owing to the shape of the corresponding neutrino mass term (2.26). As we argued, there are in principle 3 CP phases in the leptonic sector; one of these comes as a full analogy of the CKM phase $\delta_{\rm CKM}$ and, as such, it is called the Dirac CP phase; the other two (which can be transferred back and forth between $U^{\rm PMNS}$ and the eigenvalues of M_M) are called Majorana.

Both these types of BSM effects may find rather spectacular incarnation in the laboratory experiments and in cosmology.

Neutrinoless double beta decay

Perhaps the most famous of such observables is the hypothetical neutrinoless double beta decay that should be undergone by some even-even nuclei for which a sequence of

[14] Note that both terms are needed to claim LNV: without the first one $L(N_R) = 0$ leads to a conserved L while without the second term the same happens for $L(N_R) = +1$. Needless to say, $L(\tilde{H}) \neq 0$ is not an option because H triggers the electroweak symmetry breaking.

[15] It is well known that this is not true at the non-perturbative level where topologically non-trivial extended field configurations (instantons, sphalerons) may be relevant in the path integration, thus leading to both baryon and lepton number non-conservation [12,39].

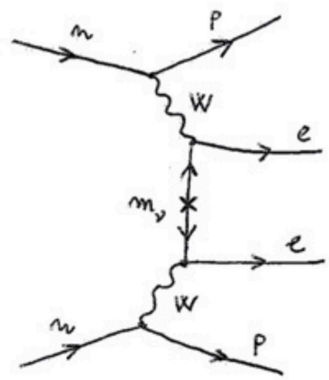


Figure 2.1: The light Majorana neturino mass contribution to the neutrinoless double beta decay hard process matrix element $\langle m_{ee} \rangle$ from the field theory perspective.

"standard" single beta decay transitions is kinematically forbidden. The most prominent of these systems are the isotopes ^{48}Ca, ^{76}Ge, ^{82}Se, ^{100}Mo, ^{130}Te and ^{136}Xe which are known to undergo a "standard" double beta decay (i.e., with two neutrinos in the final state) into their X^A_{Z+2} counterparts (or X^A_{Z-2} in the double β^+ channel) with enormous lifetimes and, at the same time, they can be found in the Earth's crust in non-negligible amounts. This makes it possible to speculate about the observability of an analogous process in which the pair of Majorana neutrinos "annihilate" in the core rather than escape the nucleus (cf. Fig. 2.1) which, in turn, may be identified by its very specific final state kinematics.

Interestingly, in most of the parameter space of the simple seesaw scheme above (as well as in other scenarios) the relevant effective leptonic matrix element $\langle m_{ee} \rangle$ exhibits a relatively universal lower limit which makes us hope that, with a bit of luck (and a steady support from the funding bodies) the relevant signal may be detected within the next few decades, see, e.g., [40] and references therein. Needless to say, this would be a phenomenal achievement as it would provide the first evidence of a glitch in the assumed perturbative-level symmetry structure of the Standard Model.

Let us finally note that this claim is often further supported by the so called "Scheckter-Valle theorem" [41] which stipulates that the observation of the neutrinoless double decay *implies* the Majorana nature of neutrinos, thus claiming their mutual equivalence! This can be heuristically understood by "dressing" the effective $d = 9$ "$0\nu 2\beta$ blackbox" operator by the SM fields which gives rise to a contribution to the light neutrino propagator of the Majorana structure, see Fig. 2.2.

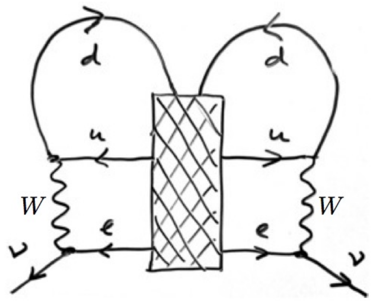


Figure 2.2: The SM four-loop dressing of the effective $d = 9$ "$0\nu2\beta$-blackbox" operator generates a Majorana mass term for light neutrinos.

Baryogenesis through leptogenesis

It is well known that the standard ΛCDM cosmology, married with the SM of particles and their interactions, fails miserably with the prediction of the baryon to photon number density ratio η_B, an all-important parameter for the modelling of the primordial nucleosynthesis during the "first three minutes" of the early Universe. Indeed, its measured value of roughly 6×10^{-9} (see e.g. [42] and references therein) is about 10 orders of magnitude above the SM upper limit based on the simple (and plausible) assumption of the baryon-antibaryon symmetry of the post-inflationary cosmic plasma.

The only reasonable way to reconcile the theory with the measurement is to revoke this assumption and, instead, look for a dynamical origin of the approximate $10^9 + 1$ to 10^9 baryon-to-antibaryon number density ratio that was to be established before the freeze-out of the baryonic populations in the early Universe. This, according to Sacharov [43], may have happened if:

1. Baryon number was not an exact symmetry of Nature;

2. C and CP were not good symmetries of the early Universe dynamics;

3. The system underwent an out-of-equilibrium (OOE) epoch.

The first of these criteria is trivial as one can cook a net baryon number from zero only if there are interactions capable of changing it. As for the second, this deals with the discrete symmetries correlating the pace of processes of a net B creation in one chirality channel with that of the B destruction in the parity-conjugated one. The third condition then ensures that the creation of B at some point is not entirely undone by a reverse process occurring at the same time at some other place.

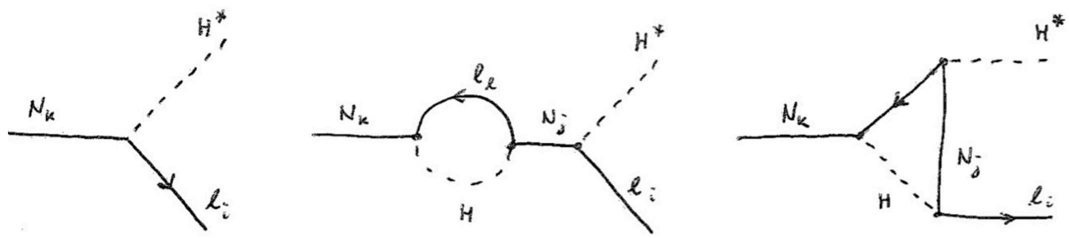


Figure 2.3: The leptonic CP-asymmetry generation in the out-of-equilibrium decay of the heavy neutral fermions in the SM extensions with RH neutrinos. The leading effect emerges from the interference between the tree-level and 1-loop diagrams.

Remarkably enough, the SM seems to fulfil all these conditions, at least qualitatively: 1. For $T > M_Z$ the sphaleron processes run fast and B violation is a standard feature of the hot SM plasma; 2. Both C and CP are, indeed, violated by the weak interactions and 3. the Universe may have fulfilled the OOE Sacharov's condition if the electroweak phase transition was strongly first order[16].

However, the situation turns out to be way less optimistic as neither condition 2. nor 3. are met in the Standard Model on the quantitative level. As for the former, CP violating dynamics turns out to be subject to strong screening in the hot and dense plasma, see e.g. [46]. Moreover, for $m_H \sim 125$ GeV the SM electroweak phase transition was very likely 2nd order and, hence, the yield of the asymmetry generated through the associated *electroweak baryogenesis* process is expected to be parametrically smaller than the desired $1 : 10^9$.

Interestingly, the heavy Majorana sector of the seesaw extension of the SM discussed in Sect. 2.1.3 offers a very elegant way out of this conundrum [47]. The basic idea is that the actual baryogenesis process may have been preceded by an epoch of *spontaneous leptogenesis* in which a net lepton number was created in the C and CP-asymmetric out-of-equilibrium decays of the heavy Majorana neutrinos (2.27) and, only later on, it was (partially) transferred into the baryons. A careful analysis of the central quantity of interest, namely, the CP asymmetry of the (lightest in most cases) heavy Majorana neutrino decays, see Fig. 2.3,

$$\varepsilon_{CP} = \frac{\Gamma(N \to \overline{L}H) - \Gamma(N \to LH^*)}{\Gamma(N \to \overline{L}H) + \Gamma(N \to LH^*)}\,, \tag{2.29}$$

reveals an upper limit [48, 49] on the CP asymmetry of the SM lepton production in

[16] In such a case the baryogenesis processes would take place at the surface of the asymmetric-phase bubbles expanding at almost the speed of light [44, 45] into the still symmetric false vacuum environment.

the form

$$\varepsilon_{CP} \leq \frac{1}{8\pi} \frac{M_1(m_3 - m_2)}{v^2}\,, \tag{2.30}$$

which, for the desired[17] $\varepsilon_{CP} \gtrsim 10^{-7}$ yields a lower limit on the mass of the lightest heavy neutrino in the form

$$M_1 \gtrsim 10^9\,\text{GeV}\,. \tag{2.31}$$

This, indeed, is perfectly compatible with the "seesaw" picture of the light neutrino mass generation discussed above, cf. Eq. (2.28).

2.2 Majorana neutrino mass as a $d = 5$ operator in the Standard Model

The arguments of the previous subsection suggest that the characteristic energy scale of the new dynamics that may be behind the light neutrino masses and other effects discussed in Sect. 2.1.4 is likely to be rather high, actually many orders of magnitude above the electroweak scale. This, at one hand, diminishes its potentially harmful impact to other precision data but, on the other hand, makes it generally difficult to test any specific shape of such a new dynamics in any but few particular channels.

2.2.1 Theories with vastly different scales

Technically, this has to do with one of the most general features of quantum physics often phrased as the "independence of the low-energy[18] observables on any 'new dynamics' with a parametrically larger characteristic energy scale". It is sometimes (not very rigorously) justified in the language of the uncertainty principle which, in principle, makes it possible for an "extremely heavy virtual particle pair"[19] to be "created" out of the quantum vacuum but, as a price for its "large mass", such an "object" may "exist" for only a very short "time" and, as such, it can not affect any observable at a substantial level.

[17] Note that the stipulated lower limit on ε_{CP} is slightly more conservative than the experimental constraint of η_B. The reason here is the presence of the so called "washout" effects corresponding to the LNV re-scattering processes undergone by the light leptons and antileptons during the asymmetry generation era which can effectively wipe out a significant portion of the asymmetric decay yield.

[18] This statement should perhaps be put into a better perspective by reminding the reader about how the notion of "low-energies" evolved over the last century - from Lord Rutherford's few MeV standpoint the CERN's LHC with its 14 TeV centre of mass energy would have likely been a machine beyond imagination, yet it may soon become a mere pre-accelerator for a yet more powerful monster - the FCC [50].

[19] The quotation marks here indicate the care required for the proper interpretation of the terms enclosed.

The Appelquist-Carazzone theorem

In the realm of the quantum field theory these expectations find their formal incarnation in the so called *decoupling theorem* by Thomas Appelquist and James Carazzone [51] which, in one of its possible formulations, stipulates the inverse-power dependency of the renormalized Green's functions of a low-energy sector of a certain theory (calculated in momentum renormalization schemes) on the masses characterising the heavy part of its spectrum. Several comments are perhaps worth making here:

1. The main scope of the theorem is to deal with UV-divergent loop graphs where the interchangeability of the integration and limit operations is not guaranteed. It is trivial though for finite graphs such as tree-level ones or loops with negative degrees of divergence.

2. The need for the renormalizability of the low-energy sector alone is easy to understand on a qualitative basis: In such a case, every singlet mass parameter of the complete theory may be viewed as a mere cut-off Λ in the momentum integration within the low-energy theory graphs; as such, it should disappear from the physics of the renormalizable low-energy sector in the $\Lambda \to \infty$ limit.

3. On the formal side, the amplitudes (and/or other quantities of interest such as masses, if such a distinction is desired[20]) of the low-energy theory entertain an explicit decoupling behaviour only in physical schemes, i.e., when the low-energy observables are parametrised in terms of physical quantities (or, at least, quantities defined in some of the momentum schemes). In schemes in which the definition of the counterterms is more or less ad-hoc (such as minimal subtraction etc.) with little reference to the underlying dynamics the relevant formulae do not need to exhibit any apparent decoupling of the heavy sector.

4. One more subtlety is perhaps worth pointing out here: it is implicitly understood that the dimensionless couplings between the light and heavy sectors do not grow more than logarithmically (as suggested by the renormalization group) with the heavy sector masses. Along with the renormalizability argument, this is the reason why, for instance, the top quark can not be formally decoupled from the Standard Model – indeed, its Yukawa coupling grows with m_t.

As an illustration of especially point 3 above consider a "light" ϕ^4 theory

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \lambda\phi^4 \tag{2.32}$$

[20] Personally, the author prefers to talk about renormalized Green's functions instead as they provide a unified language for all the aforementioned physical aspects.

coupled to a heavy version of the same

$$\mathcal{L}_\Phi = \frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}M^2\Phi^2 - \rho\Phi^4 \tag{2.33}$$

through a trilinear[21] interaction term

$$\mathcal{L}_{\phi\Phi} = \kappa\phi\Phi^2\,, \tag{2.34}$$

where κ is a dimensionfull parameter. The light-sector two-point 1PI Green's function in dimensional regularisation before imposing renormalization conditions then reads (displaying explicitly only the important heavy sector contribution[22])

$$\Gamma_{\phi\phi}(p^2) \propto p^2 - m^2 + \frac{\kappa^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \int_0^1 \mathrm{d}x \log\frac{M^2 - x(1-x)p^2}{4\pi\mu^2}\right) + \ldots + \delta Z_\phi p^2 + \delta m^2\,, \tag{2.35}$$

where δZ_ϕ and δm^2 are the "light" theory counterterms. These are in general different in different renormalization schemes:

- The on-shell scheme renormalization conditions

$$\Gamma(p^2 = m^2) = 0\,, \quad \frac{\mathrm{d}}{\mathrm{d}p^2}\Gamma(p^2 = m^2) = 1 \tag{2.36}$$

  which, among other things, immediately identify the physical mass of the light scalar with m, yield

$$\begin{aligned}\delta Z_\phi &= -\frac{\kappa^2}{16\pi^2}\log\frac{x(1-x)}{M^2 - x(1-x)m^2}\,,\\ \delta m^2 &= -\frac{\kappa^2}{16\pi^2}\left[\frac{1}{\varepsilon} - \int_0^1 \mathrm{d}x\left(\log\frac{M^2 - x(1-x)m^2}{4\pi\mu^2} - \log\frac{m^2 x(1-x)}{M^2 - x(1-x)m^2}\right)\right] + \ldots\,,\end{aligned}$$

  and thus

$$\Gamma^{\mathrm{OS}}_{\phi\phi}(p^2) \propto p^2 - m^2 - \frac{\kappa^2}{16\pi^2}\int_0^1 \mathrm{d}x \log\frac{M^2 - x(1-x)p^2}{M^2 - x(1-x)m^2} + \frac{(p^2 - m^2)x(1-x)}{M^2 - x(1-x)m^2} + \ldots\,. \tag{2.37}$$

  It is trivial to see from here that in the $M \to \infty$ limit the contribution of the heavy sector entirely drops from Γ^{OS}, i.e., the decoupling of the heavy sector from the light one is apparent.

[21] Note that the specific form (2.34) of the cross-talk between the light and heavy sectors of the scheme has been chosen only for the calculational simplicity; other choices (such as $\phi^2\Phi^2$) lead to the same conclusions.

[22] Here and in what follows we shall often use "rationalised natural units" in which all the multiplicative factors irrelevant for the merit of the argument are set to 1 and the additive ones to 0.

- On the contrary, in the minimal subtraction (MS) scheme $\delta Z_\phi = 0$ and

$$\delta m^2 = -\frac{\kappa^2}{16\pi^2}\frac{1}{\varepsilon}\,. \tag{2.38}$$

Hence, the renormalized two-point 1PI Green's function in the MS scheme reads

$$\Gamma^{\rm MS}_{\phi\phi}(p^2) \propto p^2 - m^2 - \frac{\kappa^2}{16\pi^2}\int_0^1 {\rm d}x \log\frac{M^2 - x(1-x)p^2}{4\pi\mu^2} + \ldots \tag{2.39}$$

For $M^2 \gg p^2, m^2$ the root of this function (corresponding to the one-loop physical mass of ϕ) is clearly far from $p^2 \sim m^2$; actually, in the $M \to \infty$ limit it can be readily isolated at around

$$p^2 \sim m^2 + \frac{\kappa^2}{16\pi^2}\log\frac{M^2}{4\pi\mu^2} \equiv m^2_{\rm phys} \tag{2.40}$$

which, formally, grows with M. Note, however, that the decoupling theorem still works here, only in a slightly different manner: writing the RHS of Eq. (2.39) in terms of $m^2_{\rm phys}$ rather than m^2 one is taken to the situation which (up to a redefinition of certain parameters) is qualitatively identical to the on-shell setting above.

Low-energy effective theories and their symmetries

From a wider perspective, the A-C theorem suggests a very practical approximative approach to the theories with vastly different scales. The inverse-power proportionality to the heavy sector masses in the low-energy Green's functions of the light sector (as observed in (2.37) with the log expanded) justifies an effective parametrisation of the low-energy amplitudes in terms of local operators with positive powers of M in denominators. These, due to the gauge-singlet nature of the contractions corresponding to the complete-theory internal heavy field propagators must respect all the symmetries of the full theory. These operators are equipped with generic numerical coefficients to be determined by matching such an *effective Lagrangian* [52] to the complete theory amplitudes in the low-energy regime.

In general, such a matching procedure consists in integrating out the heavy degrees of freedom from the complete theory in its path-integral formulation[23]. This, besides providing relations between the fundamental and the effective couplings also highlights those classes of effective operator structures which are "available" within any specific scheme of interest.

[23] In the canonical formalism, this is usually approached by solving the (classical) equations of motion for the heavy fields and substituting the result back into the Lagrangian. The error committed by working at the classical level here can be shown to be comparable with the effects of higher orders in the perturbative expansion, cf. [53].

2.2.2 Standard Model as a low-energy effective theory

From this perspective, any perturbative high-energy extension of the SM can be mapped – in the low-energy regime – to (a specific subset of) the complete list of $d > 4$ effective operators. At each level of dimensionality, such a list contains a number of structures which are supposedly independent with respect to the transformations that should leave the physics contents of the theory intact (e.g., total-derivative completion, use of the equations of motion, Fierz transformation etc.). A systematic classification of such "independent operator bases" for various dimensions can be found in, e.g., [54–56].

d=5 Weinberg's operator and its renormalizable tree-level openings

As an example of the practical application of these principles let us turn our attention back to the situation discussed in Sect. 2.1.3, namely, to the low-energy behaviour of the SM with three RH neutrinos with masses way above the EW scale, cf. (2.28) or (2.31). The relevant part of the Lagrangian density

$$\mathcal{L}_{N_R} = i\overline{N_R}\gamma^\mu\partial_\mu N_R - \tfrac{1}{2}M_M N_R^T C N_R - Y_\nu \overline{L_L} N_R \tilde{H} + h.c. \tag{2.41}$$

yields the Euler-Lagrange equations of motion for N_R in the low-energy regime (i.e., ignoring the kinetic term contribution) in the form[24]

$$N_R = -M_M^{-1} Y_\nu^T (H^T i\sigma_2 L_L)^c + \ldots \tag{2.42}$$

This, substituted back into Eq. (2.41), provides a $d = 5$ effective structure

$$\mathcal{L}_{\text{eff}} = \kappa (L_L i\sigma_2 H)^T C (L_L i\sigma_2 H) + h.c.\,, \tag{2.43}$$

where the brackets encompas the $SU(2)$ doublet contractions (with the transposition acting in the spinorial space) and $\kappa \equiv -Y_\nu M_M^{-1} Y_\nu^T/2$. Equation (2.43) corresponds to the famous non-renormalizable operator identified by Weinberg in his seminal paper [9].

Remarkably enough, in the same study it was also shown that this structure is *unique* at the $d = 5$ level of the higher-dimensional operator ladder. This, however, means that in the broken phase one should recover the $d = 3$ Majorana operator for the light neutrinos in the form (2.26) which, given (2.22) and (2.24), is indeed the case.

[24] Here it turns out to be way more convenient to rewrite all the Dirac-type contractions into the Majorana formalism.

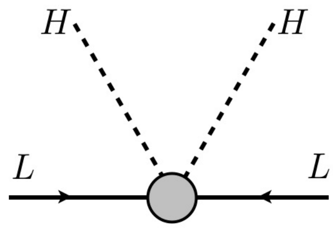


Figure 2.4: A Feynman diagram visualisation of the $d = 5$ Weinberg's operator.

Seesaw type I: From this perspective one can view the SM extension with the RH neutrinos as a specific example of the high-scale renormalizable dynamics that, in the low-energy regime, exhibits itself (predominantly) in the form of the $d = 5$ interaction (2.43) which, among other things, generates a Majorana mass term for the light neutrinos. Historically, this scheme comes under the name of the "Type-I seesaw mechanism" [38, 57]. It is interesting (though perhaps not surprising) that the type-I

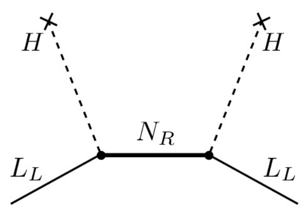


Figure 2.5: The "microscopic" structure of the light neutrino mass within the type-I seesaw scheme.The N_R field is an $SU(2)_L$-singlet fermion with hypercharge 0.

seesaw dynamics is not the only way to realize the Weinberg's $d = 5$ operator in Fig. 2.4 within a renormalizable high-energy dynamical scheme. At the level of its tree-level "openings" there are actually two more options.

Seesaw type II: Instead of a singlet fermion of Sect. 2.1.3 one can employ a scalar [58, 59] in a "t-channel" type of structure as depicted in Fig. 2.6. An $SU(2)_L$ scalar triplet Δ_L with hypercharge +1 can be coupled to the pair of L_L's via a Yukawa term like (neglecting generation structure for simplicity)

$$\mathcal{L}^{\mathrm{Y}}_{\Delta_L} = Y_\Delta L_L^T C i\sigma_2 \vec{\sigma} L_L \vec{\Delta}_L + h.c.\,, \tag{2.44}$$

which, together with a super-renormalizable coupling to the Higgs doublet pair of the form

$$\mathcal{L}^{\mathrm{H}}_{\Delta_L} = \mu H^T i\sigma_2 \vec{\sigma} H \vec{\Delta}_L + h.c. \tag{2.45}$$

and the corresponding kinetic and mass terms

$$\mathcal{L}_{\Delta_L}^{\text{free}} = \partial^\mu \Delta_L^\dagger \partial_\mu \Delta_L - M_\Delta^2 \Delta_L^\dagger \Delta_L, \tag{2.46}$$

yield again[25] a structure like that in Eq. (2.43) with $c = \mu Y_\Delta^2 / M_\Delta^2$.

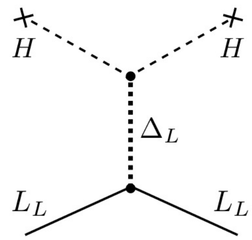


Figure 2.6: The "microscopic" structure of the light neutrino mass within the type-II seesaw scheme. The Δ_L field is an $SU(2)_L$-triplet scalar with hypercharge +1.

It is worth noting that the triplet nature of Δ_L with respect to the $SU(2)_L$ is enforced by the doublet nature of L and H in (2.44) and (2.45) - the antisymmetry of singlet contractions is incompatible with the gauge structure of either of these terms.

Another comment concerning the role of different scales in the game is perhaps worth here: unlike in the type-I case the type-II seesaw effective light neutrino mass is driven by two mass parameters (μ and M_Δ). This, for $\mathcal{O}(1)$ couplings, makes it possible to get m_ν in the eV ballpark even if M_Δ is relatively light (TeV-scale) as long as μ is appropriately small. Hence, the type-II seesaw mechanism may work in a mode in which it leaves behind interesting collider signatures! An interested reader can find more information in, e.g., [60] and references therein.

Seesaw type III: Interestingly, yet another tree-level renormalizable opening of the Weinberg's operator (2.43) can be devised with an $SU(2)_L$-triplet fermion [61] instead of the type-I singlet. This so called type-III seesaw scheme (see Fig. 2.7) differs from the type-I variant in two aspects only:

- The "mediator" is a gauge non-singlet field and, as such, its excitations can be produced at the colliders if they are light enough (i.e., for M_F in the TeV ballpark). This, however, comes for the price of very tiny Yukawa couplings.
- With a perturbatively conserved lepton number assigned to F the $B-L$ local anomaly does not vanish. Hence, $U(1)_{B-L}$ can not be gauged in this scheme.

[25] Note that, to this end, the Pauli matrices in (2.44) and (2.45 can be eliminated by means of the relevant completeness relations.

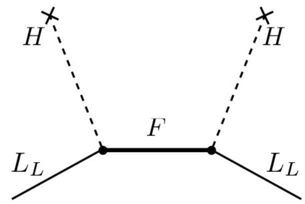


Figure 2.7: The "microscopic" structure of the light neutrino mass within the type-III seesaw scheme. The F field is an $SU(2)_L$-triplet fermion with hypercharge 0.

Naturalness of perturbative lepton number violation in BSM physics

From a wider perspective, it is very interesting that the evidence of non-zero neutrino masses – the first solid experimental signal of physics beyond the Standard Model – together with the desire to preserve the beautiful mechanism of charge quantization by anomalies, point straight towards the Weinberg's operator (2.43), the lowest-dimensional structure in the effective-operator ladder of the SM. In this respect, the perturbative lepton number violation, as arcane as it looks at the renormalizable SM level, seems to be absolutely natural for its simplest phenomenological extensions.

2.3 Standard Model at $d = 6$ level and perturbative baryon number violation

Without much of exaggeration one can perhaps even say that the observation of the LFV effects as the first signal of the BSM physics was, in fact, *expected* as it corresponds to the least suppressed $d = 5$ layer revealed upon opening up the Pandora's box of the SM effective operators. There is no reason, however, for the story to stop there, the more that the number of independent gauge-invariant effective structures grows exponentially with d. Actually, there are almost 60 different operator types in the SM at the $d = 6$ level, about 1000 at $d = 8$ and so on [56].

However, in what follows we shall dive just one more level beneath the Weinberg operator discussed so far, i.e., we shall discuss (a carefully selected subset[26] of) the $d = 6$ SM effective operators only. For our current needs it will be sufficient to divide them into three basic classes:

1. *Operators generating effects which are strictly forbidden (at the perturbative level) in the SM.* With essentially no SM background these structures have a great

[26] A full discussion of all classes of the SM $d = 6$ operator and the specific structures within would go way beyond the scope of this thesis; an interested reader can find further information in, e.g., [54, 55].

potential to provide “clean” signals in the experimental environment of any kind at any of the three major fronts of the HEP research (energy, intensity, precision). As we shall see, they typically break accidental global symmetries of the SM (such as baryon and lepton numbers). Actually, it is quite remarkable that this set is not empty!

2. *Operators corresponding to effects which are tree-level forbidden in the SM.* In this class there are typically operators triggering, e.g, flavour-changing neutral currents or CP effects (such as electric dipole moments) beyond those in the Standard Model. As for the former, these correspond to effective 4-fermion interactions while the latter come in the form similar to the Pauli operator in QED (modulo the extra Higgs field necessary to satisfy the $SU(2)_L \otimes U(1)_Y$ symmetry requirements). The associated effects are often heavily suppressed in the SM[27] and, hence, the backgrounds in the corresponding experimental searches are likely to be under control.

3. *Operators corresponding to effects with significant SM backgrounds.* This is the richest of the three sets. However, from the perturbative expansion perspective, the operators here typically yield higher-order corrections to the SM amplitudes and, hence, their effects are usually buried under a significant burden of the SM background.

For practical reasons, in the rest of this writing we shall fully focus on the first class above. The point is that the absence of any perturbative backgrounds makes it possible to look for the associated physics even if the relevant suppression scale is very far above the electroweak scale, perhaps even close to the Planck mass.

As for the others, let us just note that the second type is a typical encounter in supersymmetric theories [62–64] where the effective suppression scale is, for various reasons, expected to be in the TeV ballpark (and, hence, often becomes a problem rather than a feature of interest). Finally, almost any new physics with coloured heavy fields contributes to the third class.

2.3.1 B and $L-$ violating $d = 6$ operators and their phenomenology

There are basically four different types of $d = 6$ baryon and lepton number violating structures quoted in the recent[28] classification studies [55,56], see Table 2.1. From their structure, it is clear that these operators can have non-zero matrix elements between

[27]Recall, for instance, that FCNC's appear at the one-loop level in the SM while the leptonic EDMs

name	structure
$\mathcal{O}_1$	$\overline{u^c_L}\gamma^\mu Q_L\ \overline{d^c_L}\gamma_\mu L_L$
$\mathcal{O}_2$	$\overline{u^c_L}\gamma^\mu Q_L\ \overline{e^c_L}\gamma_\mu Q_L$
$\mathcal{O}_3$	$Q^T_L C Q_L\ Q^T_L C L_L$
$\mathcal{O}_4$	$u^{cT}_L C u^c_L\ d^{cT}_L C e^c_L$

Table 2.1: One of many possible representations of the gauge structure of the four different types of BLNV terms one can write into the SM effective Lagrangian at the $d = 6$ level. For simplicity, all fields have been written in the left-handed chirality convention and all indices have been suppressed; C is the charge conjugation matrix. Note that the above structures may be recast in many different ways using, e.g., Fierz identities or utilising the Majorana instead of the Dirac language; for instance, $\mathcal{O}_1$ is identical to Q_{duq} in [55] etc.

baryons and leptons with, possibly, extra mesons on each side. Thus, they change baryon and lepton numbers by 1 unit each, with $B - L$ left intact[29].

Proton decay as a spectacular signal of $d = 6$ perturbative BLNV

On the phenomenology side, this selection rule suggests that baryons can transmute into leptons (+ objects with zero B and L) at the $d = 6$ level of the effective operator ladder! The most spectacular of such processes is then the decay of the lightest of (color-singlet) baryons, the proton, into an antilepton and a meson. While the first is required by the statistics and phase space considerations, the latter is needed for the energy-momentum conservation. Kinematically, the processes with the maximum phase space available for the products are those with $\pi^+\overline{\nu}$ and $\pi^0 e^+$ in the final states, but there is a plethora of other options stretching up to $K^*\mu^+$ or even $\eta' e^+$ two-body decays. Note also that proton decay into τ + *anything* is impossible!

can be generated at four loops only, cf. Sect. 2.1.1.

[28] Note that, until recently, the situation concerning the correct counting of individual operators, their "types", or the numbers of independent terms needed for encompassing all of their effects at the level of Lagrangians was far from clear even for such a low dimensionality as $d = 6$; in this respect, the steady progress on new techniques such as those based on the Hilbert series (see, e.g., [65] and references therein) entertained in the last decade eventually lead to a full reconciliation among different claims.

[29] It is interesting to compare this with the $\Delta B = 0$ and $\Delta L = 2$ structure of the $d = 5$ Weinberg operator.

Proton decay searches and lifetime limits

The history of (so far unsuccessful) proton decay searches and the corresponding lower limits on its lifetime is extremely rich and there is not enough room here to pass through it. Instead, in what follows, we shall recapitulate just the main phases of this undertaking which stretch from the first simple (essentially table-top) experiments up to the current monstrous facilities such as Super-Kamiokande [27].

Perhaps the first solid quantitative information about proton lifetime came in 1950's from Maurice Goldhaber who noticed that an overly fast proton disintegration in living tissues would have devastating effects. His estimate of $\tau_{p^+} \gtrsim 10^{19}$ years (based on the contemporary estimate for lethal dose for humans at the level of about 10^{14-15}MeV absorbed per year) made it clear that backgrounds (radioactive as well as cosmic) would play a significant role in any dedicated p-decay searches; subsequently, experimentalists were "driven under the ground" for decades. Among the first dedicated searches one should mention namely the experiment by F. Reines and M. Crouch [66] from early 1970's which took place in a 3200m deep "East Rand" Gold Mine in South Africa and pushed the naïve Goldhaber's estimate up to about 10^{30}years [67].

With such an enormous lifetime limit the only chance for progress became multi-ton-scale experiments in which, optimally, the target body is transparent enough to admit full-volume detection. Hence, the era of kiloton-scale water-Cherenkov (WC) detectors begun (with KamiokaNDe [25], IMB [26], Super-K [27] and the upcoming Hyper-K [68] representing just few of these gargantuous instruments) and, to date, we are still harvesting the results obtained with this class of machines[30]. Besides price considerations, it is namely their universality (all of them are great neutrino detectors) and relative simplicity of the p-decay final state discrimination which makes this technology the primary method of experimental research in the field.

For illustration, consider the WC signature of the "golden channel" $p \to \pi^0 e^+$ process schematically depicted in Fig. 2.8: with almost a GeV of energy available e^+ is produced highly relativistic and a characteristic Cherenkov cone points in the direction of its flight before it stops and annihilates with an electron in water, only to produce a pair of delayed gammas. On the other side, π^0 decays almost instantaneously into a pair of hard γ's which produce two more (fuzzy) cones from the associated electromagnetic cascades. It means that the event is in principle fully contained within the detector volume and the parent mass can be reconstructed with a good precision. This is

[30] It is perhaps worth mentioning that recently the liquid Argon time-projection chambers (LArTPC) were put forward (e.g., LBNE/LBNF/DUNE [69]) as viable and affordable alternative to water Cherenkov detectors with several distinct features which make them even better in some channels (especially those involving charged Kaons in the final state).

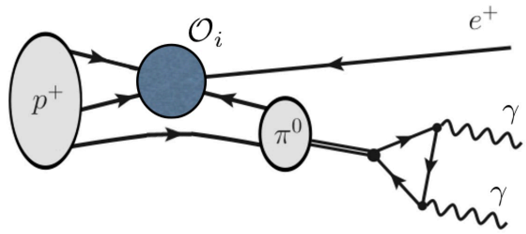


Figure 2.8: The "golden channel" proton decay $p \to \pi^0 e^+$ anatomy and evolution in the Standard Model effective theory with the underlying $d = 6$ operators $\mathcal{O}_i$ of Table 2.1.

important namely for the suppression of the backgrounds dominated by atmospheric neutrino effects.

The current best limits from the Super-K searches are at the level of 10^{34}years for $p \to \pi^0 e^+$ and about an order of magnitude milder for those with neutrinos in the final state ($p \to \pi^+ \overline{\nu}$, $p \to K^+ \overline{\nu}$ etc.); the projected sensitivity of the future almost megaton-scale facilities (such as Hyper-K) should about ten times better.

These numbers can be readily translated into very stringent upper limits on the dimensionfull coefficients of the $d = 6$ operators of Table 2.1 which turn out to be typically in the $(10^{15}\text{GeV})^{-2}$ ballpark. Hence, the searches for baryon number violation signals in particle decays represents a great opportunity to learn about possible new dynamics at unprecedentedly large scales.

2.3.2 Tree-level renormalizable openings of the $d = 6$ BLNV operators

What are the implications of the numbers above for the structure and characteristic scale(s) of the SM extensions that can possibly generate the B and L violating operators of Table 2.1? To understand this, one has to look at their renormalizable openings as we did it with the Weinberg's operator in Sect. 2.2.2.

1. $\mathcal{O}_1$: In Table 2.1, $\mathcal{O}_1$ was written in the form of a product of two vector currents so it is natural to attempt to open it "in the middle", i.e., work with $j_1^\mu \equiv \overline{u_L^c}\gamma^\mu Q_L$ and $j_2^\mu \equiv \overline{d_L^c}\gamma^\mu L_L$ which, as far as their gauge quantum numbers are concerned, transform as: $[j_1] = (\mathbf{3} \otimes \mathbf{3}, \mathbf{2}, +\frac{5}{6})$ and $[j_2] = (\mathbf{3}, \mathbf{2}, -\frac{5}{6})$. Hence, at the renormalizable level, they can both couple to a vector boson V_1 transforming like[31] $(\overline{\mathbf{3}}, \mathbf{2}, +\frac{5}{6}) \oplus (\mathbf{3}, \mathbf{2}, -\frac{5}{6})$. However, this is not the only way to generate $\mathcal{O}_1$ at the

[31] Here, in the anticipation of a future embedding of V_1 into an adjoint representation of a gauge group G of a renormalizable Yang-Mills theory we write both components of its stipulated decomposition with respect to the SM subgroup of G.

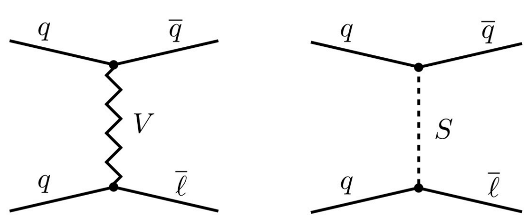
tree level – performing a Fierz transformation the same operator can be written as $u_L^{cT} C d_L^c \; Q_L^T C L_L$ and the same logic as above leads to a scalar colour triplet mediator $S_1 = (\mathbf{3}, \mathbf{1}, -\frac{1}{3})$. Actually, there is even a third option corresponding to a swapping of Q_L and L_L in the scalar contraction above (thus working with $u_L^{cT} C d_L^c \; L_L^T C Q_L$ instead) which does not change anything about the scalar mediator but, on the Fierz-equivalent side, opens another option for a vector mediator, namely, $V_2 = (\overline{\mathbf{3}}, \mathbf{2}, -\frac{1}{6}) \oplus (\mathbf{3}, \mathbf{2}, +\frac{1}{6})$. Thus, the single effective structure we begun with can be obtained in three different ways!

2. $\mathcal{O}_2$: Again, looking first at its vector × vector form one can identify V_1 as a possible mediator and, on the Fierz-transformed side, Δ_c as its scalar counterpart corresponding to the reshuffled structure like $u_L^{cT} C e_L^c \; Q_L^T C Q_L$. There is, however, no second vector option as with $\mathcal{O}_1$ because the permutation of the doublets therein does not bring anything new.

3. $\mathcal{O}_3$: Since this structure is composed of only two different fields there is not much of a room for any Fierz metamorphosis here and, thus, the only structures to look at are $Q_L^T C Q_L$ and $Q_L^T C L_L$. Since the former transforms like $(\mathbf{3} \otimes \mathbf{3}, \mathbf{2} \otimes \mathbf{2}, +\frac{1}{3})$ while the latter as $(\mathbf{3}, \mathbf{2} \otimes \mathbf{2}, +\frac{1}{3})$ there are only two types of scalar mediators that can generate it, namely, the notorious $S_1 = (\mathbf{3}, \mathbf{1}, -\frac{1}{3})$ and a new $S_2 = (\mathbf{3}, \mathbf{3}, -\frac{1}{3})$.

4. $\mathcal{O}_4$: Here, finally, we encounter a structure that can not be Fierz-transformed in any way but which offers permutation of the three fields within. Dividing it into $u_L^{cT} C u_L^c$ and $d_L^{cT} C e_L^c$ offers $S_3 = (\mathbf{3}, \mathbf{1}, -\frac{4}{3})$ as a possible mediator while in the permuted case with $u_L^{cT} C d_L^c$ and $u_L^{cT} C e_L^c$ one again recovers S_1.

All this information is comprehensively covered in Table 2.2. At low energies these fields would play the role of the internal lines of the Feynman diagrams corresponding to the quark-level proton decay hard processes such as



. (2.47)

These, at the level of amplitudes, yield a universal structure like $\mathcal{A} \sim k^2/M^2$, where k stands for the gauge or Yukawa couplings and M denotes the corresponding gauge or scalar mediator mass. At the hadronic level, this eventually boils down to

$$\Gamma(p \to \text{antilepton} + \text{meson}) \sim \frac{k^4}{M^4} m_p^5 \,, \tag{2.48}$$

mediator	spin	quantum numbers	operators generated	couplings
V_1 (X)	vector	$(\overline{\mathbf{3}},\mathbf{2},+\frac{5}{6})\oplus(\mathbf{3},\mathbf{2},-\frac{5}{6})$	$\mathcal{O}_1$, $\mathcal{O}_2$	gauge
V_2 (Y)	vector	$(\overline{\mathbf{3}},\mathbf{2},-\frac{1}{6})\oplus(\mathbf{3},\mathbf{2},+\frac{1}{6})$	$\mathcal{O}_1$	gauge
S_1 (Δ_c)	scalar	$(\mathbf{3},\mathbf{1},-\frac{1}{3})$	$\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_3$, $\mathcal{O}_4$	Yukawa
S_2 (Δ_c')	scalar	$(\mathbf{3},\mathbf{3},-\frac{1}{3})$	$\mathcal{O}_3$	Yukawa
S_3 (Δ_c'')	scalar	$(\mathbf{3},\mathbf{1},-\frac{4}{3})$	$\mathcal{O}_4$	Yukawa

Table 2.2: Possible tree-level mediators of the $d = 6$ BLNV effective operators of Table 2.1 in renormalizable extensions of the SM. The symbol in the brackets in the first column corresponds to a more traditional notation we shall conform to in the rest of the text. The words "gauge" and/or "Yukawa" refers to the type of the coupling to the matter currents the mediator entertains.

where the proton mass with the fifth power emerges on the dimensional grounds. Thus, for $\mathcal{O}(1)$ couplings[32], the experimental lower limit of the order of 10^{33-34} years on the proton lifetime quoted above translate to a generic lower limit for the masses of the mediators underpinning the $d = 6$ BLNV effective operators at the level of about 10^{15} GeV!

2.4 High-energy convergence of the SM gauge couplings

With such an enormous scale, residing some 12–13 orders of magnitude above what is currently accessible at the Earth-based accelerator experiments, there seems to be essentially no way to probe the relevant BLNV dynamics directly (at least for those schemes whose dominant effects exhibit themselves at the level of $d = 6$ operators[33]).

However, this does not mean that there is no way to say anything qualified about the likelihood of observability of the BLNV processes like p-decay or about the shape of the underlying interactions. Remarkably enough, the humongous scale of 10^{15} GeV identified from the lower limits on p decay in the preceding section is not that far from the stipulated 10^{13} GeV ballpark where the seesaw dynamics "prefers" to reside, cf. (2.28), and it is certainly compatible with the Davidson-Ibarra limit (2.31).

[32] Obviously, the lower limit on M quoted here can be reduced considerably if small couplings are involved.

[33] It is perhaps worth noting that in the case that the BLNV effects were due to $d > 6$ operators the lower limits on the associated effective scale(s) get reduced considerably. However, one should then provide a good argument for the absence/irrelevance of the $d = 6$ effects.

2.4.1 Renormalized gauge couplings and their evolution in the SM

Remarkably enough, there is a hint on such a large scale even in the very structure of the old good Standard Model itself. It is related to the evolution of its running couplings which, indeed, exhibit a nice convergence feature at about 10^{15-16} GeV!

The simplest way to see this is to write down and solve the renormalization group equations for the two non-abelian[34] gauge couplings $g_3 \equiv g_s$ and $g_2 \equiv g$

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} g_i = \beta_i \,, \quad i = 2, 3 \,, \tag{2.49}$$

with β_i denoting the relevant β-functions, together with the corresponding electroweak-scale initial conditions (imposed at $\mu = M_Z$)

$$g_3(M_Z) = \sqrt{4\pi\alpha_s(M_Z)} \,, \quad g_2(M_Z) = \frac{e(M_Z)}{\sin\theta_W(M_Z)} \,. \tag{2.50}$$

In the formulae above α_s encodes the QCD coupling strentgh, e is the electric charge and θ_W is the weak mixing angle. The β-functions are, at the lowest non-trivial order, given by the infamous formula of Gross, Wilczek and Politzer [70, 71]

$$\beta_G = \frac{1}{16\pi^2}\left[-\frac{11}{3}C_2^G + \frac{2}{3}\sum_{f_W} T_2^G(R_{f_W}) + \frac{1}{3}\sum_{s_c} T_2^G(R_{s_c})\right] g^3 + \ldots \equiv \frac{1}{16\pi^2} b_G \, g^3 + \ldots \tag{2.51}$$

where C_2^G stands for the quadratic Casimir of the group factor G and $T_2^G(R)$ denotes the index of the G-representations R_{f_W} and R_{s_c} hosting the Weyl fermions and complex scalars, respectively.

The system (2.49) receives a particularly simple form in the logarithmic coordinates $t = \frac{1}{2\pi}\log\mu/M_Z$. Defining $\alpha_i^{-1} \equiv 4\pi/g_i^2$ one arrives at

$$\frac{\mathrm{d}}{\mathrm{d}t}\alpha_i^{-1} = -b_i \,, \quad i = 2, 3 \,, \tag{2.52}$$

with a solution

$$\alpha_i^{-1}(t) = \alpha_i^{-1}(t_0) - (t - t_0) b_i \tag{2.53}$$

which, for $b_3 = -7$ and $b_2 = -\frac{19}{6}$ calculated theoretically and $\alpha_3^{-1}(M_Z) \sim 8.6$ and $\alpha_2^{-1}(M_Z) \sim 29.9$ from the experiment, yields a picture like in Fig. 2.9. This, however, is highly interesting for at least two reasons:

[34] For the time being we shall ignore the abelian (hypercharge) SM coupling g', the reason is that unlike for the non-abelian generators there is no natural normalization scheme for the abelian one and, thus, from the pure SM perspective, g' can be arbitrarily rescaled along with the hypercharge generator Y as long as their product is preserved.

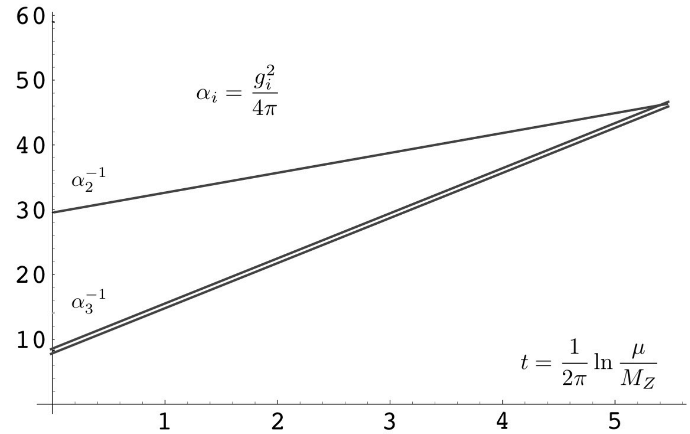


Figure 2.9: Evolution of the (inverse squared) running non-abelian gauge couplings in the Standard Model assuming no new dynamics kicking in throughout the evolution. The strengths of the two interactions converge to the same value at $t \sim 5.5$ which corresponds to roughly $\mu \sim 10^{16}$ GeV.

- The "effective strengths" of the gauge interactions governed by the $SU(2)_L$ and $SU(3)_c$ group factors tend to converge at high energies and turn out to be the same at a certain very large scale which is practically identical with the lower limit on the mass of the $d = 6$ BLNV mediators discussed above.

- Note that this feature should *not* be viewed as a trivial consequence of different slopes of two straight lines stretching in a plane; actually, in order to get an intersection in the desired semi-plane (for $t > 0$), let alone in the very interesting ballpark of $\mu \equiv M_G \sim 10^{16}$ GeV, the initial conditions (i.e., the experimental data) and the values of the relevant β-functions (encoded in the structure of the SM) must conspire to a high degree!

Nevertheless, what is the real physics content of Fig 2.9? Indeed, extending the two curves beyond $\mu \sim M_G$ makes them diverge again; from this perspective their intersection at about 10^{16} GeV would have to be interpreted as a mere coincidence. However, what if their slopes change at the point of intersection due to a change in the field content of the theory such that the two curves continue as (optically) a single one for $\mu > M_G$? Would this correspond to a next step towards the eternal dream of a fully unified description of particles and their interactions? And, if affirmative, what does it have to do with the baryon and lepton number violating operators and their phenomenology discussed at length in the preceding sections?

2.5 Grand unification of the SM interactions

In order for the unification picture sketched in the previous paragraph to make any sense one should, eventually, realise these ideas in a fully dynamical scheme. This, however, requires several ingredients:

1. A certain specific set of new fields would have to be added into the game in such a way that the two running coefficients b_2 and b_3, as different as they are in the SM, become identical at and above[35] M_G.

2. Such a change in the slopes has to occur at a very specific scale – at around $M_G \sim 10^{16}$ GeV.

3. If vector fields were to be employed for that sake one should eventually aim at a complete gauge framework with a gauge group G that would contain the entire SM one as a substructure, i.e., $G \supset SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$; in such a scheme the gluons together with the SM A and B fields would span just a part of the adjoint representation of G along with another set of vector fields that, indeed, may play the role of the extra degrees of freedom of point 1.

From the low-energy perspective, *it is like a miracle that the simplest complete picture of this kind can be obtained with just a pair of extra fields with their quantum numbers corresponding to the V_1 and S_1 mediators of the $d = 6$ BLNV effective operators in the SM!*

2.5.1 Evolution of gauge couplings in the BLNV extensions of the SM

With V_1 and S_1 at play at and above $\mu \sim M_G$ there are extra contributions to the one-loop b_i coefficients in Eqs. (2.52) driving the running gauge couplings from their intersection point onwards:

$$\Delta \begin{pmatrix} b_2 \\ b_3 \end{pmatrix}_{V_1} = \begin{pmatrix} -11 \\ -\frac{22}{3} \end{pmatrix}, \qquad \Delta \begin{pmatrix} b_2 \\ b_3 \end{pmatrix}_{S_1} = \begin{pmatrix} 0 \\ \frac{1}{6} \end{pmatrix}. \tag{2.54}$$

Remarkably, this is all that is needed for a "homogenisation" of the SM b_i's which, at and above $\mu = M_G$, change to a common value of $-85/6$. The gauge coupling evolution in this situation is depicted in Fig. 2.10.

[35] This simple phrasing is essentially consistent at the one-loop level of the RG analysis; however, one has to be way more careful if higher order corrections are to be taken into account, cf. [72,73].

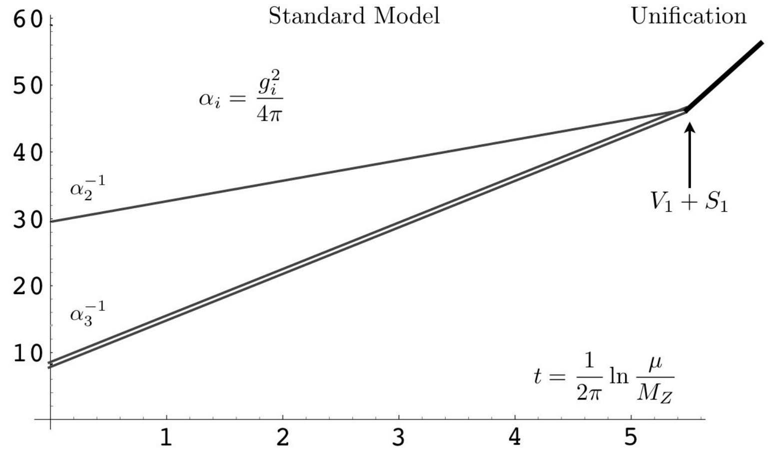


Figure 2.10: Evolution of the (inverse squared) running non-abelian gauge couplings in the Standard Model (lower t part) into which a pair of fields V_1 and S_1 from Table 2.2 is added at the scale of the $\alpha_{2,3}^{-1}$ confluence.

2.5.2 The grand unification hypothesis and its implications

Hence, a simple conspiracy of the Standard Model dynamics with just a couple of extra fields from the Table 2.2 of possible tree-level mediators of the $d = 6$ BLNV effective operators suggests a potentially renormalizable framework including a unified description of the $SU(3)_c$ and $SU(2)_L$ gauge interactions together with a new force which leads to potentially testable BLNV signals, very clear from any SM background! Hence, one encounters a *canonical incarnation of a physical theory in the best Popperian sense* – the "old" SM is fully encompassed, some of its shortcomings can be addressed (L is not a sacred perturbative-level symmetry anymore) and there is a new dynamics predicted at a very specific scale[36] with a load of spectacular new phenomena not far beyond the current experimental limits.

In some sense, the profound beauty and simplicity of such a picture may be even disturbing. Indeed, the proximity of $M_G \sim 10^{16}$ GeV and the $\sim 10^{15}$ GeV lower limit on the suppression scale associated to the $d = 6$ effective operators probed in the current proton decay searches (cf. Sect. 2.3.2) raises questions about the fragility of the whole scheme whose defining feature is the presence of the "big desert" between the electroweak scale and M_G. Is it really the case that no new dynamics is to be expected anywhere between 10^2 and 10^{16} GeV?

[36] It is difficult to overemphasise the rarity of this feature in the swampland of the SM extensions (even among the most popular ones) where the scale of the new dynamics and, thus, the size of the associated new physics effects, is typically an unknown external parameter.

To this end, it is perhaps interesting to comment briefly on the possible changes to this scheme in presence of the additional degrees of freedom associated with the renormalizable realisations of the $d = 5$ Weinberg operator in the three types of the "tree-level" seesaw mechanism discussed in Sect. 2.2:

1. As for the RH neutrinos of type-I seesaw these are complete gauge singlets and, as such, they do not affect the calculated b-coefficients above in any way, at least not at the one-loop level. From this perspective, the gauge unification pattern in Fig. 2.10 remains practically intact in this case.

2. With an extra $\Delta_L \equiv (\mathbf{1}, \mathbf{3}, +1)$ scalar at some 10^{13} GeV only small changes can be inflicted on the gauge unification pattern in type-II seesaw. However, if the triplet is to be pushed down, perhaps into the vicinity of the electroweak scale, one has to be more careful as, in such a case, the slope of the α_2^{-1} curve gets shallower for a significant portion of the running and the intersection point with α_3^{-1} is pushed well below 10^{16} GeV.

3. The type-III seesaw setting with its fermionic $SU(2)_L$ triplets at a large scale is somewhere in between the two scenarios above.

The running $U(1)_Y$ gauge coupling and the GUT-compatible hypercharge

Another very important aspect of the story is the fate of the abelian gauge coupling g' associated to the SM hypercharge. If the unification of the gauge interactions of the SM is to be complete (or "grand") then also g' should conform the g-confluence constraint at M_G as do the non-abelian couplings. On one hand, given the aforementioned arbitrariness of the SM hypercharge operator overall normalization – and thus also that of g', cf. Sect. 2.4.1– this may seem like a trivial requirement (and, in fact, it is so as long as we look at the unification from the low-energy perspective only). However, there is still something very interesting happening to the SM g' when V_1 and S_1 are integrated in at M_G. Indeed, at the level of the individual classes of contributions to the b-coefficients[37] corresponding to the three different types of terms in Eq. (2.51)

$$\begin{pmatrix} b' \\ b_2 \\ b_3 \end{pmatrix}_{\text{SM}+V_1+S_1} = -\frac{11}{3}\begin{pmatrix} \frac{25}{3} \\ 5 \\ 5 \end{pmatrix}_{\text{vectors}} + \frac{2}{3}\begin{pmatrix} \frac{10}{3} \\ 2 \\ 2 \end{pmatrix}_{\text{matter}} + \frac{1}{3}\begin{pmatrix} \frac{5}{6} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_{\text{scalars}} \tag{2.55}$$

one can see that there is a *universal scaling factor* of $\frac{5}{3}$ that, if somehow stripped from the first row above (corresponding to a redefinition of Y and g') would equalize not

[37]Note that the contributions to b' in formula (2.55) have been calculated in the usual SM hypercharge normalization fixed by Eq. (2.5).

only the resulting b's but even the individual contributions associated to the vectors, matter fields and scalars, respectively! Thus, there is a very strong indication of what the "canonical" hypercharge normalization should be in scenarios in which one may eventually wish to talk about a unified description of not only the interaction strengths but also the effective field content behind; see Sect. 2.6.2.

The final comment concerns the fate of the other SM couplings, especially those in the Yukawa sector. Another likely consequence of the effect of a "unified" description of matter and scalar fields stipulated by the observation (2.55) is a more concise account for the flavour structure within the high-energy theory which, as a by-product, may provide correlations between different effective SM Yukawa matrices and, thus, shed at least some light on the deep issue of the SM quark and lepton masses and mixing.

2.5.3 Calculability of proton lifetime in unified models

Perhaps the most important implication of the simple unification picture sketched above is the fact that it tells us a lot about the size of the inherent new physics effects (in particular the BLNV ones) which, from the SM effective theory perspective of Sect. 2.3.1, are almost entirely out of control. Indeed, with the mediator scale set by (and calculable from) the requirement of the gauge coupling unification[38] and the strong correlations among the relevant couplings governing the hard BLNV amplitudes of Fig. 2.47 one may even attempt to provide a numerical estimate of the proton lifetime. To this end, let us reiterate that the Nature has apparently been very generous to us (again!) – with the 10^{16} GeV ballpark value of M_G and for $\mathcal{O}(1)$ couplings the rough proton lifetime estimates fall just to the (logarithmic) vicinity of the current experimental limits discussed in Sect. 2.3.1!

On the practical side, however, this may be a very formidable task of a high degree of complexity. Barring the notorious difficulties associated with the "translation" of the perturbative quark-level amplitudes of Fig. (2.47) into the hadronic ones[39] there is namely the exponential sensitivity of the position of the unification scale M_G (entering in 4th power into the decay width) on most of the ingredients of the gauge running analysis including, e.g., the M_Z-scale initial conditions, the "threshold effects" associ-

[38] As we said, the unification should take place not only within the gauge couplings but also in other structures such as Yukawas etc. However, the rigidity of the gauge sector of a generic gauge theory as compared to the model-dependence of its other sectors, together with the technical simplicity of the RG evolution of the same (at least at the 1-loop level) makes it natural to determine M_G primarily from there.

[39] Note that there was a significant progress in the last decade in the lattice calculations of the B-violating hadronic matrix elements [74] which made it possible to inhibit the associated theoretical uncertainties to such a degree that they are no longer a real concern.

ated to the possible splitting of the masses of the high-scale multiplets [75] (relevant namely for higher-order calculations) and the need to account for the flavour structure of the B- and L-violating matter currents coupled to the heavy mediators. Last, but not least, there is the issue of the proximity of M_G and the Planck scale M_{Pl} [76–78] which, among other things, may inflict significant effects in the matching of the unified theory to the SM which, very often, are out of any control; see also Sect. 3.4.4. Thus, there are very few (if any) good-quality proton lifetime estimates in the existing literature.

2.5.4 Topological defects

The need to enhance the SM gauge symmetry to a higher-symmetry structure G in presence of new vector fields as stipulated in Sect. 2.5 calls, at least in the most conservative approach, for a "repetition" of the classical Higgs trick (associated, traditionally, to the phenomenon of spontaneous electroweak symmetry breaking in the SM) in the unified scenario where G must be eventually broken to $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. Besides the scale at which this should occur (M_G vs. M_Z) the main difference between the unified and the electroweak symmetry breaking is the topological structure of the corresponding coset spaces which, in the SM case, yields no stable topological defects. This, however, does not necessarily happen for the breaking of G, especially if the associated gauge group happens to be simple (i.e., when the unification is "grand"). In such a case the second homotopy class π_2 of the coset $G/SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ may be non-trivial and, hence, a topologically stable particle-like monopole solutions may exist [79–81] in the spectrum of such models. This, in fact, is even inevitable if G is also simply connected; then $\pi_2[G/SU(3)_c \otimes SU(2)_L \otimes U(1)_Y] = \pi_1[SU(3)_c \otimes SU(2)_L \otimes U(1)_Y] = \mathbb{Z}$.

Besides the topologically ensured stability (and a relatively easy production [82,83] at the "GUT-epoch" of the early Universe some 10^{-37} s after the Big Bang) such objects would have very interesting properties. These, among other things, include:

- *Very high ionisation energy loss in matter:* Indeed, the Dirac quantization condition [84,85] for the magnetic charge Q_M in the form

$$Q_E Q_M = 2\pi n \tag{2.56}$$

yields an effective "monopole fine structure constant"

$$\alpha_M \equiv \frac{Q_M^2}{4\pi} = \frac{n^2}{4\alpha} \tag{2.57}$$

which means that the monopole electromagnetic interactions with matter are characterised by a dimensionless coupling that, even for $n = 1$, is $1/4\alpha^2 \sim 4700$

times larger than the fine structure constant relevant for the QED interactions of the SM leptons or hadrons.

- *Catalysis of a rapid non-perturbative baryon number violation* (so called Callan-Rubakov effect [19, 20]): In the centre of the t'Hooft-Polyakov [79, 80] monopole solution the VEV of the Higgs field responsible for the breaking of G is effectively zero and, hence, the effects of the coset gauge fields (i.e., those carying lepto-quark charges like, e.g., V_1 of Table 2.2) are not, at least locally, inhibited by their large mass parameters. This makes the cross-section of their BLNV interactions proportional to the square of the geometrical size of this region which, for a monopole with mass $\sim M_G$, scales like M_G^{-2}; this is to be compared with the M_G^{-4} scaling of the perturbative effects discussed in Sect. 2.3.2. If a significant number of such monopoles is captured in the interior of compact astronomical objects such as white dwarfs a non-negligible increase in their luminosity can be in principle measurable. To this end, it is quite interesting that the non-observation of such effects provide orders-of-magnitude better constraints [86] on the monopole flux than direct searches based on the ionisation or magnetic effects [87].

Having already mentioned the Dirac condition (2.56) it is perhaps worth making one more conceptual comment here. Indeed, the classical Dirac's argument [84] is that if a magnetic monopole exists then the flux tube which, in the Maxwelian electrodynamics, must extend from it along a semiaxis pointing towards an antimonopole somewhere far away would be unobservable if and only if all electric charges in the entire Universe were quantized. This, in fact, is perfectly compatible with the behaviour of the irreducible representations of simple compact Lie groups as typical incarnations of the unified symmetry structure G whose generators have, necessarily, discrete spectra.

2.6 Towards a potentially realistic grand unified theory

The general concepts discussed above call for specific examples. These can be roughly classified by:

1. *The Lie group/algebra of the unified symmetry* G which, by definition, should contain that of the Standard Model (i.e., $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \equiv H$) as a substructure. At the same time, it would be nice if G was not "far larger than H", because otherwise the symmetry breaking mechanism may get rather complicated – note that, for the lowest-rank G's it can make a difference whether we demand the $H \subset G$ embedding at the level of groups or algebras - for instance, at the

group level, $SU(3) \otimes SU(2) \otimes U(1)$ is formally *not* a subgroup of $SU(5)$ but it is so as long as their algebras are concerned [40].

2. *The structure of the matter sector.* There is often no need to add anything extra on top of the SM matter fermions (like, e.g., in $SU(5)$, cf. Sect. 2.6.1) or the extensions may be very mild and well motivated (as in the $SO(10)$ models of Chapter 3).

3. *The complexity/reality of the irreducible representations of* G – the chiral structure of the SM calls for the existence of complex irreps of G - the reason is the need to maintain the sharp distinction between the LH and RH fields entertained in the SM, at least at low energies.

4. *The structure of the scalar sector* - this is by far the least constrained part of the spectrum and, hence, a defining feature of any specific unified model. Perturbativity and non-tachyonicity constraints are among the most stringent requirements imposed on its choice.

Besides these, minimality will often be our main guiding principle (because there is hardly anything better). Interestingly, as trivial as it sounds there is no universal definition of what it means - should it be the number of fields or the number of free parameters of a model? In what follows we shall entertain the second option as, indeed, the main concern of *natural philosophy* should be predictivity.

2.6.1 The minimal $SU(5)$ Georgi-Glashow model

The first work on the grand unification ever published [90] was written in 1974 by H. Georgi and S. Glashow. Though it was mainly mathematical in scope (focusing on the identification of rank=4 groups that may play the role a grand-unified gauge symmetry) it defines a paradigmatic minimal framework[41] based on an $SU(5)$ gauge group. This, back in the mid of 1970's, looked like a very interesting candidate for superseding the just-born electroweak theory and QCD.

The basic structure of the Georgi-Glashow (G-G) model is extremely simple indeed. First, there is the adjoint irrep. **24** hosting the gauge fields which, under the SM, decomposes into

$$\mathbf{24}_V = (\mathbf{8}, \mathbf{1}, 0) \oplus (\mathbf{1}, \mathbf{3}, 0) \oplus (\mathbf{1}, \mathbf{1}, 0) \oplus (\mathbf{3}, \mathbf{2}, -\tfrac{5}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, +\tfrac{5}{6}). \qquad (2.58)$$

[40] As a matter of fact, the true gauge symmetry of the SM is not $SU(3) \otimes SU(2) \otimes U(1)$ but $SU(3) \otimes SU(2) \otimes U(1)/Z_6$ [88] and only the latter does admit an embedding into $SU(5)$ even at the group level [89].

[41] As much as, for instance, the $\lambda\phi^4$ theory does within the realm of interacting QFT's.

Needless to say, the first three factors above carry the gauge quantum numbers of the SM gluons and the A and B fields, respectively. Notice also that, indeed, the last two terms in (2.58) correspond exactly to the V_1 vector ($+h.c.$) identified in Sect. 2.3.2 that was crucial later in Sect. 2.5.1 to attain the "unified" shape of the trans-M_G beta functions. In this sense, baryon and lepton number violation (BLNV) is an intrinsic feature of the G-G model.

Second, the five independent irreps hosting each family of matter fermions in the SM are embedded into just a pair of $SU(5)$ representations - a 5-dimensional vector and a 10-dimensional 2-index antisymmetric tensor:

$$\overline{\mathbf{5}}_F = (\overline{\mathbf{3}}, \mathbf{1}, +\tfrac{1}{3}) \oplus (\mathbf{1}, \overline{\mathbf{2}}, -\tfrac{1}{2}) = d_L^c \oplus L_L \,, \tag{2.59}$$

$$\mathbf{10}_F = (\overline{\mathbf{3}}, \mathbf{1}, -\tfrac{2}{3}) \oplus (\mathbf{3}, \mathbf{2}, +\tfrac{1}{6}) \oplus (\mathbf{1}, \mathbf{1}, +1) = u_L^c \oplus Q_L \oplus e_L^c \,. \tag{2.60}$$

Two comments are perhaps worth making here:

- Note that all the right-handed SM matter spinors have been conveniently written in terms of their charge-conjugated counterparts which are left-handed. First, this is a must in order to put different chirality fields into the same multiplet of a higher symmetry which is supposed to commute with the Lorentz group; second, the same transformation flips the $-\frac{1}{3}$ hypercharge of d_R into $+\frac{1}{3}$ of d_L^c which, unlike for the former, matches the $-\frac{1}{2}$ hypercharge of L_L and, thus, also the zero trace condition imposed on the $SU(5)$ generators in any representation.

- The fact that $\overline{\mathbf{5}}$ is used above for accommodating L_L and d_L^c rather than the "optically simpler" $\mathbf{5}$ is a pure convention which is set by the shape of the 24 defining generators of the natural 5-dimensional representation. This is usually chosen in such a way that the Gell-Mann matrices defining their upper-left 3×3 sub-blocks enter there without extra complex conjugation. Note that in the opposite case one would have to work with $\overline{\mathbf{10}}_F$ instead of $\mathbf{10}_F$ so a bar appears in either case.

Finally, there are 2 irreps including the two Higgs fields necessary for breaking the rank=4 $SU(5)$ gauge symmetry down to that of the $SU(3)_c \otimes U(1)_Q$ of QCD$\otimes$QED via an intermediate $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ of the SM, namely:

$$\mathbf{24}_S = (\mathbf{8}, \mathbf{1}, 0) \oplus (\mathbf{1}, \mathbf{3}, 0) \oplus (\mathbf{1}, \mathbf{1}, 0) \oplus (\mathbf{3}, \mathbf{2}, -\tfrac{5}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, +\tfrac{5}{6}) \tag{2.61}$$

and

$$\mathbf{5}_S = (\mathbf{1}, \mathbf{2}, +\tfrac{1}{2}) \oplus (\mathbf{3}, \mathbf{1}, -\tfrac{1}{3}) \,. \tag{2.62}$$

As for the former (adjoint scalar), the singlet in its SM decomposition, i.e., the 3rd term in (2.61), justifies its choice[42] as an agent to perform the first symmetry breaking step. On the same footing, the $\mathbf{5}_S$ provides the SM Higgs doublet, along with the S_1 scalar of Sects. 2.3.2 and 2.5.1. In this sense, the crucial extra fields V_1 and S_1 are supplied in the most natural way in the G-G model!

Let us also comment in brief on the possible role of the extra degrees of freedom that were not considered in Sect. 2.5.1 but which are there in the minimal $SU(5)$ theory, namely, the contents of $\mathbf{24}_S$. First, not all components of the decomposition (2.61) are propagating: indeed, the $(\mathbf{3}, \mathbf{2}, -\frac{5}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, +\frac{5}{6})$ scalars play the role of the Goldstone modes to be "eaten" in the unitary gauge by V_1 in order to give them $\mathcal{O}(M_G)$ masses so only the first three will be in the physical spectrum of the theory. Second, the effect of all these fields in the (Feynman-gauge, one loop) argument of Sect. 2.5.1 is practically irrelevant because these degrees of freedom come as components of an entire irrep of the unified $SU(5)$ group. As such, they *must* contribute homogeneously to all three beta-functions for the effective SM couplings (if the $SU(3) \otimes SU(2) \otimes U(1)$ language is preferred) and, thus, leave the scale of their intersection intact. This expectation is easy to verify by direct calculation.

The last remark concerns the Higgs mechanism triggered by a large VEV of the SM singlet in (2.61). The hypothesis that it generates masses only for the desired $V_1 = (\mathbf{3}, \mathbf{2}, -\frac{5}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, +\frac{5}{6})$ vector can be verified readily by spanning the complete set of the $SU(5)$ vector fields onto the natural basis of its adjoint representation, namely, the very generator matrices T^a:

$$|V^a\rangle = V^a T^a \text{ (no summation over } a) \, , a = 1 \ldots 24 \, . \tag{2.63}$$

Recall that the action of the generators of the adjoint representation (T^b_{adj}) on such vectors can be written in terms of the commutator of a pair of fundamental generators

$$T^b_{\text{adj}} |V^a\rangle \propto V^a \left[T^a, T^b\right] \, . \tag{2.64}$$

In this formalism, the singlet component which should bear the $SU(5)$-breaking but $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$-conserving VEV must thus be spanned over the extra Cartan operator (T^{24}) that commutes with the three SM ones. In the conventional basis in which the $SU(3)_c$ generators occupy the upper-left 3×3 corner of the corresponding 5×5 Hermitian traceless matrix space and the $SU(2)_L$ ones are accommodated in the

[42] Notice that the first step of the symmetry breaking, i.e., $SU(5) \to SU(3) \otimes SU(2) \otimes U(1)$, preserves the rank of the initial algebra. Thus, the possible choices of the irreps which can be used for that is very limited and, to this end, the adjoint is a natural candidate.

lower-right 2×2 sector T^{24} reads

$$T^{24} \propto \begin{pmatrix} \frac{1}{3}\,\mathbb{1}_{3\times 3} & \\ & -\frac{1}{2}\,\mathbb{1}_{2\times 2} \end{pmatrix}, \tag{2.65}$$

and so does the "ket" corresponding to the SM singlet: $|(\mathbf{1},\mathbf{1},0)\rangle = V_G T^{24}$. Hence, the mass matrix for the vector fields

$$M^2_{ab} = g^2 \langle 0 | T^a_{\rm adj} T^b_{\rm adj} | 0 \rangle = {\rm Tr}\, g^2 V_G^2 [T^{24}, T^a][T^b, T^{24}] \tag{2.66}$$

can receive non-zero contributions only for a and b which correspond to the fields spanned over the fundamental generators carrying both $SU(3)_c$ and $SU(2)_L$ indices, i.e., $(\mathbf{3},\mathbf{2},-\frac{5}{6}) \oplus (\overline{\mathbf{3}},\mathbf{2},+\frac{5}{6})$.

It can also be shown that the vacuum manifold defined by the scalar potential[43]

$$V = m^2 {\rm Tr}\, \mathbf{24}_S^2 + \lambda ({\rm Tr}\, \mathbf{24}_S^2)^2 + \rho {\rm Tr}\, \mathbf{24}_S^4 \tag{2.67}$$

hosts only three extrema out of which only the original $SU(5)$ and the final $SU(3) \otimes SU(2) \otimes U(1)$-symmetric one are (local) minima.

2.6.2 Phenomenology of the minimal $SU(5)$ GUT

Besides the expected new physics in the form of baryon and lepton number violation the unified theories often provide insights into the structure of their low-energy effective descendants in the form of correlations among independent parameters of the SM. To this end, there are two key observations to be made in the G-G context:

1. *The value of the SM weak mixing angle is predicted.* This has to do with the structure of the hypercharge operators in the $SU(5)$ irreps hosting (not only) the SM matter. In order to maintain simplicity all generators of simple Lie groups are conventionally orthonormalized by conditions like

$$ {\rm Tr}\, T^i_R T^j_R = T_2(R) \delta^{ij} \tag{2.68}$$

where $T_2(R)$ is a universal real number (called index) specific to the irrep R. The normalization is usually fixed so that T_2 of the vector irreps of $SU(N)$'s is $\frac{1}{2}$. Looking at $\overline{\mathbf{5}}_{\mathbf{F}}$ in (2.59) it is easy to verify that this choice indeed coincides with the usual normalization of the SM $SU(3)_c$ (anti)triplet generators (halves of the Gell-Mann matrices, $\lambda^a/2$) as well as with those of the $SU(2)_L$ doublets (halves

[43] We do not consider the $\mathbf{5}_S$ component here as its VEV necessarily reduces the rank of the gauge group and, as such, it can only be responsible for the subsequent electroweak symmetry breaking with $v \ll V_G$.

of the Pauli matrices, $\sigma^k/2$). On the other hand, the normalization of the SM hypercharge Y_{SM} as of Eq. (2.5) must be adjusted into the GUT-natural form $Y_G = nY_{SM}$ so that

$$\sum_{\overline{\mathbf{5}}_{\mathbf{F}}} Y_G^2 = n^2 \sum_{\overline{\mathbf{5}}_{\mathbf{F}}} Y_{SM}^2 = n^2 \left(3 \times \frac{1}{9} + 2 \times \frac{1}{4}\right) = \frac{5}{6} n^2 = \frac{1}{2}\,, \tag{2.69}$$

which yields

$$n = \pm\sqrt{\frac{3}{5}}\,. \tag{2.70}$$

Choosing, conventionally, the positive solution the GUT-adjusted hypercharge reads

$$Y_G = \sqrt{\frac{3}{5}} Y_{SM} \tag{2.71}$$

and the proper decomposition of $\overline{\mathbf{5}}_F$ is thus

$$\overline{\mathbf{5}}_F = (\overline{\mathbf{3}}, \mathbf{1}, +\sqrt{\tfrac{1}{15}}) \oplus (\mathbf{1}, \overline{\mathbf{2}}, -\sqrt{\tfrac{3}{20}})\,, \tag{2.72}$$

and similarly for the other irreps. The change $Y_{SM} \to Y_G$ in the structure of the covariant derivative of the theory inflicts a change in the associated gauge coupling too: On the $SU(5)$ side one has $D_\mu^{SU(5)} = \partial_\mu - ig_5 Y_G A_\mu^Y + \ldots$ which is to be matched to the SM structure $D_\mu^{SM} = \partial_\mu - ig' Y_{SM} B_\mu + \ldots$; this can be done if and only if

$$g_5 \sqrt{\tfrac{3}{5}} \equiv g' \tag{2.73}$$

at the matching scale, i.e., at M_G. Hence, there are two different languages in which one can describe the same hypercharge dynamics, either in the language of the SM with g' and Y_{SM} everywhere or with g_5 and Y_G. It is only in the latter case though that the associated gauge coupling (i.e., g_5) should be universal at M_G, not with the former g'. Thus, in the beta-functions' analysis of Sect. 2.5.2 one should have worked with $g_1 \equiv \sqrt{\frac{5}{3}} g'$ and the associated b_1 from scratch instead of b' calculated from Y_{SM}! This, indeed, supplies the extra multiplicative factor of $\frac{3}{5}$ to the first row of Eq. (2.55) and brings it into a fully universal form

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}_{\mathrm{SM}+V_1+S_1} = -\frac{11}{3} \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}_{\text{vectors}} + \frac{2}{3} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}_{\text{matter}} + \frac{1}{3} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_{\text{scalars}} . \tag{2.74}$$

Hence, the unified nature of the 3 gauge interactions in presence of V_1 and S_1 (barring the universal extra contribution from $\mathbf{24}_S$ in the full G-G setting) becomes obvious (even tautologic) in this language!

Finally, the gauge unification condition $g_1 = g_2 = g_3$ at M_G, rewritten for g and g' of the SM as $g = \sqrt{\frac{5}{3}}g'$ at M_G *makes it possible to calculate the (tree-level) SM weak mixing angle at* M_G:

$$\sin^2\theta_W(M_G) \equiv \frac{g'^2}{g^2+g'^2} = \frac{1}{\frac{5}{3}+1} = \frac{3}{8} = 0.375\,. \tag{2.75}$$

This, at one hand, looks way bigger than the measured value $\sin^2\theta_W(M_Z) \sim 0.232$ but, at the same time, it can not be discarded right away because of the large effects of running from M_G to M_Z; indeed, $b_2 = -\frac{19}{6}$ and $b' = +\frac{41}{6}$ below M_G and, thus, the RHS of (2.75) will get reduced towards M_Z. In any case, the final decision requires a more detailed analysis.

2. *Some of the SM Yukawa couplings are strongly correlated.* This is another example of a feature that provides an interesting insight into the (conceptually) free parameters of the SM, namely, its flavour structure. It has to do with the shape of the (renormalizable) Yukawa Lagrangian of the minimal $SU(5)$ which, in the most economical case of just one "Yukawa-active" scalar representation[44], can be written like

$$\mathcal{L}_Y = Y_{\overline{5}}^{ij}\overline{\mathbf{5}}_F{}^{iT}C\mathbf{10}_F{}^{j}\mathbf{5}_S^* + Y_{10}^{ij}\mathbf{10}_F{}^{iT}C\mathbf{10}_F{}^{j}\mathbf{5}_S\,, \tag{2.76}$$

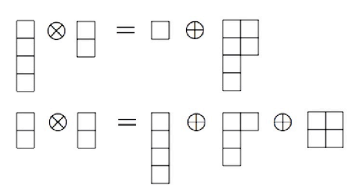
where i and j are family indices, T denotes transposition in the spinorial space and C is the associated charge conjugation matrix. In the asymmetric phase, i.e., with non-zero VEV v of the $SU(2)_L$ doublet in $\mathbf{5}_S$ one can readily identify the mass matrices of the SM fermions (at M_G):

$$M_d^T = M_l = Y_{\overline{5}}\frac{v}{\sqrt{2}}, \qquad M_u = M_u^T = Y_{10}\frac{v}{\sqrt{2}}\,. \tag{2.77}$$

Note that there is a strong correlation among the down-quark and charged-lepton sectors in the minimal $SU(5)$ GUT and that the up-quark Yukawa coupling is symmetric. This, in turn, reduces the number of independent SM Yukawa couplings considerably.

[44] It is possible to work with only a single scalar $\mathbf{5}$ because it (or its complex conjugate) can be found in both relevant tensor products

$$\begin{aligned}\overline{\mathbf{5}}\otimes\mathbf{10} &= \mathbf{5}\oplus\overline{\mathbf{45}}\\ \mathbf{10}\otimes\mathbf{10} &= \overline{\mathbf{5}}\oplus\mathbf{45}\oplus\mathbf{50}\end{aligned} \quad\leftrightarrow$$



.

2.6.3 The failure of the minimal $SU(5)$ GUT

As beautiful and natural as it looks this picture actually does not withstand a more thorough scrutiny. First, the aforementioned flavour correlations lead to a direct prediction for the (running) masses of down-quarks and charged-lepton which should – generation by generation – coincide at M_G. While this works at least to some extent for m_b and m_τ it fails miserably for the first two families.

As a matter of fact, this per se does not need to be fatal for the G-G model because the relevant Lagrangian (2.76) may be subject to large corrections associated to the $d > 4$ Planck-scale induced operators such as those discussed in [91]. Indeed, since M_G/M_{Pl} ratio is around 1% one can expect significant corrections to the two smallest eigenvalues of both M_d and M_l which may be sufficient to correct the "unrealistic" first- and second-generation $d = 4$ predictions.

The G-G scenario, however, suffers from way more severe drawbacks than that:

1. It can not (in its minimal form) account for non-zero neutrino masses. The point is the absence of the necessary structure to incarnate the Dirac option or any of the three basic seesaw variants of the neutrino mass generation of Sect. 2.3.2. The obvious solution, i.e., the addition of RH neutrinos in the form of $SU(5)$ singlets is not very satisfactory as it brings in a new Yukawa matrix which does not entertain correlations with any other. Perhaps even more importantly, the natural scale of the associated gauge-singlet Majorana mass term, given its $SU(5)$-singlet nature, exceeds M_G which is difficult to reconcile with the neutrino oscillations data.

2. The weak mixing angle (evolved from M_G to M_Z) turns out wrong. Performing the simple exercise discussed below Eq. (2.75) one recovers $\sin^2\theta_W(M_Z) \sim 0.195$ which is many (tens of) standard deviations off the measured value. In the early days of the subject [90] this has not been an issue due to the lack of good data; however, the situtation has changed drastically with the advent of LEP and its precision weak sector measurements.

Hence, the minimal $SU(5)$ grand unified model by Georgi and Glashow is nowadays considered dead and, as such, it enjoys the status of a benchmark scenario rather than a full-fledged physical theory.

2.6.4 The minimal supersymmetric $SU(5)$ GUT

There are, of course, many proposals aiming at overcoming the limitations of the original G-G model such as, for instance, settings with extra matter fermions [92] or extra scalars [93]. These, if pushed deep enough into the "GUT desert" may change the gauge running pattern considerably.

However, by far the most popular of these is the idea of TeV-scale supersymmetry (SUSY) which, with the onset of the precision electroweak era in the early 1980's, even became a BSM model-building paradigm. In this framework, the main drawbacks of the Standard Model discussed in Chapter 1 and of the minimal $SU(5)$ GUT of Sect. 2.6.3 are addressed as follows:

- *The wrong value of* $\sin^2\theta_W$ *in the Georgi-Glashow model:* The "doubling of degrees of freedom" (together with the need to add another Higgs doublet to tame the issues with holomorphy of the superpotential as well as with potential gauge anomalies) in the minimal supersymmetric extension of the SM (MSSM) bends the RG running of the SM gauge couplings so that the experimental value of $\sin^2\theta_W(M_Z) \sim 0.232$ is "miraculously" attained.

- *The absence of a suitable dark matter candidate in the SM.* It is well known that there is no room in the Standard Model to account for the cold dark matter (CDM) particle candidate favoured by the current ΛCDM Standard Model of cosmology. The light neutrinos, the only stable enough neutral particles around, could account for the observed DM critical density fraction if and only if the sum of their masses was at the level of about 10 eV, at odds with the oscillation and β-decay data. Moreover, their "hot relic" nature leads to contradictions with the small-scale structure formation data if they were to represent the dominant DM component. To this end, the low-energy SUSY can come to rescue if R-parity, one of the empirical extra symmetry requirements often imposed on its superpotential, is exact – in such a case the neutralino (if it happens to be the lightest supersymmetric particle, LSP) provides an almost ideal particle-like CDM candidate.

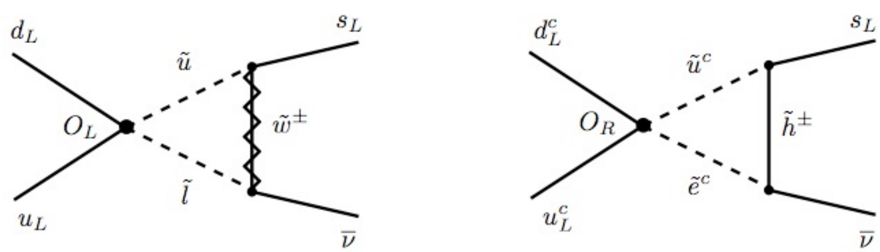
- *The apparent need for order-by-order fine-tunning among the bare and the high-scale (cut-off) contributions in the formula for the electroweak VEV.* This issue is often quoted as one of the facets of the notorious "hierarchy problem" which was stirring the HEP community[45] for at least the last 30 years. In this respect, the low-energy SUSY brings a partial relief to the issue by the radiative stabilisation of any mechanism that is eventually employed at the tree level.

In this respect, the minimal SUSY $SU(5)$ GUT [96, 97] represents the most economical grandunified completion of the MSSM. On the technical side, barring the standard elements of the SUSY model building like the need to embed all the standard $SU(5)$

[45] Not the author though whose renegade attitude to this conundrum is well documented by his recent works [94, 95].

quantum fields into superfields (chiral or vector), the presence of the soft-SUSY breaking sector etc., the basic structure of the minimal SUSY $SU(5)$ GUT closely resembles that of the Georgi-Glashow scheme of Sect. 2.6.1. The main difference is the presence of two different copies of 5-dimensional Higgs multiplets (as with the doublets in the MSSM) and, in particular, the simplicity of the $d = 3$ renormalizable superpotential which relates the higgsino couplings with matter-smatter currents to the ordinary Yukawa terms. These, in turn, play an important role in the description of the BLNV phenomena like proton decay which, unlike in the non-SUSY theories, can proceed through triplet-higgsino-mediated effective operators already at the $d = 5$ level (in the effective MSSM language), see Fig. 2.11.

Figure 2.11: The basic structure of the hard-process Feynman graphs behind proton decay in TeV-scale SUSY models. The diagrams with gluino, bino and neutral higgsino dressings are typically suppressed by the first and second generation Yukawa couplings governing the underlying $d = 5$ effective operators. The flavour structure displayed corresponds to the dominant p-decay channel which, in SUSY, is usually $p^+ \to K^+\overline{\nu}$.



2.6.5 The issues of the minimal supersymmetric $SU(5)$ GUT

This, however, indicates a potential problem with the minimal supersymmetric $SU(5)$ GUT, namely, the overly fast proton decay. Indeed, comparing the structure of the amplitudes of Fig. 2.11 with those of the non-supersymmetric variant depicted in Fig. 2.47 (especially the leading vector-mediated graph therein) the two are related by a multiplicative factor of roughly

$$R \equiv \mathcal{A}_{\rm SUSY}/\mathcal{A}_{\rm non-SUSY} \sim \frac{1}{16\pi^2}\frac{Y^2}{g^2}\frac{M_G}{m_{\rm SUSY}}\,, \tag{2.78}$$

where the first piece on the RHS correspond to the SUSY loop suppression, the second to the Yukawa vs. gauge domination of the leading-order contributions in SUSY and non-SUSY scenarios and the third to the bosonic vs. fermionic nature of the relevant high-scale mediator assisted on the SUSY side by the gaugino/higgsino propagator effect of the order of the soft-SUSY breaking scale m_{SUSY}.

Given the enormous hierarchy $M_G/m_{\mathrm{SUSY}} \sim 10^{13}$ attained within TeV-scale SUSY scenarios it looks like there is no way whatsoever to get the proton lifetime anywhere near the desired (non-supersymmetric) limit of 10^{34} years corresponding to $R \sim 10^0$. This conclusion is, however, premature as the other two contributions in (2.78) provide further suppression ($10^{-2.5}$ for the loop factor and some 10^{-8} due to the presence of the first generation Yukawa couplings). Hence, the situation of the the minimal SUSY $SU(5)$ scenario is rather severe but there is, technically, still not a strict no-go; for more details the reader is kindly deferred to [98] and references therein. Note also that the minimal SUSY $SU(5)$ fully shares the drawbacks of its non-SUSY version with accounting for the neutrino masses, see Sect. 2.6.3.

Hence, none of the two canonical versions of the minimal grand unification seems fully realistic and there is a good reason to look for better realisations of the gauge unification paradigm outside the realm of the very restrictive rank = 4 models.

Chapter 3

Radiative effects in potentially realistic unified gauge models

3.1 Rank=5 extended gauge models

Relaxing the rank = 4 requirement imposed in the previous Chapter on the defining symmetry of the minimal gauge extensions of the Standard Model the number of model-building options grows rapidly.

Perhaps the first thing that comes to mind when an extra gauge generator is admitted into the game along with the 4 Cartans of the SM is the $U(1)_{B-L}$ symmetry discussed briefly in Sect. 2.1.2 which, in the presence of 3 copies of RH neutrinos (each equipped with a unit of lepton number) becomes a good candidate for a gauge charge. There is even more to that though:

- The spectrum of such a RH-neutrino-extended variant of the SM (3NSM) becomes very symmetric as the RH neutrinos nicely "fill" the apparent vacancy in the sector of the RH leptons.

- The hypercharges of the RH sector of the 3NSM exhibit an intriguing pattern in which the two members of the "natural pairs", i.e., u_R and d_R or N_R and e_R differ by exactly the same amount ($\Delta Y_{u_R,d_R} = \Delta Y_{N_R,e_R} = +1$).

- This closely resembles the situation of the electric charges Q in the LH sector whose differences within the $SU(2)_L$ multiplets obeys the same $\Delta Q = +1$ rule formalised in the concept of isospin and the Gell-Mann-Nishijima formula (2.5).

- In attempt to replay the same game with Y instead of Q the RH variant of the weak isospin is an obvious hypothesis to put forward (with the corresponding gauge factor denoted by $SU(2)_R$ and the 3 associated generators by T^i_R). This, however, calls for an extra $U(1)$ charge X such that

$$Y = T^3_R + X\,. \tag{3.1}$$

Remarkably enough, such an additional X charge turns out to be nothing but $\frac{1}{2}(B-L)$; thus, one is left with an intriguingly symmetric relation

$$Q = T^3_L + T^3_R + \frac{B-L}{2}\,, \tag{3.2}$$

which not only connects the apparent doublet structure of the RH sector to the profound concept of the $B-L$ symmetry but it is also very aesthetically appealing. Hence, it is more than natural to take this observation as a starting point at the quest for the rank=5 gauge extensions of the SM and consider the

$$G_{LR} \equiv SU(3)_c \otimes SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L} \tag{3.3}$$

group as the first of its milestones.

3.1.1 Left-right symmetric models

Among the most appealing features of the gauge models based on the $SU(3)_c \otimes SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$ symmetry is the phenomenon of the high-scale parity restoration [58]. Indeed, in the unbroken phase the LH and RH fermions (including N_R) enjoy the same attention of their respective $SU(2)$ gauge factors and also their additional $U(1)$ charges proportional to $B-L$ are the same, see Table 3.1; the chiral structure of the SM is then revealed in the asymmetric phase only, i.e., after the symmetry breaking of

$$SU(3)_c \otimes SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L} \to SU(3)_c \otimes SU(2)_L \otimes U(1)_Y\,. \tag{3.4}$$

This is usually triggered either by a scalar field which is an $SU(2)_R$ doublet with $B-L = \pm 1$ (thus following the familiar SM pattern) or by an $SU(2)_R$ triplet $(\mathbf{1},\mathbf{1},\mathbf{3},\pm 2)$. From the phenomenology point of view the latter option is particularly interesting:

1. *In the models with LR-symmetric scalar sector the type-II seesaw mechanism for neutrino mass generation is naturally in operation.* The point is that the $SU(2)_R \otimes U(1)_{B-L}$ breaking provided by the VEV of $\Delta_R = (\mathbf{1},\mathbf{1},\mathbf{3},+2)$ gets projected onto its LH companion $\Delta_L = (\mathbf{1},\mathbf{3},\mathbf{1},+2)$ of type-II seesaw (cf. Sect. 2.2.2) through the pair of the electroweak VEVs carried by the Higgs bi-doublet $H =$

Field/Chirality	3221 quantum numbers	3RHNSM contents	$Y = T_R^3 + (B-L)/2$
Q_L	$(\mathbf{3}, \mathbf{2}, \mathbf{1}, +\frac{1}{3})$	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$+\frac{1}{6}$
Q_R	$(\mathbf{3}, \mathbf{1}, \mathbf{2}, +\frac{1}{3})$	$u_R,\, d_R$	$+\frac{2}{3},\, -\frac{1}{3}$
L_L	$(\mathbf{1}, \mathbf{2}, \mathbf{1}, -1)$	$\begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	$-\frac{1}{2}$
L_R	$(\mathbf{1}, \mathbf{1}, \mathbf{2}, -1)$	$N_R,\, e_R$	$0,\, -1$

Table 3.1: Minimal fermionic contents of the left-right symmetric extensions of the SM (single SM matter generation + RH neutrino displayed).

$(\mathbf{1}, \mathbf{2}, \mathbf{2}, 0)$ by their $\Delta_L^\dagger H^2 \Delta_R$ interaction in the scalar potential. Schematically, the vacuum condition

$$\left\langle \frac{\partial V}{\partial \Delta_L^\dagger} \right\rangle = M_\Delta^2 \langle \Delta_L \rangle + \lambda \langle H \rangle^2 \langle \Delta_R \rangle + \ldots = 0 \tag{3.5}$$

implies that Δ_L receives an "induced" VEV of the order of $\langle H \rangle^2 \langle \Delta_R \rangle / M_\Delta^2 \propto \langle H \rangle^2 / \langle \Delta_R \rangle$. Hence, the seesaw suppression of also the type-II light neutrino mass contribution is connected to the scale of the $SU(2)_R \otimes U(1)_{B-L}$ symmetry breaking as in the type I seesaw.

2. *In the supersymmetric versions of the LR models the R-parity is automatically conserved in the triplet-breaking scenarios.* The point is that the even $B - L$ charge of Δ_R makes the vacuum neutral with respect to the R-parity defined as

$$R = (-1)^{2S+3(B-L)} \tag{3.6}$$

(with $S = 0$ for spinless fields) and, hence, R remains a conserved quantum number even in the broken phase. Consequently, the neutralino LSP is a natural WIMP dark matter candidate.

Besides all this, the SM fermions entertain a high degree of correlations among their effective Yukawa couplings[1] which may, e.g., provide a rationale for the smallness of the CKM mixing and so on.

3.1.2 The Pati-Salam model

Remarkably enough, the $SU(3)_c$ and $U(1)_{B-L}$ factors of the LR gauge group of Section 3.1.1 can be further combined into a full-fledged $SU(4)$ symmetry with a Cartan

[1] This, in the simplest situation of the single bi-doublet $(\mathbf{1}, \mathbf{2}, \mathbf{2}, 0)$ scenario, is even pathological, one typically needs extra multiplets.

set including the two diagonal Gell-Mann matrices (extended by zeros to a 4×4 structure) and the 4×4 traceless $B-L$ generator with $\{+\frac{1}{3},+\frac{1}{3},+\frac{1}{3},-1\}$ on the diagonal[2]. These, together with the associated rising and lowering operators of the $SU(4)$ act naturally on a pair of 4-dimensional vector representations whose upper three components host the three colour eigenstates of quarks with the corresponding lepton on the lowest position. Thus, the gauge symmetry is further enhanced into what is usually called the Pati-Salam gauge group

$$G_{PS}\equiv SU(4)_C\otimes SU(2)_L\otimes SU(2)_R \tag{3.7}$$

and a very appealing quark-lepton unification is achieved, with "lepton number as a fourth colour" as its trademark [99].

Besides encompassing virtually all the interesting features of the aforementioned LR models (corresponding to one of its symmetry breaking chains, see below) there are other aspects of the Pati-Salam scheme worth mentioning here:

1. *There are leptoquark-type of fields in the spectrum.* This is a consequence of the extended gauge symmetry whose $(\mathbf{15},\mathbf{1},\mathbf{1})\oplus(\mathbf{1},\mathbf{3},\mathbf{1})\oplus(\mathbf{1},\mathbf{1},\mathbf{3})$ adjoint representation (especially, its first factor) decomposes under the $SU(3)_c\otimes SU(2)_L\otimes U(1)_Y$ subgroup as

$$(\mathbf{8},\mathbf{1},0)\oplus(\mathbf{1},\mathbf{1},0)\oplus(\overline{\mathbf{3}},\mathbf{1},-\tfrac{2}{3})\oplus(\mathbf{3},\mathbf{1},+\tfrac{2}{3})\oplus\ldots\ . \tag{3.8}$$

Indeed, the last two components above (to be called $X_\mu+h.c.$) are perfectly suited to provide a link between the same-chirality quarks and leptons:

$$\mathcal{L}_{PS}\ni\frac{g_4}{\sqrt{2}}\left(\overline{L_L}\gamma^\mu Q_L^i+\overline{\nu_R}\gamma^\mu u_R^i+\overline{e_R}\gamma^\mu d_R^i\right)X_{\mu\, i}+h.c. \tag{3.9}$$

Clearly, the number of SM baryons and leptons is not conserved at the level of elementary vertices in the Pati-Salam types of models. However, this does not mean that there should be any B and L violating effects observed at colliders – indeed, the B and L "leaking" from one side of a Feynman diagram containing a vertex from (3.9) is repaid on the other side of the graph where X gets transformed back to the relevant matter fermion pair. In other words, B and L are not really broken; they are just being carried around by X.

It may also be worth noting that the Pati-Salam scenario discussed here is actually not the most minimal framework where the X vectors can emerge – the $SU(4)_C\otimes SU(2)_L\otimes U(1)_R$ gauge theory, very popular recently [100], also contains these fields.

[2]Strictly speaking, the diagonal of the $B-L$ should read $\sqrt{\frac{3}{8}}\{+\frac{1}{3},+\frac{1}{3},+\frac{1}{3},-1\}$ otherwise it does not conform the same normalization as the other operators of the $SU(4)_C$ algebra.

2. *The Pati-Salam theory has a monopole.* To see this one should study the second homotopy class of the G_{PS}/SM quotient group

$$\pi_2\left(\frac{SU(4)\otimes SU(2)\otimes SU(2)}{SU(3)\otimes SU(2)\otimes U(1)}\right) = \pi_2\left(SU(4)\otimes SU(2)/SU(3)\otimes U(1)\right) \,, \tag{3.10}$$

which, thanks to the fact that

$$\pi_1(SU(4)\otimes SU(2)) = \pi_1(SU(4))\oplus\pi_1(SU(2)) = 0\,, \tag{3.11}$$

i.e., that $SU(4)\otimes SU(2)$ is simply connected, yields

$$\pi_2\left(SU(4)\otimes SU(2)/SU(3)\otimes U(1)\right) = \pi_1\left(SU(3)\otimes U(1)\right) = \mathbb{Z}\,. \tag{3.12}$$

The mass of this monopole shoud be somewhat above the scale of the Pati-Salam symmetry breaking. For more information including, e.g., cosmological constraints on its early-Universe production see [101] and references therein.

3. *The Yukawa pattern naturally accommodates the Georgi-Jarlskog texture* [102]. This concerns an interesting coincidence between the phenomenological low-energy relation $m_\mu/m_s \sim 3$ and one of the features of the electroweak symmetry-breaking pattern available in the Pati-Salam type of models. Suppose that the Higgs field of the SM contains a significant component from the $(\mathbf{15},\mathbf{2},\mathbf{2})$ Pati-Salam scalar. Since the VEV of this multiplet must be colour neutral, in the $SU(4)_C$ adjoint representation space it must be proportional to the last Cartan operator (i.e., on the $U(1)_{B-L}$ generator in this case; see also formula (2.65) of Sect. 2.6.1) which, in the traditional basis with $SU(3)_c$ spanning over the upper-left 3×3 sector of the first 8 $SU(4)_C$ generators, receives the form

$$T^{15} \propto \begin{pmatrix} \frac{1}{3}\,\mathbb{1}_{3\times 3} & \\ & -\mathbb{1} \end{pmatrix}. \tag{3.13}$$

Hence, the fraction of the electroweak VEV carried by this multiplet will naturally generate 3 times larger contributions to the effective mass matrices of leptons than to the quark ones.

3.1.3 SO(10) grand unification

With all this at hand, there is just a small final step to be made in order to identify the simplest grandunified scenario featuring the LR symmetry (3.3) as its potential low-energy descendant. Indeed, given the algebraic isomorphisms

$$su(4)\approx so(6) \qquad \text{and} \qquad su(2)\oplus su(2)\approx so(4) \tag{3.14}$$

it is clear that the entire algebra of the Pati-Salam symmetry is contained in the algebra of the simple and compact $SO(10)$ group. Thus, $SO(10)$ is often claimed to be the most natural unified rank $= 5$ gauge extension of the SM and, as a matter of fact, it can also be shown to be the only potentially realistic candidate for grand unification[3] at this level. Hence, in what follows, we shall focus primarily on the class of $SO(10)$ unified scenarios and, in particular, on their minimal SUSY and non-SUSY realisations.

One more comment concerning the anomalies in gauge extensions of the SM is perhaps worth here. Unlike in theories based on the $SU(N)$ gauge structure (or direct products of several such factors) in which the requirement of the absence of gauge anomalies provides strong constraints on their matter content (cf. Sect 2.1.2), the $SO(N)$ gauge structure is automatically anomaly free for all $N > 2$ and $N \neq 6$. This follows from the generic impossibility to construct a fully antisymmetric tensor from a product of three antisymmetric $SO(N)$ generators. The case of $N = 6$ is singular because $SO(6)$ is locally isomorphic to $SU(4)$; on a more technical level, the antisymmetry of the orthogonal groups' generators makes it possible to construct a non-vanishing anomaly structure like

$$\mathrm{Tr}\left(\{M^{ij}, M^{kl}\}M^{mn}\right) \propto \varepsilon^{ijklmn}\,, \tag{3.15}$$

with a completely antisymmetric tensor of the $SO(6)$ on the right hand side. This observation makes it also clear why the Georgi-Glashow model with matter in $\mathbf{10} \oplus \overline{\mathbf{5}}$ is anomaly free: adding a full (harmless) $SU(5)$ singlet to each generation of matter the field content of a full 16-dimensional spinor of $SO(10)$ is attained.

3.2 Minimal $SO(10)$ GUTs

Perhaps the most appealing feature of the $SO(10)$ GUTs is the fact that an entire generation of the SM matter fields can be accommodated in its single spinorial irrep, namely,

$$\mathbf{16} = \overset{Q_L}{(\mathbf{3}, \mathbf{2}, +1/6)} \oplus \overset{L_L}{(\mathbf{1}, \mathbf{2}, -1/2)} \oplus \overset{u^c_L}{(\overline{\mathbf{3}}, \mathbf{1}, -2/3)} \oplus \overset{d^c_L}{(\overline{\mathbf{3}}, \mathbf{1}, +1/3)} \oplus \overset{\nu^c_L}{(\mathbf{1}, \mathbf{1}, 0)} \oplus \overset{e^c_L}{(\mathbf{1}, \mathbf{1}, +1)} \tag{3.16}$$

and, at the same time, one RH neturino per generation is inevitable. This makes the implementation of the type-I seesaw mechanism very neatural, the more that the RH neutrino Majorana mass is forbidden in the unbroken phase (there is no $SO(10)$ singlet

[3] For a very interesting alternative rank=5 unified theory of baryon and lepton number violation which, however, is not "grand", the reader is deferred to Sect. 3.6.

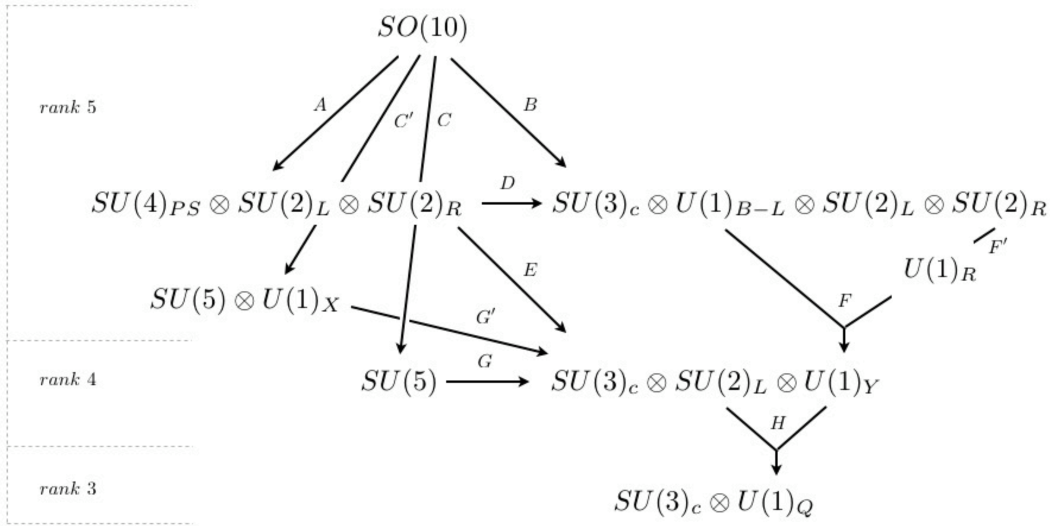


Figure 3.1: The most common breaking chains of $SO(10)$ down to the $SU(3)_c \otimes U(1)_Q$ gauge symmetry at low energies. The associated scalar irreps capable of triggering the indicated transitions are indicated along each of the lines by the symbols of Table 3.2.

in $\mathbf{16} \otimes \mathbf{16}$) and, as such, it naturally emerges below the GUT scale as a consequence of the spontaneous $SO(10)$ breaking[4].

This, however, can be triggered by many different scalar fields and may proceed through various intermediate symmetry stages, see Table 3.2 and Fig. 3.1. Barring the very exotic sequence triggered by **144**, cf. [103] at least two different scalar irreps must be employed in order to get from $SO(10)$ down to the SM; the minimal options are $\mathbf{45} \oplus \mathbf{16}$ (or $\mathbf{45} \oplus \mathbf{126}$) capable of passing through the $B-F$ chain or $\mathbf{210} \oplus \mathbf{16}$ (or $\mathbf{210} \oplus \mathbf{126}$) passing through the $B-F$, $A-D-F$ or $A-E$ sequences. Naturally, these also identify the basic model building strategies found in the literature.

The last common ingredient of all $SO(10)$ unified models is the set of gauge fields which is hosted by the 45-dimensional adjoint representation decomposing as

$$\mathbf{45} = \overset{G_\mu}{(\mathbf{8},\mathbf{1},0)} \oplus \overset{A_\mu}{(\mathbf{1},\mathbf{3},0)} \oplus \overset{B_\mu}{(\mathbf{1},\mathbf{1},0)} \oplus \overset{X_\mu}{(\overline{\mathbf{3}},\mathbf{2},+\tfrac{5}{6})} \oplus \overset{\overline{X}_\mu}{(\mathbf{3},\mathbf{2},-\tfrac{5}{6})} \oplus \overset{Y_\mu}{(\overline{\mathbf{3}},\mathbf{2},-\tfrac{1}{6})} \oplus \overset{\overline{Y}_\mu}{(\mathbf{3},\mathbf{2},+\tfrac{1}{6})} \tag{3.17}$$

under the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ subgroup.

[4]Note that this is not the case of the simple extensions of the Georgi-Glashow $SU(5)$ GUTs in which the RH neutrinos enter as full gauge singlets.

$SO(10)$ irrep	PS sub-multiplet	symmetry breaking steps
10	$(\mathbf{1},\mathbf{2},\mathbf{2})_{10}$	H
16	$(\mathbf{4},\mathbf{2},\mathbf{1})_{16}$	H
	$(\overline{\mathbf{4}},\mathbf{1},\mathbf{2})_{16}$	$C', G'; E, F$
45	$(\mathbf{1},\mathbf{3},\mathbf{1})_{45}$	$C, G'; F'$
	$(\mathbf{15},\mathbf{1},\mathbf{1})_{45}$	$B, C; D$
54	$(\mathbf{1},\mathbf{1},\mathbf{1})_{54}$	A, G
120	$(\mathbf{1},\mathbf{2},\mathbf{2})_{120}$	$G'; H$
	$(\mathbf{15},\mathbf{2},\mathbf{2})_{120}$	$G'; H$
126	$(\mathbf{15},\mathbf{2},\mathbf{2})_{126}$	H
	$(\mathbf{10},\mathbf{1},\mathbf{3})_{126}$	$E, F; G'$
144	$(\mathbf{4},\mathbf{2},\mathbf{1})_{144}$	H
	$(\overline{\mathbf{4}},\mathbf{1},\mathbf{2})_{144}$	$C', G, G'; E, F$
210	$(\mathbf{1},\mathbf{1},\mathbf{1})_{210}$	$A; C', G, G'$
	$(\mathbf{15},\mathbf{1},\mathbf{1})_{210}$	$B, D; C', G, G'$
	$(\mathbf{10},\mathbf{2},\mathbf{2})_{210}$	H
	$(\overline{\mathbf{10}},\mathbf{2},\mathbf{2})_{210}$	H

Table 3.2: The "symmetry breaking power" of various scalar $SO(10)$ irreps up to dimension 210. The quantum numbers of the submultiplets correspond to the Pati-Salam subgroup $SU(4)_C \otimes SU(2)_L \otimes SU(2)_R$. The letters denoting the different breaking steps are those used in Fig. 3.1.

3.2.1 Proton decay in $SO(10)$ GUTs

Note that the first five components of decomposition (3.17) correspond to the gauge fields of the Georgi-Glashow model and there is just one new vector field $Y \oplus \overline{Y}$ present in the $SO(10)$ case. In this respect, the gauge-driven baryon number violation phenomenology of the $SO(10)$ GUTs naturally encompasses all that has been said about it in the G-G context. The only extra effect due to the presence of $Y \oplus \overline{Y}$ consists in their capacity to provide an extra contribution to the amplitudes corresponding to the $\mathcal{O}_1$ effective $d = 6$ operator of Table 2.1 (see also Table 2.2), and, hence, loosen the correlations between the $\mathcal{O}_1$- and $\mathcal{O}_2$-driven effects characteristic to the $SU(5)$ settings. The presence of a second B and L "active" vector multiplet also brings in a second suppression scale which, in principle, can be quite different from the mass of X, especially along the symmetry breaking directions featuring an intermediate $SU(5)$ or flipped $SU(5)$ stage, cf. Sect. 3.6.2.

Concerning the minimal SUSY variants of the $SO(10)$ GUTs one would naïvely expect that they must suffer from the same drawbacks related to the insufficient suppression of the $d = 5$ Higgsino-driven BLNV amplitudes like the supersymmetric $SU(5)$ GUTs discussed at length in Sect. 2.6.4. However, this issue is typically less severe in the $SO(10)$ context as there is often more than a single Δ_c-type (cf. Table 2.2) triplet Higgsino in the theory spectrum. This, in turn, yields more room to arrange their mixing in such a way that the lightest of the corresponding mass eigenstates (i.e., the field whose contribution should kinematically dominate the BLNV amplitudes) has, for instance, suppressed couplings to the first generation (s)quarks and (s)leptons.

3.2.2 Yukawa sector of simple $SO(10)$ GUTs

In general, the proliferation of the Δ_c-type triplets in the $SO(10)$ models has to do with the complete unification of the SM matter families within $SO(10)$ spinors. As beautiful as this is, it makes it essentially impossible to accommodate all the distinct features of the SM matter spectrum and mixing in a model with just a single "Yukawa-active" irrep coupled to the matter bilinear $\mathbf{16} \otimes \mathbf{16}$ (at the renormalizable level). The corresponding decomposition[5]

$$\mathbf{16} \otimes \mathbf{16} = \mathbf{10} \oplus \mathbf{120} \oplus \mathbf{126} \tag{3.18}$$

indicates that, along with the minimal choice of **10**, the most economical renormalizable models are those in which the rank reduction (i.e., $U(1)_{B-L}$ breaking) is triggered by the scalar **126** rather than **16** because the former can help also with attaining a rich-enough flavour structure, at least at the renormalizable level. It is also worth noting that out of the three options of (3.18) it is again only **126** which contains a SM singlet and whose VEV can thus source the Majorana mass term for the RH neutrinos.

Therefore, in what follows, we shall stick to the renormalizable $SO(10)$ grandunified models with the powerful **126** tensor high in their field-contents list.

3.3 The spectacular failure of the minimal SUSY SO(10)

To this end, in the current and in the next sections we shall first discuss the minimal supersymmetric $SO(10)$ GUT and only later on switch to the non-SUSY variant. This approach closely follows the historical development of the field which, on the non-SUSY side, was hindered by the early observation of tachyonic instabilities in the spectrum of

[5] Algebraically, the three factors correspond to a one-, three- and five-index fully antisymmetric tensors where, for the last one, only the self-dual component of the full 252-dimensional maximally antisymmetric tensor is taken.

the minimal model, cf. Sect. 3.4.3; at the same time (i.e., in mid 1980's) SUSY became so widely popular[6] that there was almost no impetus to explore such issues any further, cf. Sect. 3.4.

3.3.1 The structure of the the minimal SUSY $SO(10)$ GUT

With what has been said so far, one can right away formalize the hypothesis for the structure of the (renormalizable) Yukawa part of the minimal SUSY $SO(10)$ superpotential, namely,

$$W_Y = (\mathbf{16}_M^T Y_{10} \mathbf{16}_M)\mathbf{10}_H + (\mathbf{16}_M^T Y_{126} \mathbf{16}_M)\overline{\mathbf{126}}_H\,, \tag{3.19}$$

where $\mathbf{16}_M$ stands for a vector of three generations of matter, Y_{10} and Y_{126} denote the 3×3 Yukawa matrices (both symmetric due to the structure of the $SO(10)$ contractions) and $\mathbf{10}_H$ and $\overline{\mathbf{126}}_H$ are the two Higgs multiplets identified above.

Before writing the remaining parts of the superpotential involving, among other things, the Higgs field self-interactions (and, thus, implicitly, the scalar potential of the model), one has to choose very carefully the set of the GUT-symmetry breaking fields[7]. This is subject to several important requirements:

1. SUSY should not be broken along with the GUT symmetry. The reason is that GUT-scale F- or D-terms make it very hard to obtain the soft SUSY breaking scale in the desired TeV-scale ballpark. Since, however, a VEV of the complex $\overline{\mathbf{126}}$ inevitably leads to $\langle D_{\overline{126}}\rangle \neq 0$ the only way to ensure this is to include also the complex conjugated multiplet $\mathbf{126}$ whose contribution would cancel $\langle D_{\overline{126}}\rangle$.

2. The "$SU(5)$ trap", i.e., an intermediate stage looking very much like the problematic minimal SUSY $SU(5)$ theory of Sect. 2.6.4, should be avoided by adding a Higgs superfield whose SM singlets are not simultaneously singlets of the $SU(5)$. There are several such options offered in Table 3.2 like, e.g., **45**, **54** and **210**. Interestingly, **45** does not work because the F-terms align its VEVs with that of $\mathbf{126}_H$ [104] and $SU(5)$ remains unbroken.

3. The SM Higgs boson should be spanned on the doublet components of *both* **10** and $\overline{\mathbf{126}}$, otherwise one is back to the overly rigid situation of only one effective

[6] In this respect it is interesting to note that the rocketing popularity of the low-energy SUSY in the middle of 1980's was partly fuelled by the need to find a successor of the minimal $SU(5)$ which failed miserably on the prediction of the weak mixing angle, cf. Sect. 2.6.3. Remarkably, the mainstream went for further complication (marrying $SU(5)$ with SUSY) rather than re-thinking the unification basics.

[7] We already know that **126** is not enough as the SM singlet within is also a singlet of $SU(5)$ that would remain unbroken.

Yukawa structure at play. Hence, there should be a term in the superpotential providing their mixing; out of the list above this can be done (at the renormalizable level) only by **210**.

Thus, the complete Higgs sector of the minimal potentially realistic SUSY $SO(10)$ [105–111] contains 4 irreps, namely, $\mathbf{10}\oplus\mathbf{126}\oplus\overline{\mathbf{126}}\oplus\mathbf{210}$ with the corresponding superpotential of the form

$$\begin{aligned} W_H &= M_{10}\mathbf{10}_H^2 + M_{126}\mathbf{126}_H\overline{\mathbf{126}}_H + M_{210}\mathbf{210}_H^2 + \\ &+ \lambda\mathbf{210}_H^3 + \eta\,\mathbf{210}_H\mathbf{126}_H\overline{\mathbf{126}}_H + \alpha\mathbf{10}_H\mathbf{126}_H\mathbf{210}_H + \beta\mathbf{10}_H\overline{\mathbf{126}}_H\mathbf{210}_H\,. \end{aligned} \tag{3.20}$$

In spite of the complexity of W_H above several groups succeeded in calculating its spectrum in the SUSY limit[8] and, thus, a complete analysis of the GUT-scale thresholds – a crucial ingredient of any gauge running study – became possible, see e.g. [111–113]. The scalar and Higgsino masses can be written in terms of four independent VEVs (three real ones in **210** and one complex in $\mathbf{126}\oplus\overline{\mathbf{126}}$) usually denoted by

$$\langle(\mathbf{1},\mathbf{1},\mathbf{1})_{210}\rangle\equiv p\,,\ \langle(\mathbf{15},\mathbf{1},\mathbf{1})_{210}\rangle\equiv a\,,\ \langle(\mathbf{15},\mathbf{1},\mathbf{3})_{210}\rangle\equiv\omega\,,\ \langle(\mathbf{10},\mathbf{1},\mathbf{3})_{\overline{126}}\rangle\equiv\overline{\sigma}\,. \tag{3.21}$$

Notice that $(\mathbf{10},\mathbf{1},\mathbf{3})_{\overline{126}}$ contains the Δ_R scalar of the LR setting discussed in Sect. 3.1.1 and, thus, $\overline{\sigma}$ drives the scale of the type-I+II seesaw contributions to the netrino masses in this model. Moreover, with **126** at play, the flavour structure of both the type-I+II contributions are intimately related to the same Yukawa matrix Y_{126}.

From the flavor perspective, such a seesaw structure has also got other interesting properties:

1. The type-II contribution to the light neutrino masses is correlated with the charged sector Yukawa matrices via

$$\begin{aligned} Y_d v_d &= Y_{10}v_d^{10} + Y_{126}v_d^{126} &\equiv M_d \\ Y_l v_d &= Y_{10}v_d^{10} - 3Y_{126}v_d^{126} &\equiv M_l \end{aligned} \tag{3.22}$$

(with v_d^{10} and v_d^{126} denoting the projections of the MSSM down-type Higgs doublet VEV onto the defining components in $\mathbf{10}_H$ and $\overline{\mathbf{126}}_\mathbf{H}$), which is a consequence of the simplicity of W_Y of (3.19). In particular, one has

$$Y_{126}\propto M_d - M_l\,. \tag{3.23}$$

The absolute size of this neutrino mass contribution is then proportional to σ times the product of the relevant projections of $(\mathbf{1},\mathbf{2},\mathbf{2})_{10}$ and $(\overline{\mathbf{10}},\mathbf{2},\mathbf{2})_{210}$ onto

[8] In the SUSY limit the Higgs and Higgsino masses are identical and, thus, it is sufficient to calculate the latter.

the physical MSSM Higgs doublets. Note that all these factors are in principle calculable within a complete model.

2. If the type-II contribution dominates the seesaw formula one obtains an intriguing correlation between the size of the atmospheric neutrino mixing and the stipulated GUT-scale convergence of the b-quark and τ-lepton Yukawa couplings [114]

$$\tan^2\theta_A \sim \frac{\sin 2\theta_q}{2\sin^2\theta_q - \left(1-\frac{y_\tau}{y_b}\right)}\,, \tag{3.24}$$

(with θ_q corresponding to the $2-3$ CKM mixing) justifying its tendency towards maximality indicated by the experiment. Besides that, a relatively large reactor mixing angle is strongly preferred in the physically viable parts of the parameter space [115].

3. Also the flavour structure of the RH neutrino Majorana mass is proportional to Y_{126} which, in turn, enters the type-I seesaw formula as an inverse. The overall size of this contribution to the light neutrino mass matrix is, however, inverse proportional to σ.

3.3.2 The neutrino challenge to SUSY GUTs

As promising as the initial observations look, the minimal SUSY $SO(10)$ scheme turns out to be terminally ill when it comes to the global analysis of its flavour structure together with the constraints from the gauge unification. The devil is, as always, in detail (cf. [116] and [117]):

1. It turns out that no viable complete flavour fits exist if type-I seesaw contribution to the light neturino masses is suppressed (i.e., for large σ).

2. The fits in which type-I contribution is significant thus require the $B-L$ breaking scale σ well below the GUT scale and a very specific pattern of the MSSM VEV projections onto the defining doublets in **10** and $\overline{\mathbf{126}}$.

3. This, however, pushes the model into a regime in which a set of pseudo-Goldstone modes develop several orders of magnitude below the GUT scale and, hence, ruin completely the "too good to be true" MSSM gauge coupling convergence pattern.

Hence, the minimal potentially realistic supersymmetric incarnation of the $SO(10)$ grand unification paradigm of Sect. 3.3.1 has been decisively futilised in [116] and [117]. Though there have still been later attempts to save the situation by proposing minor amendments to the original scheme (see, e.g., [118–120]) the model has been to a large degree abandoned by the community and it is no longer considered as a viable route towards a complete theory of perturbative baryon number violation.

3.4 Quantum salvation of the minimal $SO(10)$ GUT

With the strict no-go for the minimal SUSY $SO(10)$ revealed in the previous section it is more than natural to turn one's attention back to the minimal non-supersymmetric version of the $SO(10)$ GUT, the more that TeV-scale SUSY becomes less and less appealing with the latest null results of (not only) the relevant LHC searches.

3.4.1 The cons and pros of the non-SUSY $SO(10)$ GUTs

Naïvely, life gets only more complicated in the non-SUSY context. With less symmetry imposed one has to deal with, e.g., higher-order operators governing the minimal defining structures ($d = 4$ interaction Hamiltonian vs. $d = 3$ superpotential at the renormalizable level), a more complicated Yukawa sector including non-holomorphic couplings, the issues related to the need for intermediate stages in the symmetry breaking pattern (as only one of the aspects of a generally more complicated quantum structure of the theory) etc.

On the other hand, with less assumptions on the theory shape the stakes are generally higher as it is way more straightforward to learn a lesson from its eventual failure than in the SUSY context. The non-supersymmetric unifications are also way easier to treat perturbatively in the vicinity of the GUT scale as the number of degrees of freedom to be integrated over in loop diagrams is reduced considerably.

3.4.2 The tree-level vacuum of the minimal $SO(10)$ Higgs models

One of the most delicate questions to be addressed in the framework of non-SUSY unifications are those concerning the structure of their vacuum. As there is no guarantee of the scalar potential convexity in its extrema[9] the calculation of the scalar spectrum (not possessing any fermionic counterpart) is always connected to the issues of its positivity. Hence, with the non-supersymmetric GUTs, any analysis of their viability must begin with a careful inspection of the corresponding Higgs sector.

As unlikely as it sounds, a full analysis of even the minimal $SO(10)$ Higgs model may be in fact a rather formidable. Following the reasoning of Sect. 3.2 its field content can be identified readily – either it is spanned on $\mathbf{45}\oplus\mathbf{16}$ or on $\mathbf{45}\oplus\mathbf{126}$, with different

[9] Recall that in supersymmetric theories the supersymmetric minimum is always the global one.

components therein playing similar roles (i.e., avoiding the $SU(5)$ "trap", triggering the rank reduction etc.) as in the SUSY context.

In what follows we shall consider both the $\mathbf{45} \oplus \mathbf{16}$ and $\mathbf{45} \oplus \mathbf{126}$ options along with each other as their main features (especially those relevant for the operation of the symmetry breaking mechanism) are to a large degree similar in both these settings. Despite the proximity of M_G to the Planck scale we shall stick to the renormalizable version(s) of these minimal model(s) as, in most cases, the contributions of the higher order operators should play a sub-leading role (if not entirely negligible, see Sect. 3.4.4).

The tree-level scalar potentials and masses

The $\mathbf{45} \oplus \mathbf{16}$ variant of the minimal $SO(10)$ Higgs model is defined by the scalar potential of the form

$$V_{45\oplus 16} = V_{45} + V_{16} + V_{45-16}\,, \tag{3.25}$$

where, following the definitions given in Appendix A of [121] (and omitting any subscripts distingushing between different types of fileds[10])

$$\begin{aligned} V_{45} &= -\frac{\mu^2}{4}\mathrm{Tr}\,\mathbf{45}^2 + \frac{a_1}{4}(\mathrm{Tr}\,\mathbf{45}^2)^2 + \frac{a_2}{4}\mathrm{Tr}\,\mathbf{45}^4\,, \\ V_{16} &= -\frac{\nu^2}{2}\mathbf{16}^\dagger\mathbf{16} + \frac{\lambda_1}{4}(\mathbf{16}^\dagger\mathbf{16})^2 + \frac{\lambda_2}{4}(\mathbf{16}_+^\dagger\Gamma_j\mathbf{16}_-)(\mathbf{16}_-^\dagger\Gamma_j\mathbf{16}_+)\,, \end{aligned} \tag{3.26}$$

and

$$V_{45-16} = \alpha(\mathbf{16}^\dagger\mathbf{16})\mathrm{Tr}\,\mathbf{45}^2 + \beta\mathbf{16}^\dagger\mathbf{45}^2\,\mathbf{16} + \tau\mathbf{16}^\dagger\mathbf{45}\,\mathbf{16}\,. \tag{3.27}$$

In case of the $\mathbf{45} \oplus \mathbf{126}$ variant the scalar potential shares the universal V_{45} part (3.26) with that of (3.25) but differs in the other two terms, namely

$$V_{45\oplus 126} = V_{45} + V_{126} + V_{45-126}\,, \tag{3.28}$$

with

$$\begin{aligned} V_{126} &= -\frac{\nu^2}{5!}(\mathbf{126}^*\mathbf{126})_0 \\ &+ \frac{\lambda_0}{(5!)^2}(\mathbf{126}^*\mathbf{126})_0(\mathbf{126}^*\mathbf{126})_0 + \frac{\lambda_2}{(4!)^2}(\mathbf{126}^*\mathbf{126})_2(\mathbf{126}^*\mathbf{126})_2 \\ &+ \frac{\lambda_4}{(3!)^2(2!)^2}(\mathbf{126}^*\mathbf{126})_4(\mathbf{126}^*\mathbf{126})_4 + \frac{\lambda_4'}{(3!)^2}(\mathbf{126}^*\mathbf{126})_{4'}(\mathbf{126}^*\mathbf{126})_{4'} \\ &+ \frac{\eta_2}{(4!)^2}(\mathbf{126}\,\mathbf{126})_2(\mathbf{126}\,\mathbf{126})_2 + \frac{\eta_2^*}{(4!)^2}(\mathbf{126}^*\mathbf{126}^*)_2(\mathbf{126}^*\mathbf{126}^*)_2\,, \end{aligned} \tag{3.29}$$

[10]Note that, unlike in the SUSY context, all fields involved here are Lorentz scalars.

and

$$
\begin{aligned}
V_{45-126} &= \frac{i\tau}{4!}(45)_2(\mathbf{126}^*\mathbf{126})_2 + \frac{\alpha}{2\cdot 5!}(\mathbf{45}\,\mathbf{45})_0(\mathbf{126}^*\mathbf{126})_0 \\
&+ \frac{\beta_4}{4\cdot 3!}(\mathbf{45}\,\mathbf{45})_4(\mathbf{126}^*\mathbf{126})_4 + \frac{\beta_4'}{3!}(\mathbf{45}\,\mathbf{45})_{4'}(\mathbf{126}^*\mathbf{126})_{4'} \\
&+ \frac{\gamma_2}{4!}(\mathbf{45}\,\mathbf{45})_2(\mathbf{126}\,\mathbf{126})_2 + \frac{\gamma_2^*}{4!}(\mathbf{45}\,\mathbf{45})_2(\mathbf{126}^*\mathbf{126}^*)_2\,.
\end{aligned}
\tag{3.30}
$$

The brackets above correspond to the following $SO(10)$ covariant structures (with or without complex conjugation):

$$
\begin{aligned}
(\mathbf{126}^*\mathbf{126})_0 &\equiv \mathbf{126}^*_{ijklm}\mathbf{126}_{ijklm}\,, \\
(\mathbf{126}^*\mathbf{126})_2 &\equiv (\mathbf{126}^*\mathbf{126})_{mn} \equiv \mathbf{126}^*_{ijklm}\mathbf{126}_{ijkln}\,, \\
(\mathbf{126}^*\mathbf{126})_4 &\equiv (\mathbf{126}^*\mathbf{126})_{lmno} \equiv \mathbf{126}^*_{ijklm}\mathbf{126}_{ijkno}\,,
\end{aligned}
\tag{3.31}
$$

where all the latin indices run from 1 to 10 and their pairs are summed over. The contractions of these terms in Eqs. (3.29) and (3.30) are obvious with the only exceptions of the $4'$ brackets that read

$$
\begin{aligned}
(\mathbf{126}^*\mathbf{126})_{4'}(\mathbf{126}^*\mathbf{126})_{4'} &\equiv (\mathbf{126}^*\mathbf{126})_{lmno}(\mathbf{126}^*\mathbf{126})_{lnmo}\,, \\
(\mathbf{45}\,\mathbf{45})_{4'}(\mathbf{126}^*\mathbf{126})_{4'} &\equiv \mathbf{45}_{lm}\mathbf{45}_{no}(\mathbf{126}^*\mathbf{126})_{lnmo}\,.
\end{aligned}
\tag{3.32}
$$

The vacuum structure of the minimal $SO(10)$ models

The adjoint **45** of $SO(10)$ decomposes under the $SU(3)_c\otimes SU(2)_L\otimes U(1)_Y$ of the SM as

$$
\begin{aligned}
\mathbf{45} &= (\mathbf{3},\mathbf{2},-\tfrac{5}{6})\oplus(\overline{\mathbf{3}},\mathbf{2},+\tfrac{5}{6})\oplus(\mathbf{8},\mathbf{1},0)\oplus(\mathbf{1},\mathbf{3},0) \\
&\oplus (\mathbf{3},\mathbf{2},+\tfrac{1}{6})\oplus(\overline{\mathbf{3}},\mathbf{2},-\tfrac{1}{6})\oplus(\mathbf{3},\mathbf{1},+\tfrac{2}{3})\oplus(\overline{\mathbf{3}},\mathbf{1},-\tfrac{2}{3}) \\
&\oplus (\mathbf{1},\mathbf{1},+1)\oplus(\mathbf{1},\mathbf{1},-1)\oplus(\mathbf{1},\mathbf{1},0)\oplus(\mathbf{1},\mathbf{1},0)\,.
\end{aligned}
\tag{3.33}
$$

Hence, there are two SM singlets in this multiplet which, in realistic scenarios, may possess non-zero VEVs. Note, however, that these two singlets are not equivalent from the viewpoint of the intermediate symmetries: one of them descends from the $(\mathbf{15},\mathbf{1},\mathbf{1})$ Pati-Salam component of **45** while the other one resides in $(\mathbf{1},\mathbf{1},\mathbf{3})$. Hence, the VEV of the former (to be called ω_R) preserves the $SU(2)_R$ subgroup of the $SO(10)$ while the latter (ω_{BL}) leaves intact the $U(1)_{B-L}$ factor.

Concerning the complex irrep of the models (i.e., **16** or **126**) one of their common features is the presence of one SM singlet which is capable of breaking the $U(1)_{B-L}$ symmetry and, thus, set the seesaw scale. At the same time, this field is also a singlet of the $SU(5)$ subgroup of $SO(10)$ and, thus, its VEV (to be denoted σ) is not sufficient to provide the entire $SO(10)$ symmetry breaking, cf. Sect. 3.3.1.

The masses of the scalar triplet and octet of the 45

In what follows we shall focus on the masses of the fields in the first line of the decomposition (3.33) which, due to their transformation properties, do not mix with any of the fields in **16** or **126**. This means that the minimisation of the scalar potential yields the same simple formulae in either of the two settings, namely,

$$M^2_{(1,3,0)} = 2a_2(\omega_R - \omega_{BL})(\omega_{BL} + 2\omega_R)\,, \tag{3.34}$$
$$M^2_{(8,1,0)} = 2a_2(\omega_{BL} - \omega_R)(\omega_R + 2\omega_{BL})\,. \tag{3.35}$$

Remarkably enough, these expressions are simultaneously non-negative if an only if

$$a_2 > 0 \quad \text{and} \quad -2 < \frac{\omega_{BL}}{\omega_R} < -\frac{1}{2}\,. \tag{3.36}$$

Note, however, that the required proximity (up to a sign) of ω_R and ω_{BL} prefers a very specific shape of the vacuum manifold corresponding to a symmetry breaking pattern passing through the vicinity of the so called flipped $SU(5)\otimes U(1)$ intermediate stage (corresponding to the $\omega_R = -\omega_R$ situation).

However, unlike in the settings discussed below in Sect. 3.6.2 (in which only the non-abelian gauge couplings of the SM are required to unify) this observation represents a serious problem in the $SO(10)$ GUT context. Qualitatively, the issue is somewhat similar to that encountered in Sect. 3.3.2: on one hand, there is a need to have the seesaw scale σ well below $M_G \sim 10^{16}$ GeV to conform the neutrino sector constraints[11] but, on the other hand, the flipped $SU(5)\otimes U(1)$ intermediate symmetry must be broken (by the same σ) in the proximity of M_G to avoid the issues with overly fast proton decay. Hence, the non-SUSY $SO(10)$ GUTs with the first symmetry breaking step driven by the adjoint **45** irrep are very unlikely to conform even the basic phenomenological requirements including the non-tachyonicity of the scalar spectrum in the potentially realistic symmetry-breaking chains (i.e., those avoiding an intermediate-scale $SU(5)$-like gauge dynamics).

3.4.3 The minimal $SO(10)$ Higgs model(s) at the loop level

Interestingly, most of what is written above was understood already in the early 1980's [126–129] and, since then, it was generally assumed that the field of renormalizable $SO(10)$ GUTs broken by either $\mathbf{45}\oplus\mathbf{16}$ or $\mathbf{45}\oplus\mathbf{126}$ scalar fields is sterile. To this end, the situation changed completely in 2010 with the author's study [130] where it was

[11] Actually, the gauge unification constraints on σ from the consistency of the gauge unification pattern in non-SUSY settings [122–125] prefer it in the vicinity of 10^{11} GeV, i.e., even lower than the scale favoured by the seesaw!

shown that the tachyonicity of the scalar spectrum along the potentially realistic symmetry breaking chains (i.e., those passing through either the $SU(4)_C \otimes SU(2)_L \otimes U(1)_R$ or the left-right symmetric $SU(3)_c \otimes SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$ intermediate stages) is an artefact of the tree-level analysis and can be avoided if radiative corrections are taken into account.

The merit of [130] consists in the observation that the tree-level mass formulae (3.34) and (3.35) are unexpectedly simple given the fact that, in principle, all terms in V_{45} and $V_{45-16/126}$ (besides the τ piece therein) should contribute[12] already at the tree level. A thorough investigation of this phenomenon reveals that this is all due to the pseudo-Goldstone nature of the $(\mathbf{1}, \mathbf{3}, 0)$ and the $(\mathbf{8}, \mathbf{1}, 0)$ fields which, in the limit of "moduli-only" terms kept in the potential, become exact Goldstone modes of the spontaneously broken enhanced global symmetry; hence, their masses should be proportional to just a limited set of parameters which break this global symmetry explicitly. In the same perspective, the presence of only the a_2-proportional contribution in (3.34) may be attributed to the particular shape[13] of all the other potentially relevant tree-level contractions in $V_{45-16/126}$. Since, however, this does not necessarily apply to the radiative corrections, the formulae (3.34) are expected to receive different types of loop contributions which, in turn, may be capable of resolving the tachyonicity conundrum of Sect. 3.4.2.

This expectation if further justified by the fact that, in the perturbative regime, a_2 is almost automatically pushed well below 1 by one of the minimisation conditions. This, in the $\mathbf{45} \oplus \mathbf{126}$ setting, reads[14]

$$\tau = 2\beta_4'(3\omega_{BL} + 2\omega_R) + a_2 \frac{\omega_{BL}\omega_R}{|\sigma|^2}(\omega_{BL} + \omega_R)\,, \tag{3.37}$$

which, given the need to have the seesaw scale $|\sigma|$ well below M_G corresponding to the leading ω (and assuming that the other one is not parametrically smaller[15]), can be fulfilled in the perturbative regime (i.e., with loops playing a sub-leading role here) only for $a_2 \lesssim |\sigma|^2/\omega_R\omega_{BL}$.

[12] This expectation reflects the presence of at least two powers of **45** in each of these terms which, in the broken phase, should generate bilinears for the fields in **45**.

[13] For instance, the absence of the contribution from the β-proportional part of $V_{45\oplus16}$ can be understood by noticing that the relevant term has got the same group structure as the corresponding piece in the mass matrix for the gauge fields. Since, however, the VEVs in the scalar **45** can not contribute to the masses of the gluons or the SM A-fields, the same must happen (at least at the tree level) to their scalar counterparts.

[14] It can be objected that the specific form of the tadpole equation (3.37) corresponds to the tree level approximation only; it should, however, still be the leading piece even at higher loops if the theory is perturbative and, thus, the tree-level results should represent a good approximation to the full case.

[15] This is likely to be so because in the opposite case it is typically difficult to get a completely non-tachyonic rest of the scalar spectrum.

Let us illustrate this by writing down the one-loop contributions to $M^2_{(1,3,0)}$ and $M^2_{(8,1,0)}$ from the Feynman diagrams including the gauge degrees of freedom – these corrections are indeed universal to both the $\mathbf{45}\oplus\mathbf{16}$ and $\mathbf{45}\oplus\mathbf{126}$ models. A thorough analysis of the relevant piece of the one-loop effective potential à la Coleman and Weinberg [131] reveals [132]

$$\Delta^{1-\text{loop}}_{\text{gauge}} M^2_{(1,3,0)} = \frac{g^4}{16\pi^2}(19\omega_{BL}^2 + \omega_{BL}\omega_R + 16\omega_R^2) + \Delta^{\log}_{(1,3,0)}\,, \tag{3.38}$$

$$\Delta^{1-\text{loop}}_{\text{gauge}} M^2_{(8,1,0)} = \frac{g^4}{16\pi^2}(22\omega_{BL}^2 + \omega_{BL}\omega_R + 13\omega_R^2) + \Delta^{\log}_{(8,1,0)}\,, \tag{3.39}$$

where

$$\begin{aligned}\Delta^{\log}_{(1,3,0)} = \frac{3g^4}{16\pi^2\omega'}\Bigg\{ & 8\omega_{BL}\left(|\sigma|^2+\omega_{BL}{}^2\right)\log\left[\frac{2g^2\left(|\sigma|^2+\omega_{BL}{}^2\right)}{\mu^2}\right] \\ & -4\omega_R\left(|\sigma|^2+\omega_R{}^2\right)\log\left[\frac{2g^2\left(|\sigma|^2+\omega_R{}^2\right)}{\mu^2}\right] + 2\omega'^3\log\left[\frac{\frac{1}{2}g^2\omega'^2}{\mu^2}\right] \\ & -\left[4|\sigma|^2(2\omega_{BL}-\omega_R)+\omega^2(5\omega_{BL}-4\omega_R)\right]\log\left[\frac{2g^2\left(|\sigma|^2+\frac{1}{4}\omega^2\right)}{\mu^2}\right]\Bigg\},\end{aligned} \tag{3.40}$$

$$\begin{aligned}\Delta^{\log}_{(8,1,0)} = \frac{3g^4}{32\pi^2\,\omega'}\Bigg\{ & 4\left[|\sigma|^2(3\omega_{BL}+\omega_R)+\omega_{BL}^2(\omega_{BL}+3\omega_R)\right]\log\left[\frac{2g^2\left(|\sigma|^2+\omega_{BL}{}^2\right)}{\mu^2}\right] \\ & -8\omega_R\left(|\sigma|^2+\omega_R{}^2\right)\log\left[\frac{2g^2\left(|\sigma|^2+\omega_R{}^2\right)}{\mu^2}\right] + \omega'^3\log\left[\frac{\frac{1}{2}g^2\omega'^2}{\mu^2}\right] \\ & -\left[4|\sigma|^2(3\omega_{BL}-\omega_R)+\omega^2(7\omega_{BL}-5\omega_R)\right]\log\left[\frac{2g^2\left(|\sigma|^2+\frac{1}{4}\omega^2\right)}{\mu^2}\right]\Bigg\},\end{aligned} \tag{3.41}$$

provided

$$\omega \equiv \omega_R + \omega_{BL} \quad \text{and} \quad \omega' \equiv \omega_R - \omega_{BL}\,. \tag{3.42}$$

It is not difficult to show that there is indeed a lot of points in the $(\omega_{BL}, \omega_R, |\sigma|)$ space corresponding to the phenomenologically preferred symmetry breaking chains (i.e., those avoiding intermediate $SU(5)$ stages) where both expressions (3.38) and (3.39) are positive.

3.4.4 The minimal potentially realistic and testable GUTs

Hence, after almost 30 years in oblivion, the minimal $SO(10)$ Higgs model was brought back to life [133–136] as a seed of a potentially realistic theory which, however, is inherently of a quantum nature. The natural question is then whether such a theory can, in some of its parameter space point(s), accommodate all the low-energy data and, if affirmative, what would be its predictions for the new physics signals such as proton decay, leptonic CP violation, absolute neutrino mass scale etc.

Planck-scale effects in the GUT-scale determination

It is remarkable that these efforts can be further justified by another rather unique feature the minimal $SO(10)$ Higgs models possess, namely, their particular robustness with respect to the Planck-scale-induced effects in the gauge unification analysis. To see this, let us recall that whenever any unified gauge symmetry is being broken by a multiplet (to be called Φ) which admits a $d=5$ coupling to the relevant gauge-kinetic form

$$\mathcal{L}^{(5)} \ni \frac{\rho}{\Lambda} F_{\mu\nu} \Phi F^{\mu\nu} \,, \tag{3.43}$$

with ρ denoting a dimensionless (presumably $\mathcal{O}(1)$) effective coupling and Λ standing for the effective cut-off scale, one gets a non-canonical gauge-kinetic form in the broken phase

$$\mathcal{L}_{\text{kin}} \ni -\frac{1}{4}\left(1-4\rho\frac{\langle\Phi\rangle}{\Lambda}\right) F_{\mu\nu}F^{\mu\nu} \,. \tag{3.44}$$

For the sake of retaining the standard perturbative expansion, the gauge fields must be first canonically normalized by a suitable rescaling transformation which, however, depends (on the ratio of) two, in principle unknown, quantities ρ and Λ. This, in general, induces inhomogeneous shifts in the definitions of the three effective SM couplings in terms of the unified one and, hence, uncertainties in the corresponding matching between the GUT and any lower-energy theory. Surprisingly, even for as small as 1% effects of this kind (corresponding to the very natural choice of $\Lambda = M_{Pl}$ and $\rho \sim 1$ with $\langle\Phi\rangle \sim 10^{16}\,\text{GeV}$), these errors can have serious impact on the precision calculations of M_G as they enter the relevant formulae exponentially. In practice, the resulting uncertainty can easily "smear" thus obtained M_G into a domain stretching over more than an order of magnitude! This, however, sets a limit on the accuracy attainable in most proton lifetime calculations (typically at the level of several orders of magnitude) which, in turn, renders all attempts to discriminate among different models on the basis of (non)observation of proton decay essentially meaningless.

Taming the leading Planck-scale effects in the minimal $SO(10)$ GUT

Interestingly, *the irreducible theoretical uncertainties of the kind described above are absent from the minimal $SO(10)$ GUT*. The reason is that its unified-symmetry-breaking VEV resides in the scalar transforming as the adjoint **45** whose coupling to the $F_{\mu\nu}F^{\mu\nu}$ bilinear is identically zero due to the antisymmetry of $\mathbf{45}^{ab}$ in the group indices, i.e.,

$F^a_{\mu\nu}\Phi^{ab}F^{b\mu\nu}=0$. Note that this is not the case of the other popular $SO(10)$ symmetry breaking models utilising either **54** or **210** because both these multiplets are present in the symmetric product of two adjoints:

$$[\mathbf{45}\otimes\mathbf{45}]_{\rm sym}=\mathbf{54}\oplus\mathbf{210}\oplus\mathbf{770}\,. \tag{3.45}$$

In this sense, *the $SO(10)$ models with the GUT-scale symmetry breaking triggered by the adjoint are arguably very unique concerning their proton-decay predictive potential.* This, together with the ongoing construction of the new generation of large-scale detectors such as Hyper-K [68] or DUNE [137] (fuelled mainly by their potential to serve as very powerful neutrino telescopes), is the main reason why this specific class of the minimal $SO(10)$ scenarios has been receiving so much attention recently.

The minimal potentially realistic and testable $SO(10)$ GUT

The first steps along the lines of constructing and working out the potentially realistic and testable $SO(10)$ GUTs of this kind have been attempted recently in studies [133, 136]. The salient features of these settings are:

- *Scalar sector containing* $\mathbf{45}\oplus\mathbf{126}$: The choice of the 5-index antisymmetric (anti-) self-dual tensor of the $SO(10)$ rather than the spinorial **16** discussed in Sect. 3.4.2 is motivated mainly by the preference of a renormalizable (and, thus, potentially predictive) Yukawa sector. For that sake, **126** is almost ideal as it contributes not only to the effective Yukawa couplings of the charged SM matter fermions but it also generates a large mass term for the RH neturinos (and, thus, a natural type-I seesaw contribution[16] to the light neutrino masses). On top of that, the $U(1)_{B-L}$ subgroup of $SO(10)$ broken by the SM singlet of **126** leaves behind a residual Z_2 symmetry which behaves like a matter parity and, thus, can stabilize fermionic dark-matter candidates of various kinds [138]. Note also that the Yukawa sector of the minimal potentially realistic renormalizable model must be equipped with one more scalar irrep in order to smear the effective Yukawa degeneracies across different matter sectors; this is usually taken to be **10**, partly for its minimality and partly for the fact that it does not interfere with any of the findings above due to the absence of any SM singlets within.

- *Complicated vacuum structure*: The purely quantum nature of the models under consideration bring a notorious difficulty with the classification of the shapes of

[16] Recall that there is typically also a type-II seesaw piece emerging from the induced sub-electroweak-scale VEV of the scalar $SU(2)_L$ triplet in **126**, cf. Sect. 3.1.1.

the scalar spectrum conforming the conditions of non-tachyonicity and perturbativity which are obvious prerequisites of any sensible attempts of the GUT-scale determination as the precursor of the subsequent proton lifetime calculations. The effective potential methods are typically used for this purpose (cf. Sect. 3.4.3) but even with these at hand a complete chart of the regions of the model's parameter space conforming these constraints is still subject of an intensive research.

- *Intermediate scales*: Irrespective of the details of the high-energy spectrum in the fully realistic settings there are features that can be expected already at the current level of understanding. The most prominent is perhaps the need to push (at least) one of the naturally heavy scalar fields well below the GUT scale, otherwise there would be no realistic gauge unification pattern supporting a $B-L$ breaking (seesaw) scale in its phenomenologically preferred ballpark of about 10^{12-13} GeV. Pushed to the extreme, it is even conceivable to get some of these states relatively close to the LHC domain, cf. [133].[17]

For more information the reader is kindly deferred to the aforementioned studies [133, 136] and a recent review article [132].

3.5 Aspects of renormalization group evolution in theories with more than a single U(1) gauge factor

In the $SO(10)$ GUTs of Sect. 3.1.3 or their left-right-symmetric descendants discussed in Sect. 3.1.1 one often encounters a situation in which an intermediate-scale effective gauge theory features two Abelian factors (like, e.g., $U(1)_R \otimes U(1)_{B-L}$ of the breaking chains in Fig. 3.1 passing along the F' branch there). As innocent as such a situation looks the occurrence of multiple $U(1)$ gauge factors has rather non-trivial consequences for the quantum structure of the theory including a spectacular proliferation of couplings required for their formal renormalizability and also for the consistency of the overall physical picture.

In this section we shall pass through the basics of the renormalization procedure in such settings starting with a short review of the situation in the spinorial[18] QED

[17]To this end, it is perhaps worth noting that there are just two candidate multiplets in the minimal $\mathbf{45} \oplus \mathbf{126}$ $SO(10)$ Higgs model that can support viable symmetry breaking patterns without invoking any other field in the GUT desert: the $(\mathbf{8}, \mathbf{2}, +\frac{1}{2})$ and $(\mathbf{6}, \mathbf{3}, +\frac{1}{3})$ scalars. Remarkably enough, the same fields have been identified recently [139] in the low-energy SM effective field theory approach to the peculiar B-decay anomalies observed by Belle and LHCb as the most promising candidates for the new degrees of freedom underpinning the BSM dynamics behind these effects.

[18]The spinorial version of QED has been chosen only for illustration properties; the same principles

followed by a thorough discussion of the peculiarities encountered in its simplest $U(1)\otimes U(1)$ extension.

3.5.1 Renormalization of Abelian gauge theories (QED)

Let us start with the stardard QED bare Lagrangian (with the charge of the Dirac spinor ψ normalized to $Q_\psi = 1$)

$$\mathcal{L}_B = \overline{\psi}_B(\partial\!\!\!/ - m_B)\psi_B - ie_B\overline{\psi}_B A\!\!\!/_B\psi_B - \frac{1}{4}F_{B\mu\nu}F_B^{\mu\nu}\,, \tag{3.46}$$

which is conveniently redefined in terms of the renormalized quantities as $\mathcal{L}_B = \mathcal{L} + \delta\mathcal{L}$ where

$$\mathcal{L} = \overline{\psi}(\partial\!\!\!/ - m)\psi - ie\overline{\psi}A\!\!\!/\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.47}$$

and

$$\delta\mathcal{L} = \overline{\psi}_B(\partial\!\!\!/ - m_B)\psi_B - ie_B\overline{\psi}_B A\!\!\!/_B\psi_B - \frac{1}{4}F_{B\mu\nu}F_B^{\mu\nu} - \overline{\psi}(\partial\!\!\!/ - m)\psi + ie\overline{\psi}A\!\!\!/\psi + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,. \tag{3.48}$$

Defining the renormalized fields in the standard manner, i.e., $\psi_B = Z_\psi^{1/2}\psi$ and $A_B = Z_A^{1/2}A$, the fit to the "traditional form" of the counterterm Lagrangian

$$\delta\mathcal{L} = \delta Z_\psi\overline{\psi}\partial\!\!\!/\psi - \delta m\overline{\psi}\psi - i\delta e\overline{\psi}A\!\!\!/\psi - \delta Z_A\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.49}$$

yields

$$\delta Z_\psi = Z_\psi - 1 \qquad \text{and} \qquad \delta Z_A = Z_A - 1\,, \tag{3.50}$$

together with

$$m_B = Z_\psi^{-1}(m + \delta m) \equiv Z_\psi^{-1}Z_m m\,, \tag{3.51}$$

$$e_B = Z_\psi^{-1}Z_A^{-1/2}(e + \delta e) \equiv Z_\psi^{-1}Z_A^{-1/2}Z_e e\,, \tag{3.52}$$

provided

$$m + \delta m \equiv Z_m m \qquad \text{and} \qquad e + \delta e \equiv Z_e e\,. \tag{3.53}$$

Hence, the fundamental Lagrangian in terms of the renormalized quantities reads

$$\mathcal{L}_B = \overline{\psi}(\partial\!\!\!/ - m)\psi - ie\overline{\psi}A\!\!\!/\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \delta Z_\psi\overline{\psi}\partial\!\!\!/\psi - \delta Z_m m\overline{\psi}\psi - i\delta Z_e e\overline{\psi}A\!\!\!/\psi - \delta Z_A\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{3.54}$$

with $\delta Z_e = Z_e - 1$ and $\delta Z_m = Z_m - 1$; this, subsequently, leads to the standard Feynman rules one can find in the textbooks. In the QED, all the factors above are numbers

apply also in the scalar version of the theory.

and all UV divergences in all orders of the perturbative expansion can be absorbed into redefinitions of $\delta Z_\psi, \delta Z_A, \delta Z_m$ and δZ_e.

Formally, the QED beta-function is then obtained by taking a derivative with respect to the renormalization scale in Eq. (3.52) and, similarly, the running mass is governed by Eq. (3.51). The relevant Ward-Takahashi identity $Z_\psi = Z_e$ which holds to all orders in perturbation theory then ensures that, for the sake of the beta function calculation, it is sufficient to compute δZ_A in any given renormalization scheme

$$e_B = Z_A^{-1/2} e\,. \tag{3.55}$$

In order to be able to interpret the renormalization scale μ as an energy at which a certain process is considered one should take the (logarithmic) derivative of δZ_A in the class of momentum schemes; this, up to two loops, is however identical to taking the $\log \mu$-derivative of its (much simpler) MS or $\overline{\text{MS}}$ form with respect to the relevant UV-divergence structure (ε^{-1} in dimensional regularisation with $d = 4 - 2\varepsilon$).

3.5.2 Renormalization of QED squared

In QED squared, i.e., in the $U(1) \otimes U(1)$ gauge theory one should assume that there are at least 2 different fields ψ_i, $i = 1, 2$ (otherwise there is not point in talking about more than a single gauge factor) with charges Q_{ij} under the j-th $U(1)$ with an associated coupling e_j. Assuming that ψ_1 and ψ_2 carry different Q-charges (and, thus, they are not identical and can not mix) the renormalization factors δZ_ψ and δZ_m defined above should be just doubled, i.e., $\delta Z_\psi \to \delta Z_{\psi_i}$, $\delta Z_m \to \delta Z_{m_i}$. Analogously, each group factor receives its own gauge coupling e_j and, hence, $\delta Z_e \to \delta Z_{e_j}$.

The naïve definition of the Z_A counterterm in QED squared

Similarly, there are 2 gauge fields A_j so, naïvely, one is tempted to do the same, namely $\delta Z_A \to \delta Z_{A_j}$ hoping that it would suffice. *But it does not.* The reason is that there are in general three different UV-divergent diagrams corresponding to the one-loop gauge propagator corrections (a.k.a. vacuum polarisation), namely, the diagonal $A_1 - A_1$ and $A_2 - A_2$ ones whose divergences can be absorbed in $\delta Z_{A_{1,2}}$, but also a third one, $A_1 - A_2 = A_2 - A_1$ for which there is no counterterm left unless one introduces the off-diagonal $F_{1\mu\nu} F_2^{\mu\nu}$ piece into the Lagrangian[19]. The presence of such a term should be, in fact, even expected; unlike for non-Abelian field strength tensors which, by definition, carry group indices, the Abelian-case $F_{\mu\nu}$ structure is not only gauge covariant but

[19] Strictly speaking, this is true only if at least one matter field is charged under both groups, otherwise the theory trivially decays into two non-communicating sectors.

even invariant and, thus, a Lorentz contraction like $F_{1\mu\nu}F_2^{\mu\nu}$ qualifies as a Lagrangian density contribution.

The "correct" definition of counterterms in QED squared

Thus, from scratch, the gauge-kinetic part of the QED squared Lagrangian calls for a matrix structure:

$$\mathcal{L}_B = \overline{\psi}_B(\partial\!\!\!/ - m_B)\psi_B - iQe_B\overline{\psi}_B A\!\!\!/_B\psi_B - \frac{1}{4}F_{B\mu\nu}\xi_B F_B^{\mu\nu}\,. \tag{3.56}$$

As usual, $\mathcal{L}_B$ can be decomposed into $\mathcal{L}_B = \mathcal{L} + \delta\mathcal{L}$ where

$$\mathcal{L} = \overline{\psi}(\partial\!\!\!/ - m)\psi - i\overline{\psi}QeA\!\!\!/\psi - \frac{1}{4}F_{\mu\nu}\xi F^{\mu\nu}\,, \tag{3.57}$$

$$\begin{aligned}\delta\mathcal{L} &= \overline{\psi}(Z_\psi - 1)\partial\!\!\!/\psi - \overline{\psi}(Z_\psi^{1/2}m_B Z_\psi^{1/2} - m)\psi - i\overline{\psi}(Z_\psi^{1/2}Qe_B Z_A^{1/2}Z_\psi^{1/2} - Qe)A\!\!\!/\psi \\ &- \frac{1}{4}F_{\mu\nu}(Z_A^{1/2}\xi_B Z_A^{1/2} - \xi)F^{\mu\nu}\,,\end{aligned} \tag{3.58}$$

with "matrix redefinitions" of the fields

$$\psi_B = Z_\psi^{1/2}\psi \qquad \text{and} \qquad A_B = Z_A^{1/2}A\,. \tag{3.59}$$

Needless to say, everything is in principle a vector or a matrix now, and ξ in particular. Since, however, the matter fields do not mix, Z_ψ, m_B and m can be taken diagonal and one can clump these factors and define δZ_ψ and δZ_m as in Eqs. (3.50) and (3.53) (with all relevant quantities replaced by diagonal matrices) and get

$$\delta\mathcal{L} = \overline{\psi}\delta Z_\psi\partial\!\!\!/\psi - \overline{\psi}\delta Z_m m\psi - i\overline{\psi}(Z_\psi Qe_B Z_A^{1/2} - Qe)A\!\!\!/\psi - \frac{1}{4}F_{\mu\nu}(Z_A^{1/2}\xi_B Z_A^{1/2} - \xi)F^{\mu\nu}\,.$$

For the time being, we shall retain a generic ξ, i.e., the kinetic terms of the gauge fields shall not be canonically normalized. To proceed, one should define

$$Q\delta e \equiv Z_\psi Qe_B Z_A^{1/2} - Qe \quad \Leftrightarrow \quad e + \delta e = Q^{-1}Z_\psi Qe_B Z_A^{1/2}\,, \tag{3.60}$$

in full analogy with (3.53) and

$$\delta\xi \equiv Z_A^{1/2}\xi_B Z_A^{1/2} - \xi\,. \tag{3.61}$$

At this point, it may not be a-priori clear how to define the multiplicative counterterm for the gauge coupling: indeed, e has got two qualitatively different indices and one can in principle multiply from any side. However, it is much more convenient to do it from the left, i.e.,

$$e + \delta e \equiv Z_e e\,, \tag{3.62}$$

because then one reveals

$$e_B = Q^{-1} Z_\psi^{-1} Q Z_e e Z_A^{-1/2} \tag{3.63}$$

and, thanks to the Ward identity $Q^{-1} Z_\psi^{-1} Q Z_e = 1$, one obtains a matrix version of the equation (3.55) in the form

$$e_B = e Z_A^{-1/2} \,. \tag{3.64}$$

However, a multiplicative renormalization does not make much sense for the gauge kinetic term of Eq. (3.61), so for that one we rather stick to the additive convention. In any case, the counterterm to the gauge kinetic term reads simply

$$-\frac{1}{4} F_{\mu\nu} \delta\xi F^{\mu\nu} \tag{3.65}$$

and the full counterterm Lagrangian receives the final form

$$\delta\mathcal{L} = \overline{\psi} \delta Z_\psi \partial\!\!\!/ \psi - \overline{\psi} \delta Z_m m \psi - i \overline{\psi} Q \delta Z_e e A\!\!\!/ \psi - \frac{1}{4} F_{\mu\nu} \delta\xi F^{\mu\nu} \,. \tag{3.66}$$

3.5.3 The link between Z_A and $\delta\xi$ in different renormalization schemes

This is all right in principle but not very practical yet – on one side one can relatively easily determine $\delta\xi$ from the structure of the UV divergences of the gauge field propagators but, on the other hand, its correspondence to the central quantity of our interest, namely, Z_A in Eq. (3.64), is non-linear, cf. (3.61). It is, however, relatively easy to trade Z_A for $\delta\xi$ and, thus, connect the running of e to a quantity at hand. To this end, consider the $e_B \xi_B^{-1} e_B^T$ which, due to (3.61), receives a simple form[20]

$$e_B \xi_B^{-1} e_B^T = e Z_A^{-1/2} \xi_B^{-1} Z_A^{-1/2} e^T = e(\xi + \delta\xi)^{-1} e^T \,, \tag{3.67}$$

or, even more conveniently,

$$(e_B \xi_B^{-1} e_B^T)^{-1} = (e^T)^{-1} (\xi + \delta\xi) e^{-1} \,, \tag{3.68}$$

which holds to all orders in the perturbative expansion. Note that since the LHS of (3.68) is independent of μ so must be the RHS; this, in turn, correlates (in a relatively simple way) the evolution of e and ξ and the form of $\delta\xi$.

Besides that, Eq. (3.68) provides the key for the understanding of different approaches to the renormalization of such theories adopted in the literature.

[20] In what follows, we shall adopt a convention in which all real vectors are treated as column matrices; thus, their dot product can always be written as $x^T y$ while $x\, y^T$ corresponds to the outer (tensor) product.

Retaining e diagonal and keeping a non-trivial ξ at play

First, there is an option to work with e entirely diagonal but then Z_A must be diagonal too in order to retain this property throughout its running; c.f. Eqs. (3.64). However, in such a case Eq. (3.61) can be fulfilled (for a matrix-like $\delta\xi$) if and only if a non-trivial ξ is kept in the play (note that for a fixed ξ_B and a diagonal Z_A the structure $Z_A^{1/2}\xi_B Z_A^{1/2}$ therein clearly does not complement any initial and hypothetically constant matrix ξ to fulfil (3.61) for a specific $\delta\xi$ on its LHS for all μ).

Equivalently, without dynamically changing ξ one can not retain the RHS of Eq. (3.68) constant for diagonal e - note that there are 3 independent evolving combinations of running parameters in $\delta\xi$ while only 2 parameters are ready in e to compensate for their μ dependence!

In other words, one can naïvely take the renormalization-scale derivative of Eq. (3.68) and require that it vanishes; an attempt to solve such a linear system with respect to individual derivatives of e_1 and e_2 fails unless there is also an extra derivative of a component from ξ at one's disposal. Note that, in the current approach, this is also one way to derive the RGE for the ξ parameter in practice, cf. [140].

Getting rid of ξ at the expense of a matrix-like e and a matrix-like Z_A

Alternatively, one can arrange things in such a way to live without ξ altogether. This amounts to redefining first the bare gauge fields in Eq. (3.56) in order to absorb ξ_B therein, namely,

$$A_B \to \tilde{A}_B = \sqrt{\xi_B} A_B\,, \tag{3.69}$$

where the square root of the symmetric ξ_B matrix is defined in the standard manner. This changes the bare Lagrangian (3.56) into

$$\mathcal{L}_B = \overline{\psi}_B(\partial\!\!\!/ - m_B)\psi_B - iQ\overline{\psi}_B\tilde{e}_B\tilde{A\!\!\!/}_B\psi_B - \frac{1}{4}\tilde{F}_{B\mu\nu}\tilde{F}_B^{\mu\nu} \tag{3.70}$$

where the new set of gauge couplings

$$\tilde{e}_B = e_B\xi_B^{-1/2} \tag{3.71}$$

constitutes a non-diagonal matrix.

Performing the same redefinitions at the level of the renormalized Lagrangian like before one obtains formulae identical to Eqs. (3.57) and (3.58) with the following replacements:

$$e \to \tilde{e},\quad A \to \tilde{A},\quad F \to \tilde{F},\quad Z_A \to Z_{\tilde{A}},\quad \text{and, in particular,}\quad \xi \to 1,\quad \xi_B \to 1\,. \tag{3.72}$$

The rest of the analysis, i.e., Eqs. (3.59)-(3.68), follows along the same lines as above. Renaming (for optical reasons) $\delta\xi$ to $\delta Z_{\tilde{A}}$ in the analogue of the former Eq. (3.61) one ends up with

$$\delta\mathcal{L} = \overline{\psi}\delta Z_{\psi}\partial\!\!\!/\psi - \overline{\psi}\delta Z_m m\psi - i\overline{\psi}Q\delta Z_{\tilde{e}}\tilde{e}\tilde{A}\!\!\!/\psi - \frac{1}{4}\tilde{F}_{\mu\nu}\delta Z_{\tilde{A}}\tilde{F}^{\mu\nu} \tag{3.73}$$

where $\delta Z_{\tilde{e}}$ is actually identical with the former δZ_e because it is a multiplicative renormalization factor; in other words, the transformation

$$e \to \tilde{e} = e\xi^{-1/2} \tag{3.74}$$

acts homogeneously in formula (3.62) and thus the relevant multiplicative factors (Z_e in the lagrangian before reabsorption and $Z_{\tilde{e}}$ after that) are the same in both cases. This also ensures the validity of the Ward identity necessary to bring the analogue of formula (3.63) into the simple form

$$\tilde{e}_B = \tilde{e}Z_{\tilde{A}}^{-1/2}\,. \tag{3.75}$$

This scheme looks even better suited for practical purposes because the renormalization parameter $Z_{\tilde{A}}$ entering the equation above is directly connected to the gauge kinetic counterterm and, thus, easily accessible without any need to go for a non-linear formula like Eq. (3.68).

To recapitulate, we managed to get rid of the non-canonical form of the gauge kinetic term at the expense of a non-diagonal matrix of couplings $\tilde{e}$ (rather than just their pair as in e) and a matrix-like counterterm $\delta Z_{\tilde{A}}$.

3.5.4 Beta-functions in schemes with matrix gauge couplings

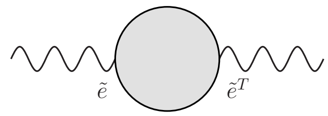
The simplicity of the latter scheme can be readily appreciated in a sample calculation of the gauge-couplings' evolution in QED-squared. The relevant analogue of formula (3.68), given (3.72) and (3.75), reads here

$$(\tilde{e}_B\tilde{e}_B^T)^{-1} = (\tilde{e}^T)^{-1}Z_{\tilde{A}}^{-1}\tilde{e}^T = (\tilde{e}^T)^{-1}(1+\delta Z_{\tilde{A}})\tilde{e}^{-1} = (\tilde{e}\tilde{e}^T)^{-1} + (\tilde{e}^T)^{-1}\delta Z_{\tilde{A}}\tilde{e}^{-1}\,. \tag{3.76}$$

Taking the logarithmic μ-derivative of both sides above with respect to the renormalization scale μ in momentum schemes (or, equivalently, with respect to the UV-pole structure in MS or $\overline{\text{MS}}$) one obtains

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}(\tilde{e}\tilde{e}^T)^{-1} = -(\tilde{e}^T)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}\log\mu}\delta Z_{\tilde{A}}\right)\tilde{e}^{-1} + \ldots \tag{3.77}$$

where the ellipsis stands for higher order terms. The structure of $\delta Z_{\tilde{A}}$ is simple to obtain from the diagrams of the type



,

which yield

$$\delta Z_{\tilde{A}}^{\mathrm{MS}} = \frac{1}{8\pi^2}\tilde{e}^T\gamma\tilde{e}\frac{1}{\varepsilon} \qquad \text{with} \qquad \gamma \equiv \frac{2}{3}QQ^T\,, \tag{3.78}$$

and, thus,

$$\frac{\mathrm{d}}{\mathrm{d}\log\mu}\delta Z_{\tilde{A}}^{\mathrm{MS}} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon^{-1}}\delta Z_{\tilde{A}}^{\mathrm{MS}} = \frac{1}{8\pi^2}\tilde{e}^T\gamma\tilde{e}\,. \tag{3.79}$$

Combining all this together one finally receives

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}(\tilde{e}\tilde{e}^T)^{-1} = -\frac{1}{8\pi^2}\gamma + \ldots\,. \tag{3.80}$$

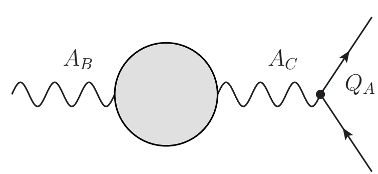
In what follows, we shall stick to the standard notation used in the literature [141,142] (i.e., replace $\tilde{e}$ by G) and work with $A \equiv GG^T/4\pi$ and $t = \frac{1}{2\pi}\log\mu/\mu_0$. In these coordinates, the evolution equation (3.79) can be recast in a particularly simple form

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1} = -\gamma + \ldots \tag{3.81}$$

which clearly resembles the situation encountered in the standard spinorial QED, see also (2.52).

Alternative derivation of the same result

Note that the same result can be obtained in a less fancy way by inspecting the μ-dependence of the three-body gauge amplitudes of the form[21]



, (3.82)

where, for the sake of simoplicity, we have been focusing on the renormalization of the g_{AB} coupling, i.e., the external gauge-field leg corresponds to A_B and the matter current

[21] Note that the Ward identities connecting the vertex corrections to the matter propagator counter-term ensuring the flavour-blindness of the charge renormalization are at work here in the same manner as in the ordinary QED so, as before, all that's needed are just the gauge propagator corrections.

is assumed to carry solely the $U(1)_A$ charge. The corresponding evolution equation for G defined as

$$G \equiv \begin{pmatrix} g_{AA} & g_{AB} \\ g_{BA} & g_{BB} \end{pmatrix} \tag{3.83}$$

can be written in a simple matrix form as

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} G = \frac{1}{(4\pi)^2} G(G^T \gamma G) \,. \tag{3.84}$$

Note that the configuration of various "building blocks" on the RHS above reflects the transformation properties of G, Q and A (with $Q^T = (Q_A, Q_B)$ and $A^T = (A_A, A_B)$) necessary for preserving the form of the covariant derivative (for ψ) $D \ni Q^T G A$, namely,

$$G \;\rightarrow\; O_1 G O_2^T \,, \tag{3.85}$$

$$Q \;\rightarrow\; O_1 Q \,, \tag{3.86}$$

$$A \;\rightarrow\; O_2 A \,, \tag{3.87}$$

where $O_{1,2}$ are independent real orthogonal matrices. Given this, the matrix shape of the RHS of Eq. (3.84) is practically enforced up to an overall numerical factor which, however, can be obtained trivially by matching to the known structure of the single-$U(1)$ case. Note that this is exactly the method used in the study [141] complementing the seminal results of Martin and Vaugn [143], see Sect. 5.4.

Finally, taking into account the symmetry properties of γ, Eq. (3.84) implies

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} (G G^T) = \frac{1}{8\pi^2} (G G^T) \gamma (G G^T) \,, \tag{3.88}$$

or, equivalently

$$(G G^T)^{-1} \left[\mu \frac{\mathrm{d}}{\mathrm{d}\mu} (G G^T) \right] (G G^T)^{-1} = \frac{1}{8\pi^2} \gamma \,. \tag{3.89}$$

which, using the general identity for regular matrix functions $A^{-1}(\frac{\mathrm{d}}{\mathrm{d}t} A) A^{-1} = -\frac{\mathrm{d}}{\mathrm{d}t} A^{-1}$ and the definitions above yields again the desired result (3.81).

3.5.5 One-loop matching in schemes with $U(1)$ mixing

So far, we have been discussing just the shapes of the renormalization group equations in different renormalization schemes traditionally adopted in theories with multiple $U(1)$ gauge factors. However, in practical calculations, these would be useless without the corresponding initial conditions obtained by matching the structure of the high and low-energy Lagrangians and, in particular, the coupling within.

In general, such a procedure closely resembles the recipe one follows even in the Standard Model case of matching the higher-energy $SU(2)_L \otimes U(1)_Y$ gauge structure to that of the effective low-energy $U(1)_Q$ one:

1. First, one should write down the high-scale (+) and low scale (−) covariant derivatives in terms of the relevant high- and low-scale theory charges (Q_+ and Q_-, respectively) and the corresponding gauge fields (A_+ and A_-), i.e. $D_+ \ni Q_+^T G_+ A_+$ and $D_- \ni Q_-^T G_- A_-$ (with all Lorentz factors suppressed for simplicity).

2. Second, one should express the A_+ fields in term of A_- and a suitable orthogonal matrix: $A_+ = OA_-$.

3. Subsequently, the identification of D_- and D_+ makes it possible to get all Q_-'s as linear combinations of Q_+'s (by looking at the A_-'s there) with coefficients corresponding to the entries of the specific "light" columns of the O matrix.

4. Next, one should combine this information with the model definitions of the Q_- charges in terms of the Q_+ ones (typically given as $Q_- = PQ_+$ where P is a suitable rectangular matrix, often with more columns than rows; in what follows we shall denote its individual rows by p^T where p's will be column vectors of coordinates of the individual operators of Q_- in the basis of operators in Q_+) and equate the coefficients of the individual Q_+ factors. This is justified by the fact that the action of the covariant derivatives must match on all fields in the model and there should be enough such fields in order to distinguish among all the charges.

5. Hence, one gets a set of equations for the "light columns" O_l of the total O matrix (i.e., those corresponding to the still massless gauge bosons) in the form $G_+ O_l = P^T G_-$.

6. The last step is to solve for the O_l matrix $O_l = G_+^{-1} P^T G_-$ and use its orthogonality $O_l^T O_l = 1$ (which, however, applies only from one side!) to get

$$1 = (G_-)^T P (G_+^{-1})^T G_+^{-1} P^T G_- = (G_-)^T P (G_+ G_+^T)^{-1} P^T G_- \,. \tag{3.90}$$

   The desired matching condition between G_- and G_+ then follows readily:

$$(G_- G_-^T)^{-1} = P (G_+ G_+^T)^{-1} P^T \,. \tag{3.91}$$

As an example, consider the simplest case of such a setting corresponding to the $U(1)_{Q_A} \otimes U(1)_{Q_B}$ gauge theory parametrised by a 2×2 matrix of gauge couplings (3.83) which gets broken down to a $U(1)_q$ scheme with the corresponding gauge coupling g. Suppose that the low-energy charge q is expressed in terms of the high-energy ones Q_A, Q_B as

$$q = p_A Q_A + p_B Q_B, \tag{3.92}$$

then $p^T = (p_A, p_B)$ so that $q = p^T Q$ with $Q = (Q_A, Q_B)^T$. Then the matching formula (3.91) yields

$$g^{-2} = p^T (GG^T)^{-1} p \,. \tag{3.93}$$

Matching in the "canonical" $U(1)_R \otimes U(1)_{B-L} \to U(1)_Y$ case

Let us apply the prescription above to the canonical example of the LR-symmetry breaking (cf. Sect. 3.1.1) passing through a $U(1)_R \otimes U(1)_{B-L}$ stage which is eventually broken to the SM hypercharge $U(1)_Y$. With $Y = \sqrt{\frac{3}{5}}R + \sqrt{\frac{2}{5}}X$ (note that Y here stands for the GUT-compatible hypercharge defined by (2.71) and X denotes the properly normalized $B-L$ charge connected to the "empirical" one via the $\sqrt{\frac{3}{8}}$ factor, see Sect. 3.1.2) one has

$$p^T = (\sqrt{\tfrac{3}{5}}, \sqrt{\tfrac{2}{5}}) \,, \tag{3.94}$$

and the procedure above yields

$$g_Y^{-2} = \frac{\frac{3}{5}\left(g_{XX}^2 + g_{XR}^2\right) + \frac{2}{5}\left(g_{RR}^2 + g_{RX}^2\right) - \frac{2}{5}\sqrt{6}\left(g_{RR}g_{XR} + g_{RX}g_{XX}\right)}{(g_{RR}g_{XX} - g_{RX}g_{XR})^2} \,. \tag{3.95}$$

Note that forgetting about off-diagonals, it trivially reduces to the naïve (and wrong) relation $\alpha_Y^{-1} = \frac{3}{5}\alpha_R^{-1} + \frac{2}{5}\alpha_X^{-1}$ which, however, often appears in the literature. Nevertheless, in most cases of physical interest the initial condition is such that the off-diagonal couplings are absent by definition (because $U(1)_R$ typically descends from a non-abelian factor) and, thus, the error due to their omission is relatively small with only negligible effects to the final results (unless the double-$U(1)$ gauge running stage is "long"[22]).

3.6 Towards the minimal renormalizable theory of perturbative baryon number violation

The renormalizable $SO(10)$ GUT studied in the preceding sections is arguably the minimal potentially realistic and calculable grand unified model and, as such, it is certainly worth a thorough scrutiny along the lines sketched above. However, it is obviously not a simple setting to deal with - the need for a detailed account for even small details of its quantum structure makes it utterly complex at the technical level, cf. [132]; hence, the number of its phenomenological analysis is very limited. The undertaking is further complicated by the presence of the humongous (anti-selfdual part of the) 5-index antisymmetric tensor whose main purpose (besides rank reduction)

[22] An extreme example of the size of the error committed by omitting the off-diagonal factors from the running can be found in the supplementary material of Sect. 5.4.

is to provide a renormalizable and tree-level Majorana mass for the RH neutrinos in the matter $\mathbf{16}_M$.

Assume for the moment that the $U(1)_{B-L}$ subgroup of $SO(10)$ is rather broken by the minimal scalar multiplet with this capacity, namely, the scalar $\mathbf{16}_S$. It is well known that, in such a case, one can still obtain a contribution to the RH neutrino masses from a $d = 5$ contractions of the type

$$\mathcal{L}^{(5)} \ni \frac{\lambda^{ij}}{\Lambda} \mathbf{16}^i_M \mathbf{16}^j_M \mathbf{16}^2_S \,, \tag{3.96}$$

where i, j are generation indices, λ is a 3×3 symmetric matrix of dimensionless couplings and Λ denotes an effective cut-off scale. In the asymmetric phase this operator yields $M_R \sim \lambda\langle\mathbf{16}_S\rangle^2/\Lambda$ which, assuming[23] $\Lambda \sim M_{Pl}$, gives M_R in the desired $10^{12-13}\,$GeV ballpark for $\langle\mathbf{16}_S\rangle \sim 10^{16}\,$GeV. At first glance, this is exactly what one needs in order to obtain the correct light neutrino masses via type-I seesaw; however, a closer inspection reveals at least two serious drawbacks of this scenario:

1. *One gets no information about the flavour structure of M_R.* Let us recall that this was one of the great benefits of having $\mathbf{126}_S$ rather than $\mathbf{16}_S$ which, indeed, made the minimal $SO(10)$ GUTs of Sections 3.3 and 3.4.4 (potentially) testable and, thus, so interesting.

2. *There are issues with the GUT-scale unification of the gauge couplings.* With $\langle\mathbf{16}_S\rangle \sim 10^{16}\,$GeV one typically ends up with a situation in which the $SO(10)$ symmetry is broken completely in the vicinity of $10^{16}\,$GeV and, hence, the theory looks entirely like the SM below M_G. Note, however, that this does not need to be a problem if the model is supersymmetric down to the TeV scale!

However, in the early days of GUTs, none of these two points was taken very seriously because nothing was known about neutrino masses in point 1., and there were no good electroweak-scale data to appreciate the severity of the tension in point 2.

3.6.1 Witten's loop in the $SO(10)$ GUTs

To this end, a very nice trick was pulled out by Witten [144] in 1980 which made it possible to generate the RH neutrino masses in the $SO(10)$ GUT context with $\mathbf{16}_S$ instead of $\mathbf{126}_S$ without the need to resort to the non-renormalizable mechanism above. The key consists in giving up the tree-level origin of M_R. Indeed, the two-loop diagram in Fig. 3.2 provides the necessary contraction of the ferminonic matter bilinear

[23] Note that $\Lambda \sim M_{Pl}$ is the most natural assumption here as M_{Pl} is the only scale of BSM physics which is almost universally accepted to exist.

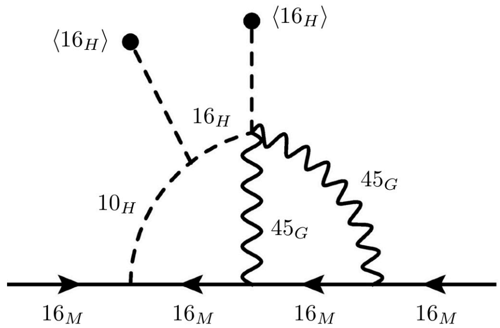


Figure 3.2: A sample two-loop Feynman diagram contributing to the Majorana masses for the RH neutrinos (residing in $\mathbf{16}_M$) in $SO(10)$ GUTs. Note that the graph corresponds to just one of many two-loop contributions in the broken-phase perturbative expansion with VEVs kept in the interaction part of the Hamiltonian.

$\mathbf{16}_M \otimes \mathbf{16}_M$ to a pair of $\mathbf{16}_S$ VEVs (arranged in a way to resemble an effective 5-index antisymmetric tensor structure) through a very specific contraction of the internal lines of the relevant Feynman graph. The consistency of this picture can be verified readily:

- There are, indeed, two units of $B-L$ carried into the vacuum by the VEVs of $\mathbf{16}_S$ in Fig. 3.2.
- The algebraic structure of the loop propagators ($\mathbf{10}_S \otimes \mathbf{45}_G \otimes \mathbf{45}_G$) attached to the fermionic line does admit a fully antisymmetric contraction and, thus, can mimic the 5-index antisymmetric tensor structure of $\mathbf{126}$.
- The matrix of Yukawa coupling of $\mathbf{10}_S$ is fully symmetric in the flavour space and so is also the M_R generated through the graph in Fig. 3.2.

This looks like a perfect alternative to both the tree-level M_R generation entertained in the models with $\mathbf{126}_S$ and the non-renormalizable mechanism one would naïvely have to resort to in scenarios with $\mathbf{16}_S$ instead.

Witten's mechanism $SO(10)$ model building issues

On the practical side, however, the beautiful Witten's mechanism of the preceding section has never been implemented in a fully compelling unified setting. There are two main reasons having to do with point 2. in the second paragraph of Sect. 3.6.2, namely, the difficulty to obtain a potentially viable gauge unification pattern:

- In the traditional TeV-scale SUSY scenarios in which $\langle \mathbf{16}_S \rangle \sim 10^{16}$ GeV is not a problem for gauge unification the scale of M_R turns out to be strongly suppressed with respect to the natural expectation of $M_R \sim \lambda\langle \mathbf{16}_S\rangle^2/16\pi^2 M_G \sim M_G/16\pi^2$ by the $m_{\rm soft}/M_G$ extra factor due to the SUSY non-renormalization theorems [145, 146].

- In non-SUSY theories the gauge running with $\langle \mathbf{16}_S\rangle \sim 10^{16}\,$GeV cries for an intermediate scale. However, since the residual symmetry left intact by $\langle \mathbf{16}_S\rangle$ is $SU(5)$ there is not much of a room for such a scale to emerge from the subsequent gauge symmetry breaking and, hence, the only potentially viable scenarios are those featuring extra fine-tunning(s). In the extreme case, one can consider the split-SUSY [147] variant of the Witten's scenario [148, 149] in which the squarks and sleptons are maximally diverted from the TeV scale where the gauginos and higgsinos ensure the MSSM-like gauge unification.

3.6.2 Witten's loop in the flipped $SU(5)$ unification

In what was written above there is a clear indication that a remedy may eventually come from relaxing the stringent grand unification constraints imposed (among other sectors) on the gauge couplings of the model. This, however, calls for a "non-grand" unified scenarios, i.e., those based on non-simple symmetry groups. At the same time, it would be very welcome if the predictive power of such models was not entirely ruined concerning the effects of our main interest, i.e., perturbative baryon and lepton number violation.

Remarkably, there is indeed a very interesting gauge model just half way between the fragility of the $SO(10)$ GUTs and baryon number triviality of its simple descendants such as Pati-Salam or LR-symmetric models discussed in Sects. 3.1.2 and 3.1.1. It is based on the maximal $SU(5)\otimes U(1)$ subgroup of $SO(10)$ and, as such, it stands out of the usual mantra of the need for the LR-symmetrisation of the gauge symmetry in presence of the RH neutrinos. It is not similar to the $SU(5)$ GUTs of Sect. 2.6.1 either as the SM hypercharge is not fully contained in the $SU(5)$ gauge factor.

Flipped $SU(5)$ unification overview

Let's just recapitulate the salient features of the traditional approach to the "marriage" between the $SU(5)$ and $U(1)$ gauge symmetries (the latter to be from now on called $U(1)_Z$): Assigning a unit of the Z charge to the $\mathbf{10}$ of $SU(5)$ the following set of charges is anomaly free:

$$(\mathbf{10},+1)\oplus(\overline{\mathbf{5}},-3)\oplus(\mathbf{1},+5)\,. \tag{3.97}$$

Hence, 16 matter fields of each of the SM generations + 3 RH neutrinos can be accommodated in such a (reducible) representation of $SU(5)\otimes U(1)$. However, the trivial matter embedding along the lines of the minimal $SU(5)$ à la Georgi and Glashow discussed in Sect. 2.6.1, namely, L and d^c in $(\overline{\mathbf{5}}, -3)$, Q, u^c and e^c in $(\mathbf{10}, +1)$ and N^c in $(\mathbf{1}, +5)$ with the hypercharge generator identified with (a properly normalized) T_{24} of $SU(5)$, is not what we are after here.

Interestingly, there is a second option [150,151] corresponding to swapping u^c with d^c and ν^c with e^c that, however, requires a non-trivial ("flipped") embedding of Y into the full $SU(5) \otimes U(1)_Z$, namely

$$Y = \frac{1}{5}(Z - T_{24}). \tag{3.98}$$

This also means that one can use $\mathbf{10}_S = (\mathbf{10}, +1)$ to break the symmetry down to the SM[24] instead of the larger adjoint $\mathbf{24}$ of the standard $SU(5)$ and, thus, save a number of scalar degrees of freedom[25].

The "flipped" hypercharge embedding also induces important changes in the baryon number violation phenomenology. Indeed, the leptoquark degrees of freedom in the gauge sector spanning over $(\mathbf{24}, 0) \oplus (\mathbf{1}, 0)$ transform as $(\mathbf{3}, \mathbf{2}, +\frac{1}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, -\frac{1}{6})$ rather than $(\mathbf{3}, \mathbf{2}, -\frac{5}{6}) \oplus (\overline{\mathbf{3}}, \mathbf{2}, +\frac{5}{6})$ of the Georgi-Glashow model and, hence, different pattern of $d = 6$ BLNV operators is generated at the SM effective theory level.

The devil, as usual, is in detail. Unlike in the "standard" $SU(5)$ where the RH neutrino mass underpinning the type-I seesaw is trivially introduced as a direct mass term in the relevant $\mathbf{1}_M \mathbf{1}_M$ matter singlet bilinear, one needs a large 50-dimensional two-index symmetric tensor transforming as $(\mathbf{50}, -2)$ to do the same in the current scenario, thus enlarging the so far limited number of degrees of freedom by about a factor of 4. Hence, in spite of its overall simplicity, the flipped $SU(5)$ framework does not seem to provide more insight into the flavour aspects of lepton number violation than its "standard" $SU(5)$ counterpart.

The minimal renormalizable flipped $SU(5)$ unification

The important observation made[26] in [154] was that the drawbacks of the original Witten's loop mechanism in the $SO(10)$ context and the baroqueness of the minimal

[24] Note that the simplified symmetry breaking mechanism based on $\mathbf{10}_S = (\mathbf{10}, +1)$ is in the core of the "missing partner" doublet-triplet splitting mechanism that brought the flipped $SU(5)$ scenario a lot of attention in 1980's.

[25] The associated simplification of the scalar sector necessary for the proper symmetry breaking down to the SM is one of the typical benefits of the "flipped" scenarios; to this end, one can quote the situation in the flipped SUSY $SO(10)$ scenarios (see, e.g., [152]) in which the notoriously cumbersome $\mathbf{210} \oplus \mathbf{126} \oplus \overline{\mathbf{126}}$ minimal Higgs sector (cf. Sect. 3.3.1) or its $\mathbf{45} \oplus \mathbf{54} \oplus \mathbf{126} \oplus \overline{\mathbf{126}}$ variant can be replaced with just $\mathbf{45} \oplus 2 \times (\mathbf{16} \oplus \overline{\mathbf{16}})$.

[26] For the sake of completeness let us note that a similar scheme was considered in the string theory context already in 1991 in [153].

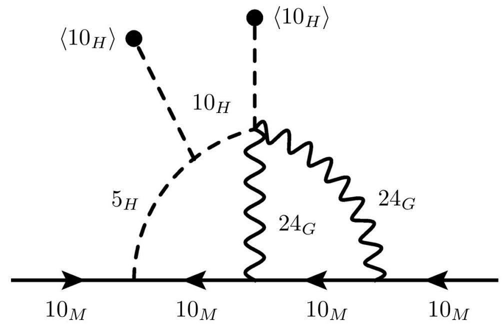


Figure 3.3: A variant of the Witten's two-loop diagram in Fig. 3.2 contributing to the right-handed neutrino masses in the minimal flipped $SU(5)$ setting.

potentially realistic renormalizable flipped $SU(5)$ model sketched above can cure each other. Indeed, the Feynman diagram in Fig. 3.3 can be viewed as that of Witten in Fig. 3.2 minimally adopted to the flipped $SU(5)$ context. The main point is that the VEV of the (Hermitean conjugate of) $(\mathbf{10},+1)$ sticking out of the graph can be as large as M_G with no tension in the gauge running because the g_Z coupling associated to the $U(1)_Z$ factor is not required to unify with that of the $SU(5)$ part where only g_3 and g_2 of the SM are fully contained. Note also that, algebraically, the effect of **50** is mimicked by the tensor product of[27] $\mathbf{5}\otimes\mathbf{24}\otimes\mathbf{24}$ though perhaps not as clearly as it was for the **126** within $\mathbf{10}\otimes\mathbf{45}\otimes\mathbf{45}$ of $SO(10)$.

3.6.3 The minimal renormalizable theory of perturbative B violation

Hence, one arrives at a very attractive scenario which shares the nice features of both worlds, namely, the BLNV sector's predictive power of the minimal $SO(10)$ GUTs and the relative simplicity of the $SU(5)$ settings, with the extra benefit of providing a potentially realistic framework for the implementation of the beautiful Witten's idea.

In the minimal version the high-scale spectrum and the symmetry breaking pattern of the model is encoded in the scalar potential

$$\begin{aligned} V &= \frac{1}{2}m_{10}^2\mathrm{Tr}(\mathbf{10}_S^\dagger\mathbf{10}_S) + m_5^2\mathbf{5}_S^\dagger\mathbf{5}_S + \frac{1}{8}(\mu\varepsilon_{ijklm}\mathbf{10}_S^{ij}\mathbf{10}_S^{kl}\mathbf{5}_S^m + h.c.) \\ &+ \frac{1}{4}\lambda_1[\mathrm{Tr}(\mathbf{10}_S^\dagger\mathbf{10}_S)]^2 + \frac{1}{4}\lambda_2\mathrm{Tr}(\mathbf{10}_S^\dagger\mathbf{10}_S\mathbf{10}_S^\dagger\mathbf{10}_S) + \lambda_3(\mathbf{5}_S^\dagger\mathbf{5}_S)^2 \\ &+ \frac{1}{2}\lambda_4\mathrm{Tr}(\mathbf{10}_S^\dagger\mathbf{10}_S)(\mathbf{5}_S^\dagger\mathbf{5}_S) + \lambda_5\mathbf{5}_S^\dagger\mathbf{10}_S\mathbf{10}_S^\dagger\mathbf{5}_S\,, \end{aligned} \tag{3.99}$$

[27] Note that the quantum numbers of the 5-dimentional scalar generating the Yukawa contraction of the type $\mathbf{10}_M\mathbf{10}_M\mathbf{5}_S$ are trivially $(\mathbf{5},-2)$.

where $\lambda_{1,..,5}$ are real numerical couplings and μ provides the necessary mixing between $\mathbf{10}_S$ and $\mathbf{5}_S$, cf. Fig. 3.3. In the broken phase (triggered by a unification-scale VEV of the SM singlet component of $\mathbf{10}_S$ accompanied by the electroweak VEV of $\mathbf{5}_S$) the scalar spectrum of the theory comprises of 16 real Goldstone modes (corresponding to $25 - 9 = 16$ massive vector bosons of $SU(5) \otimes U(1)$ broken down to the low-scale $SU(3)_c \otimes U(1)_Q$), a light Higgs boson, a heavy SM singlet S from $\mathbf{10}_S$ and a pair of heavy SM colour-triplet leptoquarks $\Delta_{1,2}$ admixed (via μ) from the two $(\mathbf{3}, \mathbf{1}, -\frac{1}{3})$ components in $\mathbf{5}_S \oplus \mathbf{10}_S$. Besides that, the physical heavy spectrum contains a vector leptoquark X^μ with the SM quantum numbers $(\mathbf{3}, \mathbf{2}, +\frac{1}{6}) + h.c.$ and a heavy SM singlet. Note also that the unified symmetry breaking occurs in one-step, i.e., the $SU(5) \otimes U(1)$ is broken directly into the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ of the Standard Model and one can expect the characteristic scale of the heavy degrees of freedom to be in the 10^{16} GeV ballpark. For further details, an interested reader is kindly deferred to the original work [154].

Phenomenology of the flipped $SU(5)$ unification with Witten's loop

In what follows, the central quantity of interest will be namely the Yukawa Lagrangian which, in the minimal case, reads

$$\mathcal{L} \ni Y_{10} \mathbf{10}_M^{+1} \mathbf{10}_M^{+1} \mathbf{5}_S^{-2} + Y_5 \mathbf{10}_M^{+1} \overline{\mathbf{5}}_M^{-3} (\mathbf{5}_S^{-2})^* + Y_1 \overline{\mathbf{5}}_M^{-3} \mathbf{1}_M^{+5} \mathbf{5}_S^{-2} + h.c. \,, \tag{3.100}$$

where Y_{10}, Y_5 and Y_1 are (complex) 3×3 matrices of Yukawa couplings (Y_{10} is symmetric). It yields the following expressions and correlations for the (high-scale) effective matter mass/Yukawa matrices:

$$\begin{aligned} M^d = (M^d)^T &\propto Y_{10} \,, \\ M^u = (M_D^\nu)^T &\propto Y_5 \,, \\ M^e &\propto Y_1 \,, \\ M_M^\nu &= 0 \,. \end{aligned} \tag{3.101}$$

Note that, as expected, no Majorana mass matrix for the RH neutrinos is generated at the tree level. Moreover, it is the down-quark mass matrix that turns out to be symmetric rather than M_u in the standard $SU(5)$ case, cf. (2.77). Instead, M_u is tightly correlated to the Dirac neutrino mass matrix; this is actually the seed of the model's significant predictive power in the flavour sector.

As anticipated, the Majorana RH neutrino mass is generated at the quantum level by the loops of the kind depicted in Fig. 3.3. A short inspection reveals that its flavour structure must be driven by Y_{10} as it is the only Yukawa therein and its overall scale should be governed by two powers of $\langle \mathbf{10}_S \rangle$. Moreover, the relevant expression should vanish for $\mu \to 0$; hence, on purely dimensional grounds, one expects

$$M_M^\nu \propto \frac{1}{16\pi^2} Y_{10} g^4 \mu \frac{\langle \mathbf{10}_S \rangle^2}{M_X^2}\,, \tag{3.102}$$

where M_X is the scale of the heavy gauge boson mass.

With this at hand, the type-I seesaw formula combined with the non-tachyonicity conditions for the scalar spectrum (for details see Sect. 5.5) yields

$$D_u U_\nu^\dagger (m^\nu_{\text{diag.}})^{-1} U_\nu^* D_u \leq \frac{\alpha_U}{64\pi^4} \sqrt{\lambda_2 \lambda_5} |Y_{10}| V_G F\,, \tag{3.103}$$

where D_u is the diagonal form of the up-quark mass matrix, U_ν is the unitary transformation diagonalising the light neutrino masses m^ν (i.e., $m^\nu = U_\nu^T m^\nu_{\text{diag.}} U_\nu$), α_U is the $SU(3) \otimes SU(2)$-unification-scale value of the associated "generalised fine structure constant" of the model and F is an $\mathcal{O}(1)$ factor calculable from the relevant Feynman graphs, see next section.

Note that for a fixed light neutrino spectrum (conveniently parametrised by the mass of the lightest neutrino m_1) and with the assumption of perturbativity of the $\lambda_{2,5}$ and Y_{10} couplings[28] in formula (3.103) one obtains *a strong constraint on the possible shapes of* U_ν. For instance, for small enough m_1 large 1–3 angle in U_ν is clearly forbidden as it would propagate the big 11 element of the $(m^\nu_{\text{diag.}})^{-1}$ matrix to the 33-element of $U_\nu^\dagger (m^\nu_{\text{diag.}})^{-1} U_\nu$ which would, subsequently, pick up the pair of the top quark masses in (3.103) and, hence, violate the desired inequality there by orders of magnitude. At the same time, *heavier light neutrino spectrum will be preferred* as it would also have the tendency to alleviate this issue. Both these observation are extremely welcome as they provide non-trivial constraints on both the lepton and baryon number violating phenomena in this setting.

Lepton number violation phenomenology

As for L violation, some first (unification-scale dependent) lower limits on the absolute scale of the light neutrino Majorana masses were derived in [155], with a clear preference of relatively heavy light neutrino spectra. Besides that, scans through the parameter space (including CP phases) for the spots which do support thermal leptogenesis (see Sect. 2.1.3) impose further limits for the absolute light neutrino scale. This is work in progress, to appear soon.

[28] Perturbativity constraints are generally very tricky. For the sake of simplicity, all that we shall assume here is that none of the dimensionless couplings involved exceeds 4π.

Baryon number violation phenomenology

Concerning perturbative B violation, U_ν is a structure governing practically all the leading order gauge-leptoquark contributions to the two-body proton decay amplitudes in the current scenario. This is due to the flavour structure of the $d = 6$ BLNV operator $\overline{d^c}Q\overline{u^c}L$ (a Fierz-transform of O_1 in Table 2.1) which, barring its strongly suppressed companion involving heavy RH neutrinos, is the only relevant piece of $\mathcal{L}_{\rm int}$ here

$$\mathcal{L}_{\rm int} \ni (U_{d^c})_{a1}(U_e^\dagger)_{1b}\overline{u^c_{(1)}}\gamma_\mu u_{(1)}\overline{d^c_{(a)}}\gamma^\mu e_{(b)} + (U_{d^c}U_d^\dagger)_{ba}(U_\nu^\dagger)_{1c}\overline{u^c_{(1)}}\gamma_\mu d_{(a)}\overline{d^c_{(b)}}\gamma^\mu \nu_{(c)} \,. \quad (3.104)$$

In (3.104) the bracketed indices label generations, U_f and U_{f^c} for $f = u, d, e$ are defined as $M_f = U_f^T D_f U_{f^c}$ and all is written in the basis in which the top-quark mass matrix is diagonal. From here it is clear that the amplitudes for the p-decay neutral-meson final states factorize $(U_e^\dagger)_{1b}$ (where b stands for the flavour of the final state lepton) and $U_{d^c} = U_d$ following from (3.101) is nothing but the CKM matrix here. Thus, the ratios of the same-lepton partial decay widths are fully calculable in the minimal flipped $SU(5)$ scenario! Moreover, for $U_{d^c} = U_d$ the flavour structure in the second term above, i.e., the piece governing the decays into charged mesons+neutrinos, depends only on U_ν. Hence, *the minimal flipped $SU(5)$ à la Witten represents a framework in which the gauge contributions to the two-body proton decay amplitudes are very strongly correlated in all channels.* As such, the model may be viewed as a genuine theory of perturbative baryon number violation.

3.6.4 Calculating Witten's loop in the flipped $SU(5)$ context

Needless to say, a decisive numerical analysis of this attractive scenario[29] requires a detailed calculation of the F-factor in formula (3.103). This was recently performed in the study [155] (enclosed as a Supplementary material in Sect. 5.5). Perhaps the most interesting aspect of the calculation worth elaborating on here is the IR- and UV-divergence structure of the relevant Feynman graphs which can be conveniently written in the massive perturbation theory employing the unitary gauge, see Fig. 3.4. There are few observations one can make right away:

- The sum of the two two-loop topologies depicted in Fig. 3.4 should be UV-finite. This is clear as there is no tree-level counterterm in the model that can tame any UV divergence.

[29] It is worth noting that in order to make the model potentially realistic a second copy of the scalar $\overline{\mathbf{5}}_S$ has to be added. The main reason is that, in the minimal setting, Y_{10} in (3.103) is overly constrained by Eq. (3.101) and a realistic hierarchy of light neutrino masses can not be attained, cf. [154]. Remarkably enough, even with the second copy of $\overline{\mathbf{5}}_S$ all the features underpinning the predictivity of the model are preserved: M^d remains symmetric, M^u is still equal to M_D^ν and M^e is as unconstrained as before.

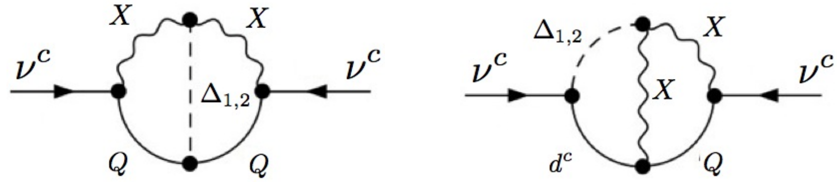


Figure 3.4: The complete set of unitary-gauge Feynman graphs generating the RH neutrino masses in the minimal flipped $SU(5)$ setting in the "massive" perturbative expansion (i.e., with the VEVs absorbed into the massive propagators of all the relevant fields). Note that this approach minimises the total number of graphs for the price of their higher complexity.

- There should be no UV sub-divergences either that would require insertion of the first-order counterterms anywhere. The point is that none of the graphs with the trilinear local counterterms (replacing any of the loops involved) exists because the first non-trivial contraction corresponding to the given configuration of the external lines emerges only at two loops.
- One should be free to set the light matter fermion masses in Fig. 3.4 to zero as these should play no role in such loops and all the IR divergences potentially emerging in such a limit should eventually disappear.

This is, indeed, what eventually happens when the calculations are performed in detail. For each of the two Δ-scalars the UV-divergent structure of the diagram on the left in Fig. 3.4 reads (in $d = 4 - 2\varepsilon$)

$$\Sigma_1^{\mathrm{UV}}(0) = -\frac{1}{(4\pi)^4}\left[\frac{3}{2\varepsilon} - \frac{m_\Delta^4}{2m_X^4}\left(\frac{1}{2\varepsilon^2} + \frac{3}{2\varepsilon} - \frac{1}{\varepsilon}\log\frac{m_\Delta^2}{M^2}\right)\right], \tag{3.105}$$

(with M denoting the regularisation scale) while that of the graph on the right turns out to be

$$\Sigma_2^{\mathrm{UV}}(0) = \frac{1}{(4\pi)^4}\left[\frac{3}{4\varepsilon} + \frac{m_\Delta^4}{4m_X^4}\left(\frac{1}{2\varepsilon^2} + \frac{3}{2\varepsilon} - \frac{1}{\varepsilon}\log\frac{m_\Delta^2}{M^2}\right)\right]. \tag{3.106}$$

Since, however, these contributions sum with a relative factor of 2 in the total integral

$$I_\Delta \equiv -(4\pi)^4\left[\Sigma_1(0) + 2\Sigma_2(0)\right] \tag{3.107}$$

the overall cancellation of the UV divergences follows for each cluster of graphs. As one can expect, the finite part of I_Δ which should be only a function of $s \equiv m_\Delta^2/m_X^2$

behaves regularly[30] even in the asymptotic limits with s large or small, namely

$$
\begin{aligned}
I_\Delta &\sim -3 + O(s^{-1}\log^2 s) \qquad \text{for } s \to \infty\,, \\
I_\Delta &\sim 3 + s\left(3\log s + \pi^2 - \frac{15}{2}\right) + O(s^2\log^2 s) \qquad \text{for } s \to 0\,.
\end{aligned}
$$

In conclusion, the absolute value of the F factor in Eq. (3.103) which is just a weighted sum of I_Δ's for the two $\Delta_{1,2}$ scalars, is bounded from above. This, in turn, provides a clear rationale for the phenomenological expectations spelled out in the previous section.

[30]Interestingly, I_Δ is even a monotonic function of s, see Sect. 5.5.

Chapter 4

Conclusions and outlook

After almost 50 years since the discovery of an internally consistent framework for description of the elementary constituents of matter and their interactions – the Standard Model – the Higgs boson, the last missing piece of the underlying structure, has eventually been discovered. As beautiful and soothing the picture became there is still a number of indications that this has not been the final word the elementary particle physics had to say about the world around. Indeed, in the years to come our confidence is likely to be challenged by:

- Clear experimental signals of physics that the SM can not account for; among the most prominent of these are the neutrino flavour oscillation effects.

- Theoretical issues related to the structural aspects of the SM such as flavour and other things; for instance, we have no clue on what is behind the peculiar matter generation pattern, why there is so little CP violation in the strong interactions etc. Remarkably enough, we do not know even such a basic thing as whether the very SM vacuum is stable or not.

Even more trouble is likely to be encountered if the Standard Model is to be eventually married with gravity:

- On the observational side, the SM provides no clue for the nature and origin of the peculiar matter-like but invisible gravitating component of the Universe's energy density budget (dark matter), let alone its negative-pressure counterpart (often called dark energy) which, together, overwhelm the part accounted for by the SM by almost a factor of 20. We do not understand the number of baryons left behind the initial annihilation inferno if the initial conditions were B-symmetric; if the were not, we do not know why.

- As far as the theory is concerned, there is no realistic (even potentially) and calculable framework that can consistently account for the intrinsically quantum nature of the SM together with the general covariance of the Einstein gravity. We have no clue about what is behind the vast disparity between the electroweak and Planck scales; actually, we do not even know whether there is anything to be understood there at the first place.

The discovery of neutrino masses with all its basic implications as described in this thesis paves a particularly interesting (yet pretty conservative) avenue for a further systematic exploration of the *Terra Incognita* slowly emerging on the horizons of namely the intensity frontier experimental activities. In this struggle, the effects of baryon and lepton number violation are likely to play a prominent role either as the very objectives of the research or at least as irreducible elements of the undertaking.

From this perspective, the candidate's research activities on the frontier of perhaps the most natural theoretical approach to the conundrum – the BSM theories with extended gauge symmetries – represent a valuable and relevant contribution to the current understanding of especially the quantum structure of unified gauge models. Among the most important of these one should perhaps mention the series of works in which the paradigmatic minimal SUSY GUT model has been decisively refuted, the original idea that the tachyonic vacuum instabilities of the minimal $SO(10)$ GUT may be lifted by quantum effects or the full logical completion of the two-loop renormalization-group programme set out by Martin and Vaughn [4] for softly broken supersymmetric theories has been achieved.

It is always the time that eventually decides on whether anything of this would be relevant for the future of the beautiful subject at stakes. At the moment, one may only hope that the quest for the unified description of matter and its interactions, underlying all efforts elaborated on in this thesis, will be one day rewarded by a clear signal of baryon and/or lepton number violation. If we were ready and lucky this may even be the case within the upcoming generation of experimental facilities.

Chapter 5

Supplementary material

Here we shall present and comment upon six selected candidate's publications reflecting the extent of his contribution to the subject of Grand unified theories and their quantum structure.

5.1 The novel $SO(10)$ seesaw mechanism

M. Malinský, J. Romao, J.W.F. Valle, Physical Review Letters 95, (2005) 161801, DOI: 10.1103/PhysRevLett.95.161801

This article is enclosed as the candidate's most cited work (having earned about 280 citations within the `inspirehep.net` database as of January 2020). It elaborates on the unexpected possibility to disentangle the seesaw scale from the scale of the $SU(2)_R \otimes U(1)_{B-L}$ breaking in the SUSY GUT context exploiting the interesting features of the inverse seesaw scheme put forward in [156, 157]. The key to a viable and internally consistent setting is the minimality of the $SO(10)$ symmetry breaking pattern passing through an intermediate $U(1)_R \otimes U(1)_{B-L}$ symmetry stage (see Section 3.5 of this thesis for details) which is, subsequently, broken by a *single scalar field* that does not contain any SM-charged components. Thus, the $B-L$ breaking scale is not only decoupled from the $SU(2)_R$ one but it becomes essentially free from the gauge unification perspective. At the same time, the $B-L$ breaking scale drops from the seesaw formula (at the leading order) due to an interplay between the size of the induced $SU(2)_L \otimes U(1)_Y$-breaking VEV of an additional $SU(2)_L$ doublet in the scalar **16** of $SO(10)$ and the shape of the relevant linear-seesaw contribution to the light neutrino masses.

Besides having written the major part of the manuscript the candidate contributed to the study by two central ingredients, namely, by noticing the presence of the induced VEV of the $SU(2)_L$ doublet in the scalar **16** enforced by the SUSY F-flatness conditions and by providing the entire tedium of the renormalization group analysis.

5.2 The minimal renormalizable SUSY $SO(10)$ no-go

S. Bertolini, M. Malinský, T. Schwetz, Physical Review D 73 (2006) 115012, DOI: 10.1103/PhysRevD.73.115012

This article is enclosed as the candidate's first game-changing contribution to the subject of grand unified theories.

As described in Sect. 3.3.1, at the beginning of the first decade of this century the minimal renormalizable SUSY $SO(10)$ model [105–111] was widely recognised as one of the most promising potentially realistic supersymmetric grand unified theories on the market. Besides its unprecedented simplicity owing to the low number of Higgs-type chiral supermultiplets it was known for providing interesting predictions for the flavour structure of its low-energy effective descendants (MSSM in most cases), namely, i) an automatic conservation of R-parity, ii) a rationale for the near-maximality of the atmospheric mixing in the lepton sector (attributed to the phenomenological high-energy convergence of the tau-lepton and bottom-quark Yukawa couplings) and iii) a lower limit for the reactor mixing angle (close to its recently measured value).

Originally, the latter two observations were made under a simplifying assumption of the dominance of the type-II seesaw contribution to the light neutrino masses which, however, could not be justified without a particularly tedious calculation of the induced VEVs of the $SU(2)_L$ triplets of $\mathbf{126}\oplus\overline{\mathbf{126}}$ (as a by-product of a thorough investigation of the high-energy spectrum of the theory). The situation, luckily, changed with the publication of the studies [111, 158] which provided the missing information. Thus, all of a sudden, it was possible to perform a thorough scrutiny of the minimal model including, in particular, a complete χ^2-analysis of the matter fermion mass, mixing and CP patterns and a comprehensive study of the relevant gauge unification chains (revealing, in particular, the positions of the seesaw scale), all that in connection with the detailed understanding of the heavy part of the model's spectrum.

With all this at hand, it was decisevely concluded (cf. Sect. 3.3.2 of this thesis) that no point in (the perturbative part of) the parameter space exists that would support all the imposed phenomenological constraints. The importance of this result has been widely acknowledged by the HEP community, earning to the article about 140 citations (in the `inspirehep.net` database as of 1/2020).

Besides having written a significant part of the manuscript the candidate's contribution to the study consisted of providing all the necessary information related to the shape of the high-energy spectrum and the complete renormalization group running analysis in the cases of main interest.

5.3 The minimal $SO(10)$ GUT resurrection

S. Bertolini, L. di Luzio, M. Malinský, Physical Review D 81, (2010) 035015, DOI: 10.1103/PhysRevD.81.035015

In this article the candidate's hypothesis that quantum corrections can resolve the notorious tachyonicity issues of the minimal adjoint $SO(10)$ Higgs model (see Sect. 3.4.3) has been worked out. This textbook-level result has brought back to life an entire class of minimal grand unified models which were considered non-realistic for almost quarter of a century.

As described in Sect. 3.4 of this thesis a serious issue with the minimal implementation of the Higgs mechanism in the realm of the $SO(10)$ gauge theories has been revealed at the beginning of 1980's. In sharp contrast to the naïve expectation based on the contemporary studies of other models of spontaneous symmetry breaking the minimal $SO(10)$ Higgs scheme has been shown to possess no stable classical vacuum state that could support potentially realistic symmetry breaking chains.

The merit of the enclosed study consists in the observation that the surprising formal simplicity of the mathematical expressions for certain scalar masses underpinning this issue can be attributed to the pseudo-Goldstone nature of these fields. Several sets of spurion couplings defining various accidental global symmetries of the classical potential have been identified and it was argued that these symmetries should be explicitly broken by higher-order corrections, thus providing a room for the radiative vacuum stabilisation mechanism à la Coleman and Weinberg. A representative set of leading quantum corrections generated by the gauge fields has been calculated and a potentially realistic locally stable quantum vacuum configuration has been found.

These ground-breaking results sparked a renaissance of the minimal $SO(10)$ GUT with many subsequent works following the logic outlined in the article. The paper and its follow-up studies co-authored by the candidate have, to date, earned in total over 160 citations in the `inspirehep.net` database.

Besides formulating the central idea of understanding the critical states in terms of the Goldstone theorem language the candidate has worked out many details of the calculation including, e.g., a convenient expansion for the second derivatives of the effective potential, identification of the mechanism of cancellation of spurious IR divergences in the quantum-level pole masses etc. He has also contributed by writing a significant part of the manuscript.

5.4 Two-loop SUSY RGEs with U(1) mixing

R.M. Fonseca, M. Malinský, W. Porod, F. Staub, Nuclear Physics B 854 (2012) 28, DOI: 10.1016/j.nuclphysb.2011.08.017

The main scope of this article was to fill an unpleasant "hole" in the literature on the renormalization group techniques in supersymmetric gauge field theories in which the corresponding gauge symmetry contains more than a single Abelian group factor, cf. Sect. 3.5 of this thesis. By doing that, the programme set out by the seminal works of Y. Yamada and S. Martin and M. Vaughn in the early 1990's has been finally completed and a self-contained toolbox facilitating the construction of two-loop beta-functions for any softly broken $\mathcal{N} = 1$ renormalizable supersymmetric gauge field theory has been provided.

Perhaps the most interesting aspect of the study consists in the method by which the desired results have been obtained. Indeed, the obvious brute-force approach which, on one hand, was guaranteed to yield the result but, on the other hand, represented an enormous tedium with only limited cross-checking options, was almost completely avoided by exploiting the symmetries of the covariant derivative, the basic building element of the theory, cf. Sect. 3.5.4. With this technique, one can attempt to construct the two-loop beta-functions from the covariant blocks including gauge couplings organised into matrices and the other structures (group invariants, soft terms etc.) by a mere matching to the known results for the theories with a single $U(1)$ gauge factor. Remarkably, this method admitted to resolve almost all ambiguities up to a very limited set of residual Feynman graphs' topologies which still required the old-fashioned treatment. Besides providing the final completion of the "vocabulary" of Martin and Vaughn's Physical Review D 50 (1994) 2282 the article elaborates also on the issue of the matching for the gauge couplings and gaugino masses and illustrates the basic principles on a set of specific examples.

The importance of these results has been widely recognised by the HEP community; this can be clearly demonstrated on the fact that the study itself and its sequel Physics Letters B726, 882 (2013) have, to date, attracted over 100 citations in the `inspirehep.net` database.

The candidate's main contribution to the study was the crucial observation that the symmetry properties of the gauge-coupling- and charge matrices, together with matching to the existing single-$U(1)$ results, provide a very powerful method for circumventing the swampland of the ab-initio approach. Besides that, he contributed by the "Methods" of Sect. 2, examples of Sect. 4 and by Appendix A therein.

5.5 The minimal theory of perturbative B violation

C. Arbeláez Rodríguez, H. Kolešová, M. Malinský, Physical Review D89 (2014) 055003, DOI: 10.1103/PhysRevD.89.055003

D. Harries, M. Malinský, M. Zdráhal, Physical Review D 98 (2018) 095015, DOI: 10.1103/PhysRevD.98.095015

The first of these articles represents a ground-breaking research of the candidate and his students in the field of minimal models of perturbative B violation; the second study is a sequel including mainly technical details not given in the first one.

The Witten's mechanism for the radiative generation of neutrino masses in the $SO(10)$ GUTs, as simple and beautiful as it is (cf. Sect. 3.6.1), has been for a long time taken as a mere curiosity without much of a practical use in the realm of potentially realistic unified models. This was mainly due to the dichotomy between the gauge unification constraints in the non-supersymmetric settings and the need to keep the $B-L$ symmetry-breaking VEVs of the relevant scalars in the vicinity of the GUT scale in order for the effective type-I seesaw scale to be generated in the desired 10^{12-13} GeV ballpark.

The central point of the study below was the observation that there exist more economical frameworks in which the original Witten's idea can be naturally realised without the need to conform the draconic grand-unification constraints of the $SO(10)$ settings. It was argued that the flipped $SU(5)$ gauge framework equipped with the Witten's loop is not only way simpler than any other renormalizable flipped $SU(5)$ scenario considered before but, thanks to the high degree of flavour correlations within, it may be even viewed as the most minimal model of perturbative baryon number violation ever conceived as it has something to say about virtually all the gauge-induced two-body decay channels of nucleons. Besides that, it has a potential to provide fundamental insights into the absolute light neutrino mass scale and the early-Universe baryon number asymmetry generation via the leptogenesis mechanism.

The candidate's contribution to this research was cenral as he was not only the inventor of the ideas sketched above (and in Sect. 3.6.2) but also a driving force of the subsequent research conducted on this ground. Besides most of the manuscript writing, he contributed by the explicit calculation and discussion of the two-loop diagrams studied in the second paper and by deriving the p-decay partial widths in the representative case presented therein. At the moment of writing he is supervising the preparation of a third study in which a complete numerical synthesis of the existing proton lifetime, absolute neutrino mass scale, gauge running and leptogenesis constraints should be accomplished, with possible predictions for the leptonic CP and the associated neutrinoless double-beta-decay observables.

Bibliography

[1] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).

[2] A. Salam and J. C. Ward, Phys. Lett. **13**, 168 (1964).

[3] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[4] Super-Kamiokande, S. Fukuda *et al.*, Phys. Lett. **B539**, 179 (2002), arXiv:hep-ex/0205075.

[5] SNO, B. Aharmim *et al.*, Phys. Rev. **C72**, 055502 (2005), arXiv:nucl-ex/0502021.

[6] KamLAND, S. Abe *et al.*, Phys. Rev. Lett. **100**, 221803 (2008), arXiv:0801.4589.

[7] E. W. Otten and C. Weinheimer, Rept. Prog. Phys. **71**, 086201 (2008), arXiv:0909.2104.

[8] K. N. Abazajian *et al.*, Astropart. Phys. **35**, 177 (2011), arXiv:1103.5083.

[9] S. Weinberg, Phys. Rev. **D22**, 1694 (1980).

[10] S. L. Adler, Phys. Rev. **177**, 2426 (1969).

[11] J. S. Bell and R. Jackiw, Nuovo Cim. **A60**, 47 (1969).

[12] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, Phys. Lett. **B155**, 36 (1985).

[13] W. Rodejohann, Int. J. Mod. Phys. **E20**, 1833 (2011), arXiv:1106.1334.

[14] A. S. Barabash, Double beta decay experiments: beginning of a new era, in *Proceedings, 4th International Conference Current Problems in Nuclear Physics and Atomic Energy (NPAE-2012): Kyiv, Ukraine, September 3-7, 2012*, pp. 13–20, 2012, arXiv:1209.4241.

[15] W.-Y. Keung and G. Senjanovic, Phys. Rev. Lett. **50**, 1427 (1983).

[16] S. Weinberg, Phys. Rev. Lett. **43**, 1566 (1979).

[17] S. Weinberg, Phys. Rev. **D26**, 287 (1982).

[18] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[19] V. A. Rubakov, JETP Lett. **33**, 644 (1981), [Pisma Zh. Eksp. Teor. Fiz.33,658(1981)].

[20] C. G. Callan, Jr., Phys. Rev. **D26**, 2058 (1982).

[21] K. A. Milton, Rept. Prog. Phys. **69**, 1637 (2006), arXiv:hep-ex/0602040.

[22] G. Battistoni *et al.*, Nucl. Instrum. Meth. **A245**, 277 (1986).

[23] FREJUS, C. Berger *et al.*, Nucl. Instrum. Meth. **A262**, 463 (1987).

[24] J. L. Thron, Nucl. Instrum. Meth. **A283**, 642 (1989).

[25] Kamiokande-II, K. S. Hirata *et al.*, Phys. Lett. **B220**, 308 (1989).

[26] R. Becker-Szendy *et al.*, Nucl. Instrum. Meth. **A324**, 363 (1993).

[27] Super-Kamiokande, H. Nishino *et al.*, Phys. Rev. Lett. **102**, 141801 (2009), arXiv:0903.0676.

[28] MEG, A. M. Baldini *et al.*, Eur. Phys. J. **C76**, 434 (2016), arXiv:1605.05081.

[29] T. P. Cheng and L.-F. Li, Phys. Rev. Lett. **38**, 381 (1977).

[30] S. T. Petcov, Sov. J. Nucl. Phys. **25**, 340 (1977), [Yad. Fiz.25,641(1977); Erratum: Sov. J. Nucl. Phys.25,698(1977); Erratum: Yad. Fiz.25,1336(1977)].

[31] M. Pospelov and A. Ritz, Annals Phys. **318**, 119 (2005), arXiv:hep-ph/0504231.

[32] ACME, J. Baron *et al.*, Science **343**, 269 (2014), arXiv:1310.7534.

[33] D. Ng and J. N. Ng, Mod. Phys. Lett. **A11**, 211 (1996), arXiv:hep-ph/9510306.

[34] G. Bressi *et al.*, Phys. Rev. **A83**, 052101 (2011), arXiv:1102.2766.

[35] J. A. Minahan, P. Ramond, and R. C. Warner, Phys. Rev. **D41**, 715 (1990).

[36] R. Foot, H. Lew, and R. Volkas, J.Phys.G **G19**, 361 (1993), arXiv:hep-ph/9209259.

[37] E. Majorana, Il Nuovo Cimento (1924-1942) **14**, 171 (2008).

[38] P. Minkowski, Phys. Lett. **B67**, 421 (1977).

[39] V. Rubakov and M. Shaposhnikov, Usp.Fiz.Nauk **166**, 493 (1996), arXiv:hep-ph/9603208.

[40] A. Wright, J. Phys. Conf. Ser. **888**, 012036 (2017).

[41] J. Schechter and J. W. F. Valle, Phys. Rev. **D25**, 2951 (1982), [,289(1981)].

[42] S. Sarkar, Measuring the baryon content of the universe, in *Proceedings, XIIIth Rencontres de Blois on Frontiers of the Universe: Blois, France, June 17-23, 2001*, pp. 53–63, 2004, arXiv:astro-ph/0205116.

[43] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967).

[44] S. R. Coleman, Phys.Rev. **D15**, 2929 (1977).

[45] J. Callan, Curtis G. and S. R. Coleman, Phys.Rev. **D16**, 1762 (1977).

[46] T. Brauner, O. Taanila, A. Tranberg, and A. Vuorinen, Phys. Rev. Lett. **108**, 041601 (2012), arXiv:1110.6818.

[47] M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).

[48] S. Davidson and A. Ibarra, Phys. Lett. **B535**, 25 (2002), arXiv:hep-ph/0202239.

[49] Y. Nir, Introduction to leptogenesis, in *6th Rencontres du Vietnam: Challenges in Particle Astrophysics Hanoi, Vietnam, August 6-12, 2006*, 2007, arXiv:hep-ph/0702199.

[50] FCC, A. Abada *et al.*, Eur. Phys. J. **C79**, 474 (2019).

[51] T. Appelquist and J. Carazzone, Phys. Rev. **D11**, 2856 (1975).

[52] S. Weinberg, Physica **A96**, 327 (1979).

[53] S. Scherer and H. W. Fearing, Phys. Rev. **D52**, 6445 (1995), arXiv:hep-ph/9408298.

[54] W. Buchmuller and D. Wyler, Nucl. Phys. **B268**, 621 (1986).

[55] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, JHEP **1010**, 085 (2010), arXiv:1008.4884.

[56] B. Henning, X. Lu, T. Melia, and H. Murayama, JHEP **08**, 016 (2017), arXiv:1512.03433, [Erratum: JHEP09,019(2019)].

[57] T. Yanagida, Horizontal gauge symmetry and masses of neutrinos, in *Proc. Workshop on the Baryon Number of the Universe and Unified Theories*, edited by O. Sawada and A. Sugamoto, p. 95, 1979.

[58] R. N. Mohapatra and G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980).

[59] M. Magg and C. Wetterich, Phys. Lett. **B94**, 61 (1980).

[60] P. Nath *et al.*, Nucl. Phys. Proc. Suppl. **200-202**, 185 (2010), arXiv:1001.2693.

[61] R. Foot, H. Lew, X. He, and G. C. Joshi, Z.Phys. **C44**, 441 (1989).

[62] Yu. A. Golfand and E. P. Likhtman, JETP Lett. **13**, 323 (1971), [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[63] D. V. Volkov and V. P. Akulov, Phys. Lett. **46B**, 109 (1973).

[64] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).

[65] B. Henning, X. Lu, T. Melia, and H. Murayama, Commun. Math. Phys. **347**, 363 (2016), arXiv:1507.07240.

[66] F. Reines *et al.*, Phys. Rev. **D4**, 80 (1971).

[67] F. Reines and M. F. Crouch, Phys. Rev. Lett. **32**, 493 (1974).

[68] K. Abe *et al.*, (2011), arXiv:1109.3262 [hep-ex].

[69] LBNE Collaboration, T. Akiri *et al.*, (2011), arXiv:1110.6249 [hep-ex].

[70] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).

[71] H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).

[72] S. Weinberg, Phys.Lett. **B91**, 51 (1980).

[73] L. J. Hall, Nucl.Phys. **B178**, 75 (1981).

[74] Y. Aoki, E. Shintani, and A. Soni, Phys. Rev. **D89**, 014505 (2014), arXiv:1304.7424.

[75] V. V. Dixit and M. Sher, Phys. Rev. **D40**, 3765 (1989).

[76] P. Nath and P. F. Perez, Phys. Rept. **441**, 191 (2007), arXiv:hep-ph/0601023.

[77] J. Chakrabortty and A. Raychaudhuri, Phys.Lett. **B673**, 57 (2009), arXiv:0812.2783 [hep-ph].

[78] X. Calmet, S. D. Hsu, and D. Reeb, Phys.Rev.Lett. **101**, 171802 (2008), arXiv:0805.0145 [hep-ph].

[79] G. 't Hooft, Nucl. Phys. **B79**, 276 (1974).

[80] A. M. Polyakov, JETP Lett. **20**, 194 (1974).

[81] J. A. Bryan, S. M. Carroll, and T. Pyne, Phys.Rev. **D50**, 2806 (1994), arXiv:hep-ph/9312254.

[82] T. W. B. Kibble, J. Phys. **A9**, 1387 (1976).

[83] W. H. Zurek, Nature **317**, 505 (1985).

[84] P. A. M. Dirac, Proc. Roy. Soc. Lond. **A133**, 60 (1931).

[85] P. A. M. Dirac, Phys. Rev. **74**, 817 (1948).

[86] K. Freese and E. Krasteva, Phys. Rev. **D59**, 063007 (1999), arXiv:astro-ph/9804148.

[87] L. Patrizii and M. Spurio, Ann. Rev. Nucl. Part. Sci. **65**, 279 (2015), arXiv:1510.07125.

[88] B. L. G. Bakker, A. I. Veselov, and M. A. Zubkov, Phys. Lett. **B583**, 379 (2004), arXiv:hep-lat/0301011.

[89] J. C. Baez and J. Huerta, Bull. Am. Math. Soc. **47**, 483 (2010), arXiv:0904.1556.

[90] H. Georgi, H. R. Quinn, and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).

[91] H. Kolešová and M. Malinský, (2016), arXiv:1612.09178.

[92] B. Bajc and G. Senjanovic, JHEP **0708**, 014 (2007), arXiv:hep-ph/0612029.

[93] I. Dorsner, P. Fileviez Perez, and R. Gonzalez Felipe, Nucl.Phys. **B747**, 312 (2006), arXiv:hep-ph/0512068.

[94] M. Malinsky, Eur.Phys.J. **C73**, 2415 (2013), arXiv:1212.4660.

[95] M. Hudec and M. Malinský, J. Phys. **G47**, 015004 (2020), arXiv:1902.04470.

[96] J. R. Ellis, D. V. Nanopoulos, and S. Rudaz, Nucl. Phys. **B202**, 43 (1982).

[97] P. Nath, A. H. Chamseddine, and R. L. Arnowitt, Phys. Rev. **D32**, 2348 (1985).

[98] B. Bajc, S. Lavignac, and T. Mede, JHEP **01**, 044 (2016), arXiv:1509.06680.

[99] J. C. Pati and A. Salam, Phys. Rev. **D10**, 275 (1974).

[100] P. Fileviez Perez and M. B. Wise, Phys. Rev. **D88**, 057703 (2013), arXiv:1307.6213.

[101] H. Murayama and J. Shu, Phys. Lett. **B686**, 162 (2010), arXiv:0905.1720.

[102] H. Georgi and C. Jarlskog, Phys. Lett. **B86**, 297 (1979).

[103] K. S. Babu, I. Gogoladze, P. Nath, and R. M. Syed, Phys. Rev. **D72**, 095011 (2005), arXiv:hep-ph/0506312.

[104] C. S. Aulakh, B. Bajc, A. Melfo, A. Rasin, and G. Senjanovic, Nucl. Phys. **B597**, 89 (2001), arXiv:hep-ph/0004031.

[105] T. E. Clark, T. K. Kuo, and N. Nakagawa, Phys. Lett. **B115**, 26 (1982).

[106] C. S. Aulakh and R. N. Mohapatra, Phys. Rev. **D28**, 217 (1983).

[107] T. Fukuyama and N. Okada, JHEP **11**, 011 (2002), arXiv:hep-ph/0205066.

[108] T. Fukuyama, A. Ilakovac, T. Kikuchi, S. Meljanac, and N. Okada, Phys. Rev. **D72**, 051701 (2005), hep-ph/0412348.

[109] K. S. Babu and C. Macesanu, (2005), arXiv:hep-ph/0505200.

[110] C. S. Aulakh, B. Bajc, A. Melfo, G. Senjanovic, and F. Vissani, Phys. Lett. **B588**, 196 (2004), arXiv:hep-ph/0306242.

[111] B. Bajc, A. Melfo, G. Senjanovic, and F. Vissani, Phys. Rev. **D70**, 035007 (2004), arXiv:hep-ph/0402122.

[112] M. Malinský, arXiv:0807.0591 [hep-ph].

[113] C. S. Aulakh and S. K. Garg, (2006), arXiv:hep-ph/0612021.

[114] B. Bajc, G. Senjanovic, and F. Vissani, Phys. Rev. Lett. **90**, 051802 (2003), arXiv:hep-ph/0210207.

[115] H. S. Goh, R. N. Mohapatra, and S.-P. Ng, Phys. Rev. **D68**, 115008 (2003), arXiv:hep-ph/0308197.

[116] S. Bertolini, T. Schwetz, and M. Malinský, Phys. Rev. **D73**, 115012 (2006), arXiv:hep-ph/0605006.

[117] C. S. Aulakh and S. K. Garg, Nucl. Phys. **B757**, 47 (2006), arXiv:hep-ph/0512224.

[118] S. Bertolini, M. Frigerio, and M. Malinský, Phys. Rev. **D70**, 095002 (2004), arXiv:hep-ph/0406117.

[119] C. S. Aulakh, arXiv:hep-ph/0607252.

[120] B. Bajc, I. Dorsner, and M. Nemevsek, JHEP **11**, 007 (2008), arXiv:0809.1069.

[121] S. Bertolini, L. Di Luzio, and M. Malinský, Phys. Rev. **D81**, 035015 (2010), arXiv:0912.1796 [hep-ph].

[122] S. Bertolini, L. Di Luzio, and M. Malinský, Phys. Rev. **D80**, 015013 (2009), arXiv:0903.4049 [hep-ph].

[123] D. Chang, R. N. Mohapatra, J. Gipson, R. E. Marshak, and M. K. Parida, Phys. Rev. **D31**, 1718 (1985).

[124] J. M. Gipson and R. E. Marshak, Phys. Rev. **D31**, 1705 (1985).

[125] N. G. Deshpande, E. Keith, and P. B. Pal, Phys. Rev. **D47**, 2892 (1993), arXiv:hep-ph/9211232.

[126] M. Yasue, Phys. Lett. **B103**, 33 (1981).

[127] M. Yasue, Phys. Rev. **D24**, 1005 (1981).

[128] G. Anastaze, J. P. Derendinger, and F. Buccella, Z. Phys. **C20**, 269 (1983).

[129] K. S. Babu and E. Ma, Phys. Rev. **D31**, 2316 (1985).

[130] S. Bertolini, L. Di Luzio, and M. Malinský, J.Phys.Conf.Ser. **259**, 012098 (2010), arXiv: 1010.0338 [hep-ph].

[131] S. R. Coleman and E. J. Weinberg, Phys. Rev. **D7**, 1888 (1973).

[132] L. Gráf, M. Malinský, T. Mede, and V. Susič, Phys. Rev. **D95**, 075007 (2017), arXiv:1611.01021.

[133] S. Bertolini, L. Di Luzio, and M. Malinský, Phys.Rev. **D87**, 085020 (2013), arXiv:1302.3401 [hep-ph].

[134] S. Bertolini, L. Di Luzio, and M. Malinský, AIP Conf. Proc. **1467**, 37 (2012), arXiv:1205.5637.

[135] S. Bertolini, L. Di Luzio, and M. Malinský, Phys. Rev. **D85**, 095014 (2012), arXiv:1202.0807 [hep-ph].

[136] H. Kolešová and M. Malinský, Phys. Rev. **D90**, 115001 (2014), arXiv:1409.4961.

[137] DUNE, B. Abi *et al.*, (2018), arXiv:1807.10334.

[138] Y. Mambrini, N. Nagata, K. A. Olive, J. Quevillon, and J. Zheng, Phys. Rev. **D91**, 095010 (2015), arXiv:1502.06929.

[139] I. Dorsner, S. Fajfer, A. Greljo, and J. F. Kamenik, JHEP **1211**, 130 (2012), arXiv:1208.1266.

[140] M.-x. Luo and Y. Xiao, Phys. Lett. **B555**, 279 (2003), arXiv:hep-ph/0212152.

[141] R. M. Fonseca, M. Malinský, W. Porod, and F. Staub, Nucl. Phys. **B854**, 28 (2012), arXiv:1107.2670 [hep-ph].

[142] R. M. Fonseca, M. Malinský, and F. Staub, Phys.Lett. **B726**, 882 (2013), arXiv:1308.1674 [hep-ph].

[143] S. P. Martin and M. T. Vaughn, Phys.Rev. **D50**, 2282 (1994), arXiv:hep-ph/9311340.

[144] E. Witten, Phys. Lett. **B91**, 81 (1980).

[145] A. Salam and J. A. Strathdee, Phys. Rev. **D11**, 1521 (1975).

[146] M. T. Grisaru, W. Siegel, and M. Rocek, Nucl. Phys. **B159**, 429 (1979).

[147] N. Arkani-Hamed, S. Dimopoulos, G. F. Giudice, and A. Romanino, Nucl. Phys. **B709**, 3 (2005), hep-ph/0409232.

[148] B. Bajc and G. Senjanovic, Phys. Lett. **B610**, 80 (2005), hep-ph/0411193.

[149] B. Bajc and G. Senjanovic, Phys.Rev.Lett. **95**, 261804 (2005), arXiv:hep-ph/0507169.

[150] S. M. Barr, Phys. Lett. **B112**, 219 (1982).

[151] J. Derendinger, J. E. Kim, and D. V. Nanopoulos, Phys.Lett. **B139**, 170 (1984).

[152] S. Bertolini, L. Di Luzio, and M. Malinský, Phys. Rev. **D83**, 035002 (2011), arXiv:1011.1821 [hep-ph].

[153] G. K. Leontaris and J. D. Vergados, Phys. Lett. **B258**, 111 (1991).

[154] C. Arbeláez Rodríguez, H. Kolešová, and M. Malinský, Phys. Rev. **D89**, 055003 (2014), arXiv:1309.6743.

[155] D. Harries, M. Malinský, and M. Zdráhal, Phys. Rev. **D98**, 095015 (2018), arXiv:1808.02339.

[156] R. N. Mohapatra, Phys. Rev. Lett. **56**, 561 (1986).

[157] R. N. Mohapatra and J. W. F. Valle, Phys. Rev. **D34**, 1642 (1986).

[158] B. Bajc, A. Melfo, G. Senjanovic, and F. Vissani, Phys. Lett. **B634**, 272 (2006), arXiv:hep-ph/0511352.