Tracing New Physics: From Symmetry to Observation

Thesis by
Samuel Patrone

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended May 27, 2026

Samuel Patrone
ORCID: 0000-0001-5319-2729

To the City of Angels,
for giving me back to myself.

ACKNOWLEDGEMENTS

"A journey of a thousand miles begins with a single step"
attributed to Lao Tzu

It was a rainy day at the beginning of fall when my high school physics teacher entered the classroom and wrote these words on the blackboard. That day, while learning about the International System of Units, I took my first step into physics. Sixteen years later, as I complete my education with a PhD, some reflection and acknowledgments are in order.

First, I want to thank all my mentors and advisors, who guided me along the path: steadying the helm when I doubted myself, trusting me when I needed confidence, leading me when I was lost, and letting me go when I needed independence. A special mention goes to Luciano, who was there for the very first steps; Antonello, who taught me the beauty of quantum mechanics; Alan and Francesco, who assisted me in my first baby steps into original research; and finally Mark, my advisor with a capital A, for guiding me through this six-year-long journey. Through their preparation, wit, intellectual generosity, patience, and contagious enthusiasm for physics, they have been a source of inspiration and guidance: my lighthouse while I was navigating turbulent waters, and examples of the kind of scientist, mentor, and educator I hope to become.

Science is a common effort of humankind to overcome our limits and push the boundaries of our knowledge, and everybody knows that *"everything is cool when you are part of a team"*. That is why I am grateful to all my more senior collaborators, who were amazing companions in this adventure and, in many ways, big siblings to me: Kim, Nikita, Pavel, and Ryan.

A special mention goes to Clarita. I still remember the first time we talked outside your office, and the rainbow bracelet that immediately caught my attention. Thank you for our endless discussions about physics and life, for the crazy SEC nights in LA, and for being the amazing woman you are: passionate, hard-working, full of enthusiasm, and forever a point of reference for the person I wish to become. Thank you for both your mentorship and your friendship.

These years would not have been the same without all the wonderful people I met along the way, and to whom I am deeply grateful.

Thank you Adriano, for being my closest friend along my PhD adventure. For our calls during COVID, with an ocean between us; for picking me up at LAX, with a Mia Martini playlist, on my very first day in this City of Angels; for sharing countless days and nights talking about everything and more; for standing by me throughout the years, at the blackboard or in front of a Winchell's donut, getting mangos, launching rockets, and planting potatoes together. Thank you for being my collaborator, friend, confidant, and chosen family; for helping me through moments of panic and uncertainty; for walking with me through heartbreak and change; and for all the conversations about life, physics, and the future. You have been one of the people who made these years special.

Thank you Francesco, for being one of the closest and most constant presences in my life over the last two years. For the songs you gifted me, for your understanding, patience, love, and for becoming one of the people with whom I could be myself, with no filters. Thank you for your intelligence, sensitivity, and creativity; for the conversations in which we could talk about anything, from music to politics, from the meaning of life to the thousand facets of who we are; and, of course, for your help with the color palette of my paper (see Fig. 4.9). Our bond has been one of the most amazing gifts life has given me.

I would like to thank the *Italians*: Federico, Michele, and Francesco. Our pizza nights and pasta dinners made me feel at home, even miles away. Our chats in front of the Jura machine, when it was not broken, were definitely the most "productive" moments of the day. Thank you guys for your friendship, kindness, and for bringing a little piece of Italy into my life in Pasadena.

Thank you *Grand Rabbi* Yakov Landau, for being the best officemate I could have asked for. I am going to miss our dinners together, your very loud coffee sipping, and our endless fights over which lights to keep on. Every conversation with you was an adventure and an opportunity to learn something new: about physics, Aramaic, etymology, investments, life, faith, or some obscure detail that only you could somehow make fascinating. You are one of the most intelligent, multifaceted, and wonderfully curious people I have ever met, and I will forever hold our conversations dear.

I also want to thank the Closson family: Tim, Clemmy, Aubrey, and Matt, for being my West Coast American family throughout these years. Thank you for opening the doors of your home and your hearts to this Italian student, and for welcoming me as one of your own. Your warmth, generosity, and love made California feel like home.

Thank you Elliott, for teaching me the patience of sharing my space with someone else, and for being a mentor in one of the most important parts of my self-discovery.

Thank you Joe, for giving me the gift of my first love. The end of our story taught me how to grow through loss, how to let go, and how to find myself again. Thank you for your help in one of the darkest moments of my life, and for all our memories together: *I'll Always Remember Us This Way*, on a Malibu beach, staring at the sunset and holding hands.

Thank you Maxx, for entering my life during the very final stretch of this journey, and for being *my boy* in this unexpected new chapter.

There are some people who have been there since before my PhD, and to whom I owe much of the man I have become.

Thank you Gioacchino, my *compa'* in this physics journey, for all the long calls about life and for all the adventures we shared around the world, from Paris to Rome, from LA to New York. Your unique irony, sarcasm, and intelligence have made me smile and lifted my spirit even in the darkest moments. Thank you for being one of those rare companions with whom the road always feels lighter, no matter how far apart we are.

I am grateful to my lifelong friends, Alessio, Leonardo, and Pierluigi, for their endless support, even with an ocean between us. Thank you for making the time (or at least plans, right Leo?) to visit me and share in my new life on the West Coast. Thank you for understanding me, for being happy for me, and for encouraging me, always, to be myself, to pursue my dreams, and to embrace the freedom I found here.

Thank you to my parents, Silde and Amato, who watched me change and grow throughout these years, leaving the nest and learning to spread my wings. Thank you for accepting me for who I am and whom I love, without judgment. Gibran wrote that parents are the bows from which their children are sent forth as living arrows. You have been that bow for me: steady, loving, and brave enough to let me fly. I know it must have been difficult at times to let me go, to see me build a life so far away, and to embrace parts of me that took time to understand. But you never made your struggle my burden, and your unconditional love gave me the freedom to become myself.

Thank you to Zia Rosy, who has always been on my side, no matter what; who knows me better than I know myself; and who can sense my feelings from six thousand miles away. Thank you for being always present in my life, even when you were

going through the storm. For our morning calls, and for our ritual before every exam: your mock-French *dans la bouche du lupe* and my *creps, creps con la Nutella*, without which I am not sure I would have made it this far. Thank you for making me feel protected even from across the ocean, and for the endless generosity with which you helped me build my life here. In many ways, we are made of the same fire: stubborn, fierce, protective, and full of that unmistakable *cazzimma*. You are the person I know would do anything for me, and I hope you know I would do anything for you too. You taught me how to love intensely, how to care deeply, and how to stay close to the people who matter, even from far away.

Thank you to my grandparents, Nonno Mimmo and Nonna Pea, for still being present in my life in the most beautiful ways: with verses, with artichoke lasagna, with our videocalls through Alexa on a Sunday morning, and with the kind of love that needs no explanation.

Thank you to my East Coast American family, Zia Sara, Zio Roel, Tatjana, and Ylenia, for being one of my anchors in this foreign country. Thank you for loving me since I was born, for making New York feel like another home, and for welcoming me with such generosity every time I arrive, often with a friend in tow.

Thank you Zio Tore, for all the love you have given me throughout these years. For our Tuesday dinners at Mercato Centrale during my university days; for coming back from the States always with a gift; for all the times you spoke to me about science and physics and made me look at the world with curiosity; for our unforgettable road trip through the land of canyons and Las Vegas; and for your punctual Wednesday calls at 9 in the morning, exactly when I have therapy and cannot answer. Thank you for being affectionate, present, and generous. You have been family in one of the most beautiful meanings of the word, and I have always felt accepted and loved exactly as I am.

My final thought goes to Nonna Mena and Viviana, whom I like to imagine waiting for me just around the corner, close in ways that spacetime cannot capture.

And to all the people who crossed my life along the way: those who were always there, those who stayed for a season or a night, those who left too early, and those who will remain forever in some corner of my heart. Thank you for being part of the journey. It was never mine alone.

ABSTRACT

Despite the success of the Standard Model of particle physics and the ΛCDM model of cosmology in describing many aspects of nature, several fundamental questions still demand extensions of our current understanding. This thesis studies how effective descriptions and symmetry principles can be used to trace possible new physics and understand their effect on low-energy observables.

The first part focuses on large-scale structure as a probe of primordial physics. Since galaxies are biased tracers of the underlying matter density field, extracting primordial information from galaxy surveys requires a controlled treatment of nonlinear gravitational evolution, galaxy bias, and short-distance sensitivity. We study the renormalization of composite operators in the galaxy bias expansion in the presence of local primordial non-Gaussianity, showing how different ultraviolet regulators affect counterterm coefficients while preserving the structures fixed by symmetry.

The second part turns to physics beyond the Standard Model. We study finite naturalness in Pati-Salam theories of quark-lepton unification, deriving constraints from finite threshold corrections to the Higgs mass. We then investigate long-lived axion-like particles in beam-dump experiments, showing that electromagnetic cascades can substantially enhance the sensitivity of such searches. Finally, we study leptogenesis in automatic Nelson-Barr models, identifying viable regions where spontaneous CP violation can both protect the quality of the strong CP solution and generate the observed baryon asymmetry.

Together, these results trace new physics from symmetry to observation, showing how effective descriptions turn microscopic principles into cosmological and experimental targets.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] S. Patrone, A. Testa, and M. B. Wise. "*Regularization scheme dependence of the counterterms in the galaxy bias expansion*". JCAP 11 (2023). DOI: 10.1088/1475-7516/2023/11/087. arXiv: 2306.08025 [astro-ph.CO].

[2] P. Fileviez Pérez, C. Murgui, S. Patrone, A. Testa, and M. B. Wise. "*Finite naturalness and quark-lepton unification*". Phys. Rev. D 109.1 (2024). DOI: 10.1103/PhysRevD.109.015011. arXiv: 2308.07367 [hep-ph].

[3] C. Murgui and S. Patrone. "*Leptogenesis in automatic Nelson-Barr models*". JHEP 09 (2025). DOI: 10.1007/JHEP09(2025)113. arXiv: 2506.18963 [hep-ph].

[4] S. Patrone, N. Blinov, and R. Plestid. "*Long-lived axionlike particles from electromagnetic cascades*". Phys. Rev. D 113.7 (2026). DOI: 10.1103/38hq-qk7k. arXiv: 2509.14310 [hep-ph].

Chapters in this thesis reproduce results and content from the publications listed above. All authors contributed equally.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632 .

TABLE OF CONTENTS

Acknowledgements iv
Abstract viii
Published Content and Contributions ix
Table of Contents ix
List of Illustrations xii
List of Tables xix
Introduction 1

I Formal Aspects of Effective Descriptions in Cosmology **5**

Chapter I: Large-Scale Structure, Primordial Non-Gaussianity, and Renormalized Bias 6
1.1 Biased tracers for large-scale structure: a historical perspective . . . 6
1.2 From local bias to the general perturbative bias expansion 8
1.3 Primordial non-Gaussianity and scale-dependent bias 10
1.4 Renormalization and physical bias parameters 12
1.5 Summary and outlook 15
Chapter II: Regularization Scheme Dependence of the Counterterms in the Galaxy Bias Expansion 16
2.1 Introduction 16
2.2 Theoretical Framework 17
2.3 Renormalization of the operator δ_ρ^2 at one loop 21
2.4 Concluding Remarks 29
Appendix 2.A Conformal Invariance of δ_ζ correlators at tree-level . . . 29
Appendix 2.B Feynman Rules for the Galaxy Bias expansion 33

II Symmetry, Naturalness, and Hidden Sectors Beyond the Standard Model **37**

Chapter III: Finite Naturalness and Quark-Lepton Unification 38
3.1 Introduction 38
3.2 Minimal Theory For Quark-Lepton Unification 39
3.3 Concluding Remarks 50
Chapter IV: Long-Lived Axionlike Particles from Electromagnetic Cascades . 52
4.1 Introduction 52
4.2 Axion couplings at low scales 55
4.3 Axion production mechanisms 57
4.4 Monte Carlo Methods 63
4.5 Experimental projections 66

4.6 Conclusions . . . 75
Appendix 4.A Inelastic Primakoff-like scattering . . . 76
Appendix 4.B Monte Carlo production weights . . . 77
Appendix 4.C Energy thresholds at SHiP . . . 80
Appendix 4.D N_{sig} dependence . . . 83
Appendix 4.E Primary photon spectra used to seed the SHiP cascade . . . 83
Chapter V: Leptogenesis in Automatic Nelson-Barr Models . . . 86
5.1 Introduction . . . 86
5.2 The Nelson-Barr mechanism in a Nutshell . . . 89
5.3 Quality of Nelson-Barr Solution . . . 91
5.4 Leptogenesis . . . 96
5.5 Methods and Results . . . 107
5.6 Potential signatures . . . 114
5.7 Conclusions . . . 116
Appendix 5.A Automatic Nelson-Barr UV completions . . . 118
Appendix 5.B The lightest heavy neutrino mass . . . 122
Appendix 5.C The Davidson-Ibarra bound for more than three heavy neutrinos . . . 123
Bibliography . . . 126

LIST OF ILLUSTRATIONS

Number *Page*

0.1 Schematic summary of the four regimes explored in this thesis. New physics may be searched for in the distant past through large-scale structure, at high energies through naturalness constraints, at weak coupling through long-lived hidden particles, and at high precision through CP-violating observables. The four research projects presented in this thesis provide examples of these complementary windows. 2
2.1 Generic one-loop divergent diagrams at first order in gravitational evolution. 22
2.2 Diagrams with a single F_2 vertex used to determine the counterterm for δ_ρ^2 linear in δ_ρ. 24
2.3 Two routings of the loop diagram with a single F_2 vertex used to determine the counterterm for δ_ρ^2 linear in δ_ρ with Gaussian primordial fluctuations. 25
2.4 Diagrams with one F_3 and two F_2's used to determine the counterterms for δ_ρ^2 quadratic in δ_ρ with Gaussian primordial fluctuations. 27
2.5 Generic chain diagram. 30
2.6 Contributions to $T^{(4)}$. 31
2.7 Feynman rules for the vertices and the lines used in the computation of galaxy overdensity field correlators. 33
2.8 Diagram contributing to the three point function of the galaxy overdensity field. 35
2.9 Example of an infrared divergent diagram. 36

4.1 Comparison to past literature – PBC [53] (green shaded region), ALPINIST [62] (red shaded region) – of our projections for the sensitivity of SHiP (solid black line) to axion-photon couplings, with updated geometry and nominal protons-on-target (POT) [63, 64]. For reference, we also plot our sensitivity projections with a lower POT, used before in the literature (dashed black line). We defined $g_{a\gamma\gamma} \equiv f^{-1} c_{a\gamma\gamma} \alpha / \pi$, and we set $c_{aee} = 0$. Accounting for the full electromagnetic cascade substantially enhances the reach of SHiP at lower couplings and lighter masses. See Fig. 4.11 for an illustration of how this result depends on the energy threshold used at SHiP. . . . 54

4.2 Electromagnetic production mechanisms for axion-like particles in a beam dump 58

4.3 Benchmark experimental parameters used in our SHiP and BDX projections. For BDX there is no decay pipe; detectable decays must occur within the calorimeter fiducial volume. 66

4.4 Event rate of axions per protons-on-target (POT), weighted by f^4, for an ALP with mass $m_a = 100$ MeV and photon-dominated couplings. The setup simulates SHiP photons emerging from meson decays impinging on a molybdenum target, with the corresponding SHiP experimental geometry. The flux is calculated for ALPs directed towards the detector. The individual production channels are: Primakoff conversion from shower photons (green); bremsstrahlung from shower electrons and positrons (orange); e^+e^- annihilation and Compton-like scattering of shower $e^\pm$ (blue); Primakoff conversion from the primary photon beam from neutral meson decays (red). 68

4.5 Event rate of axions produced per electrons-on-target (EOT), weighted by f^4, for an ALP with mass $m_a = 100$ MeV and electron-dominated couplings. The setup simulates a 10.6 GeV electron beam on an aluminum target, with a geometry representative of the BDX experiment. The flux is restricted to ALPs with trajectories pointing towards the forward detector. The contributions from different production mechanisms are shown separately: Primakoff conversion from shower photons (green); bremsstrahlung from shower electrons and positron (orange); e^+e^- annihilation and Compton-like scattering from shower particles (blue); bremsstrahlung (red) and photon fusion (purple) from the primary electron beam. 68

4.6 Comparison between the primary-only and the shower-induced coupling-independent weighted event rates in the LLP limit for different axion masses for SHiP (**left**) and BDX (**right**). As above, the axion count has been weighted by decay (in the LLP limit), geometry, and energy acceptance. In both panels, as the mass is increased, the ability of softer shower particles to produce axions is diminished, and production rates from the showers becomes smaller (in the case of SHiP) than the primary-only estimates for masses greater than 1 GeV. . . . 69

4.7 Projected sensitivity for 5 signal events for the SHiP experiment for the electron-dominated (**left**) and photon-dominated (**right**) ALP coupling scenarios, assuming an integrated 6×10^{20} POT and an energy acceptance threshold $E_{\text{cut}} = 200$ MeV. The green shaded region shows the sensitivity for ALPs generated by primary photons arising from neutral meson decays. The solid lines show the individual contributions from shower-induced processes: Primakoff (orange), bremsstrahlung (blue), and resonant annihilation/Compton scattering (purple). The total sensitivity, combining primary and shower contributions, is shown by the dashed black line, with the gray shaded area indicating the full parameter space covered. For the photon-dominated case, we do not show the bremsstrahlung and resonant annihilation/Compton scattering lines since they are sub-dominant with respect to all the other processes. 71

4.8 Projected sensitivity for 5 signal events for the BDX experiment for the electron-dominated (**left**) and photon-dominated (**right**) ALP coupling scenarios, assuming an integrated 10^{22} EOT and an energy acceptance threshold $E_{\text{cut}} = 300$ MeV. The green shaded region represents the contribution from ALPs generated by the initial electron beam, and it is obtained by summing the contributions from the bremsstrahlung (dotted green line) and the EPA γ-fusion process (dot-dashed green line). The solid lines show the contributions from the subsequent electromagnetic shower: Primakoff (orange), bremsstrahlung (blue), and annihilation/Compton (purple). The total sensitivity is shown by the dashed black line, which encompasses the gray shaded region. For the photon-dominated scenario, we do not show the bremsstrahlung and resonant annihilation/Compton scattering lines since they are sub-dominant with respect to the Primakoff channel. 72

4.9 Comparison between our sensitivity projections for SHiP (black solid line) and BDX (black dashed line) with other relevant bounds (colored shaded regions) and projections (colored dotted lines) in the literature. We use $g_{a\gamma\gamma} \equiv f^{-1} c_{a\gamma\gamma}\alpha/\pi$, and we set $c_{aee} = 0$ (the same conventions as Fig. 4.1). Starred bounds were computed without taking into account any enhancement from secondary production. Sources: PrimEx (projection only, current bound not shown) [98], BaBar [99], NA64 [100], FASER (57.7 fb^{-1}) [101], EIC Prompt (10 fb^{-1} line) and EIC Displaced ($S_a = 100$, $[10, 100]$ cm line) [102], Belle II [103], NuCal [62, 104, 105], DarkQuest (10^{18} POT) [71], MiniBooNE Beam Dump [59, 106], E137 [99], DUNE (GAr, 1 yr line) [57], SNe Cooling and CCSNe (1 B line) [107], BBN+N_{eff} [108], GW170817 fireball bound [109]. LUXE-NPOD Phase-I [110] projection curve is not shown since it is very close to DarkQuest. . . 74

4.10 Ratio of the projected downstream ALP flux (in the LLP limit) at SHiP calculated with and without the inelastic Primakoff form factor ($G_{2,\text{inel}}$), for γ-dominated couplings. The enhancement is shown separately for ALPs produced by primary photons from neutral meson decays (blue) and secondary photons from electromagnetic showers (orange). The inclusion of inelastic scattering provides an $\mathcal{O}(1)$ enhancement to the flux, which becomes increasingly significant at higher ALP masses. 77

4.11 Sensitivity curves of SHiP for 5 signal event in the electron-dominated (**left panel**) and photon-dominated (**right panel**) benchmarks with different energy cuts applied. The green shaded region shows the sensitivity for ALPs generated by primary photons arising from neutral meson decays. The solid, dashed, and dotted black line, are indicating the full parameter space covered including the shower contributions for $E_{\text{cut}} = 100,\ 300,\ 1000$ MeV respectively. We observe (almost) no dependence on the energy cut for the primary reach. 81

4.12 Evolution of the ALP flux through successive application of weights for $m_a = \{3, 10, 30, 100, 300, 1000\}$ MeV; note that we work in the LLP limit, and scale out $1/f^4$ such that weights can be greater than unity. **Left panel:** Raw production flux, i.e. with $w_{\rm prod}$ only, from Primakoff conversion (from both primaries photons arising from neutral meson decay and secondary photons arising from electromagnetic showers) showing enhanced production for lighter ALPs due to the abundance of soft photons in the showers. **Middle panel:** Decay-weighted flux in the LLP limit ($w_{\rm prod} \times w_{\rm decay}$), with high-energy suppression from ALPs escaping the decay pipe; note that $w_{\rm decay}$ is greater than unity because $1/f$ has been factored out. **Right panel:** Downstream predicted flux including the angular acceptance of the detector ($w_{\rm prod} \times w_{\rm decay} \times w_{\rm geom}$), with no energy cut applied. The interplay between geometric acceptance and decay kinematics creates a characteristic double-cutoff structure that shifts to higher energies for heavier ALPs. All fluxes are normalized per proton on target (POT) and made coupling-independent by rescaling by the appropriate powers of f. . . . 82
4.13 $N_{\rm sig}$ dependence of the projected SHiP sensitivity for the electron-dominated (**left panel**) and photon-dominated (**right panel**) ALP coupling scenarios. We show contours corresponding to $N_{\rm sig} = 5$ (solid), $N_{\rm sig} = 10$ (dashed), and $N_{\rm sig} = 100$ (dotted) signal events, assuming an integrated 6×10^{20} POT and an energy acceptance threshold $E_{\rm cut} = 200$ MeV. 83
4.14 $N_{\rm sig}$ dependence of the projected BDX sensitivity for the electron-dominated (**left panel**) and photon-dominated (**right panel**) ALP coupling scenarios. We show contours corresponding to $N_{\rm sig} = 5$ (solid), $N_{\rm sig} = 10$ (dashed), and $N_{\rm sig} = 100$ (dotted) signal events, assuming an integrated 10^{22} EOT and an energy acceptance threshold $E_{\rm cut} = 300$ MeV. 84

4.15 Primary-photon spectra used to seed the SHiP cascade simulation. The photon list shown is a random subsample of $N_{\text{samp}} = 5 \times 10^3$ photons drawn from a parent population of size N_{parent}, which constitute all the photons obtained from the 10^5 pp `PYTHIA`'s collision simulations. **Left panel**: photon energy spectrum. **Right panel**: photon transverse-momentum spectrum. Histograms show the observed subsample counts, while the curves show the expected counts obtained by scaling the parent-population distributions to the subsample size, $N_{\text{samp}}/N_{\text{parent}}$. Both panels use logarithmic binning on the horizontal axis. Photons with $E \leq 50$ MeV have been removed from the sampled list. 85

5.1 Upper panel: Radiative corrections at 1-loop contributing to $\bar{\theta}_{\text{QCD}}$ by inducing complex quark masses. Lower panel: Tree-level and 1-loop corrections to the quartic coupling $H^\dagger H X_a X_b^*$. In square brackets we add the would-be contribution from extending the Nelson-Barr mechanism to the neutral lepton sector (including right-handed neutrinos). 95

5.2 Possible heavy neutrino physical mass spectra for different hierarchies. We follow the notation from the main text. Panel A corresponds to the hierarchies $m_N \gg [\mathsf{M}_R] \gg |\mu_i|$ and $m_N \gg |\mu_i| \gg [\mathsf{M}_R]$. In both cases $[\mu]$ does not impact the heavy neutrino masses. Panel B represents the mass spectrum when $[\mathsf{M}_R] \gg m_N \gg |\mu_i|$. Panel C involves the hierarchies $|\mu_i| \gg m_N \gg [\mathsf{M}_R]$, where the lightest heavy neutrino mass is either m_{R_1} or $m_{R_i} m_N^2/|\mu_i|^2$, and $|\mu_i| \gg [\mathsf{M}_R] \gg m_N$, with the lightest heavy neutrino mass always being given by $m_{R_i} m_N^2/|\mu_i|^2$. We note that in the latter case, m_N does not impact the heavy neutrino masses. Panel D corresponds to the hierarchy $[\mathsf{M}_R] \gg |\mu_i| \gg m_N$ (except for very hierarchical M_R, which lead to the spectrum from Panel B instead). Although we assumed three decoupled scales to reproduce the panels above, similar hierarchies can be found relaxing this condition. 102

5.3 Tree-level and one-loop level diagrams for the decays of the lightest heavy neutrino N_1 that can contribute to the CP asymmetry. 103

5.4 Scatter plots showing the parameter space that lead to successful leptogenesis, obtained by randomly sampling the magnitude of the Yukawa couplings, Y, from log-uniform distributions with varying ranges. Specifically, the log values of Y are drawn from a uniform distribution $\log \mathsf{Y} \sim \mathcal{U}(n, 0)$, where $n = -1$, -2, or -5 sets the lower bound of the distribution. This corresponds to logarithmically sampling Y over one (green points), two (orange points), or five (blue points) orders of magnitude below the upper bound, which is fixed at $\mathsf{Y} = 1$ (see discussion in footnote 13). The points satisfy $\eta_B^{\rm obs} < |\eta_B| < 1.5\,\eta_B^{\rm obs}$, $v_{\rm CP} > T_i$, and $M_{N_{j\neq 1}} > 5 M_{N_1}$. 109

5.5 Scatter plots of the successful points for leptogenesis as a function of M_{N_1} and the ratio $v_{\rm CP}/T_i$ (required to be greater than one by our selection procedure), for a TIA (left) and DIA (right) of N_1s. In the first row, we show all successful points from our numerical analysis. In the second and third row, we show which of these succesful points are compatible with the quality bounds for automatic Nelson-Barr models with new vector-like and chiral quarks, respectively. Different colors of the points indicated different spreads taken for the Yukawa couplings, according to the legends. 112

5.6 Scatter plots of the successful leptogenesis points and their induced shift in $\bar{\theta}_{QCD}$ computed with Eq. (5.47) as a function of $v_{\rm CP}$ (first row), $v_{\rm CP}/T_i$ (second row), and M_{N_1} (third row), for a TIA (left) and DIA (right) of N_1s. In red, the zone currently excluded from electric dipole moment experiments. In gray, delimitated by dashed lines, we show the projective bounds from future experiments on the nEDM [202–205], and the proton electric dipole moment proton electric dipole moment (p-EDM) [206]. 115

LIST OF TABLES

Number *Page*

5.1 Quantum numbers of the field content under the $SU(3) \otimes SU(2)_L \otimes U(1)_Y \otimes U(1)_B \otimes U(1)_L$ gauge group. The first box contains just the SM fields. B and L are unconstrained baryon and lepton charges, respectively. 121

INTRODUCTION

The Standard Model of particle physics and the standard cosmological model encoded in the ΛCDM paradigm have been extraordinarily successful in describing a wide range of phenomena in nature. The Standard Model provides a quantum field theoretic description of ordinary matter and its interactions, while ΛCDM gives a remarkably economical account of the expansion history of the universe and the formation of cosmic structure. Together, they form the backbone of our present understanding of fundamental physics.

Yet their success is also sharply incomplete. The Standard Model describes only the ordinary matter component of the cosmic energy budget, leaving dark matter and dark energy outside its microscopic description. The origin of primordial fluctuations, the physics of inflation, the possible presence of primordial non-Gaussianities, and the nature of the dark sector all point to questions beyond the minimal cosmological model. Conversely, within particle physics, neutrino masses, the baryon asymmetry of the universe, the stability of the electroweak scale, the flavor structure of the Standard Model, and the absence of observed strong CP violation all suggest that the Standard Model is not the final microscopic theory.

These open questions motivate what we collectively call *New Physics*: new particles, forces, symmetries, mechanisms, or effective degrees of freedom that extend our current paradigms. New physics, however, need not appear in only one way. It may be hidden in the very early universe, encoded in correlations among primordial fluctuations and later traced by the distribution of galaxies. It may be hidden at high energies, where heavy particles leave threshold corrections to low-energy parameters. It may be hidden through weak couplings, producing long-lived particles that escape ordinary collider searches. It may also be hidden in high-precision observables, where tiny symmetry-violating effects reveal the structure of physics at inaccessible scales.

The common theme of this thesis is the role of effective descriptions in tracing new physics from symmetry principles to observable signatures. Effective theories are not merely approximations to more fundamental descriptions. They are the language in which low-energy observables become insensitive to microscopic details that cannot be resolved directly. Short-distance physics then appears through a finite set of parameters, counterterms, Wilson coefficients, threshold corrections,

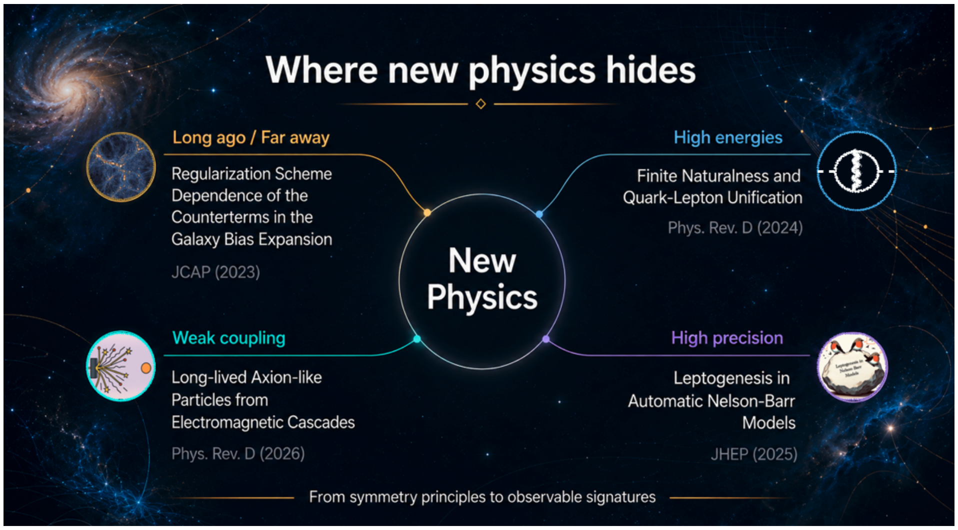


Figure 0.1: Schematic summary of the four regimes explored in this thesis. New physics may be searched for in the distant past through large-scale structure, at high energies through naturalness constraints, at weak coupling through long-lived hidden particles, and at high precision through CP-violating observables. The four research projects presented in this thesis provide examples of these complementary windows.

or symmetry-protected textures. Symmetries determine which terms are allowed, which quantities are protected, and which observables can retain memory of physics beyond the scales directly probed.

The first part of this thesis focuses on the oldest and largest available laboratory: the large-scale structure of the universe. The distribution of galaxies observed at late times carries information about both the initial conditions of the universe and their subsequent nonlinear gravitational evolution. In particular, primordial non-Gaussianities provide one of the sharpest probes of inflationary dynamics. While current Cosmic Microwave Background measurements already place stringent bounds on their size, large-scale structure offers a complementary observational window through the clustering of biased tracers.

This perspective is especially timely. Current and upcoming galaxy surveys are producing increasingly detailed maps of the large-scale distribution of matter. These datasets make it essential to develop a theoretical framework in which subtle signatures of primordial physics can be robustly separated from late-time gravitational evolution, galaxy formation, and short-distance nonlinearities. In this context, the effective theory approach provides the necessary organization principle: the impact of unresolved short-distance modes is absorbed into renormalized bias parameters, while the observable predictions remain cutoff independent.

In Chapter 1, we introduce the formalism of biased tracers, which relates the observed galaxy overdensity to the underlying matter density field. We review how primordial curvature fluctuations are mapped into late-time matter perturbations, how nonlinear gravitational evolution generates additional non-Gaussianities, and how the galaxy bias expansion organizes the relation between galaxies and matter. Since this expansion involves composite operators, its correlators are sensitive to short-distance modes. We therefore explain why regularization and renormalization are necessary in order to define physical observables.

In Chapter 2, based on *Regularization Scheme Dependence of the Counterterms in the Galaxy Bias Expansion* [1], we study this issue explicitly. We examine how different ultraviolet regulators affect the counterterms needed to renormalize the galaxy bias expansion, focusing on the operator δ_ρ^2 in the presence of local primordial non-Gaussianity and nonlinear gravitational evolution. We show which counterterms are regulator dependent and which structures are fixed by symmetry. This chapter illustrates, in a concrete cosmological setting, one of the main lessons of effective theory: unphysical short-distance ambiguities can be absorbed into parameters, while symmetry controls the structures that enter observable predictions.

The second part of the thesis turns from cosmological large-scale structure to physics beyond the Standard Model. The organizing logic remains the same. The Standard Model leaves several structural puzzles unanswered, and each chapter studies one way in which symmetry principles and effective descriptions connect possible microscopic extensions to measurable consequences.

In Chapter 3, based on *Finite Naturalness and Quark-Lepton Unification* [2], we study new physics hidden at high energies. We examine Pati-Salam theories of quark-lepton unification and ask whether low-scale quark-lepton unification can remain compatible with the observed lightness of the Higgs. The relevant low-energy trace of the heavy sector is a finite, physical threshold correction to the Higgs mass. This chapter uses finite naturalness as a diagnostic of whether a symmetry-motivated extension of the Standard Model can remain compatible with the electroweak scale.

In Chapter 4, based on *Long-Lived Axion-like Particles from Electromagnetic Cascades* [3], we study new physics hidden by weak couplings. Axion-like particles are naturally light when protected by approximate shift symmetries, and their feeble interactions can make them long lived. Beam dump experiments are therefore powerful probes of such hidden sectors. We show that electromagnetic cascades inside thick targets can substantially enhance the production of visibly decaying axion-like particles, improving the projected reach of SHiP and BDX. In this case, the effective low-energy couplings of a light pseudoscalar determine how a hidden sector may become experimentally visible.

In Chapter 5, based on *Leptogenesis in Automatic Nelson-Barr Models* [4], we study new physics hidden in high-precision observables and early-universe dynamics. The strong CP problem suggests that CP violation is more constrained than a generic Standard Model extension would allow, while the baryon asymmetry of the universe requires new CP-violating dynamics. Automatic Nelson-Barr models provide a framework in which CP is an exact symmetry in the ultraviolet and is broken spontaneously. We investigate whether the same spontaneous source of CP violation can generate both the observed quark CP phase and the baryon asymmetry through high-scale leptogenesis, while preserving the quality of the strong CP solution.

Taken together, the chapters of this thesis explore four complementary ways in which new physics may hide: in the distant past, at high energies, through weak couplings, and in high-precision symmetry tests. Across these examples, the central idea is the same. Symmetry principles constrain the microscopic possibilities, effective descriptions translate them into low-energy language, and observations determine which traces of new physics can survive. In this sense, this thesis traces new physics from symmetry to observation, showing how theoretical structure becomes a guide to the physics still to be discovered.

Part I

Formal Aspects of Effective Descriptions in Cosmology

Chapter 1

LARGE-SCALE STRUCTURE, PRIMORDIAL NON-GAUSSIANITY, AND RENORMALIZED BIAS

In this introductory chapter, we will briefly review the emergence of *biased tracers* as an effective description of the large-scale structure of the Universe, mostly following the historical introduction given in [5]. We will then pedagogically build the general perturbative bias expansion for both Gaussian and non-Gaussian primordial conditions. Finally, we present the issue of renormalization of the bias parameters and operators, completing the needed background knowledge to understand the material presented in the following chapter.

1.1 Biased tracers for large-scale structure: a historical perspective

At the foundation of our knowledge of cosmic history lies the *large-scale structure* of the Universe, namely the observed distribution of visible matter in the form of galaxies, quasars, clusters, and, conversely, its absence in voids. Throughout this chapter, we will refer to these objects collectively as *tracers* of the underlying matter density field, which is not directly accessible to observation. If the full matter distribution could be observed directly, it would provide a powerful probe of the composition of the Universe, the nature of dark energy and gravity on large scales, and the mechanism that generated the primordial seeds of structure. The need to infer this information from visible tracers instead leads naturally to the notion of **bias**: the effective description of the statistical relation between the observed tracer distribution and the underlying matter density field.

One of the earliest cosmological conclusions from sky surveys came from the observed distribution of radio sources [6], which was inconsistent with a static homogeneous Universe. However, it soon became apparent that we needed to go beyond one-point functions to extract more information from visible matter. The first analyses of the two-point correlation function of the galaxy density field, and of its Fourier transform, the *power spectrum*, showed that galaxies in our Universe behaved differently with respect to clusters: they do not share the same correlation function. Since they both trace the same underlying matter distribution, this was the first empirical evidence that they cannot both be unbiased tracers of the matter distribution. This simple observation was the origin of the first phenomenological model of bias, where the number of galaxies $n_g(\mathbf{x})$ at a certain fixed time is no longer a linear function of the underlying matter field $\rho_m(\mathbf{x})$, but the two respective overdensities, $\delta_g(\mathbf{x})$ and $\delta_\rho(\mathbf{x})$, are related through a bias parameter b_1. In formulae,

$$\delta_g(\mathbf{x}) \equiv \frac{n_g(\mathbf{x})}{\bar{n}_g} - 1 = b_1\, \delta_\rho(\mathbf{x}) = b_1 \left(\frac{\rho_m(\mathbf{x})}{\bar{\rho}_m} - 1 \right), \tag{1.1}$$

where $\bar{n}_g$ ($\bar{\rho}_m$) is the mean comoving number density of galaxies (background matter density). This simple relation allows the two-point function of galaxies (or clusters) to have an enhancement factor b_1^2 with respect to the analogous statistical correlator of the matter density, hence explaining the observations.

The first attempt to give biased clustering a physical interpretation was made by Kaiser [7] in the context of rich clusters of galaxies and later made quantitative with Gaussian random-field statistics by Bardeen and collaborators in [8]. In the *high-peak bias* picture, one separates the density field into small-scale fluctuations and a long-wavelength background mode. Regions where the small-scale density exceeds a threshold are identified as the sites where proto-objects form. Because a long-wavelength overdensity raises the local background, it makes threshold crossing more likely there, so these rare peaks are more strongly clustered than the matter itself. This means the selected objects are already spatially biased in the *Lagrangian* sense, even before gravitational evolution. For very massive objects like clusters, the threshold is physically tied to the requirement that collapse occur by today, which implies selecting rare peaks. For galaxies, this simple mechanism is less automatic, and additional physics is needed to explain why star formation should preferentially occur in the earliest-collapsing regions. These seminal works consequently led to the development of the *peak-background split* (PBS) framework [9–11], which first established the connection between the bias parameters and the mean abundance of tracers. Already by the early 1990s, it had become clear that bias could be generalized beyond the naive peak/threshold approach, in a more systematic expansion: non-linear bias was pioneered by Fry and Gaztañaga [12].

Galaxies are not the only relevant tracers. Groups and clusters of galaxies, voids, quasars, the Lyman-α forest, line-intensity maps, the 21cm signal, and diffuse backgrounds can all be treated in the same broad framework of biased tracers. For simplicity, in the rest of the chapter we will use “galaxies” as shorthand for any tracer population.

The theory of galaxy clustering is a theory for matter plus bias, not for matter alone. Indeed, only a theory that predicts both the statistics of the underlying matter field and describes how the observed tracers are related to it enables us to interpret clustering data correctly. For the theoretical description of the distribution of matter, we can use gravitational perturbation theory to study gravitational collapse starting from the initial perturbations. The validity of such an approach depends on the scale considered: at redshift $z \lesssim 10$, the perturbations in the matter density field are mostly non-linear. However, on *quasi-linear scales*, perturbation theory converges to the right result if carried out to sufficiently high order. The first half of the theoretical work is done once we reach a consistent framework to relate the initial conditions to the late-time density and tidal fields (second derivatives of the gravitational potential) at such scales. Then, the goal of the theory of galaxy bias is to write the local number of galaxies, n_g, as a function of the large-scale environment allowed by the general covariance under coordinate transformations. Remarkably, a finite number of *bias parameters* is sufficient at each order in perturbation theory and at a fixed time to marginalize over the unknown details of galaxy formation while still extracting robust cosmological information. The modern approach to biased tracers is therefore the development of a systematic expansion in operators, each with its own bias coefficient, in formulae

$$\delta_g(\mathbf{x},\tau) = \sum_{O} b_O(\tau) O(\mathbf{x},\tau)\,. \tag{1.2}$$

The rest of the chapter will be devoted to describing which operators are allowed to enter such an expansion and how to relate the bias parameters to their physical values.

In conclusion, the problem of galaxy clustering on large scales is reduced to two coupled tasks: predicting the quasi-linear matter field perturbatively, and expressing the tracer density in terms of the most general perturbative bias expansion consistent with symmetry. This is the winning approach to confidently extract information about our Universe's history and content from *biased tracers*.

1.2 From local bias to the general perturbative bias expansion

The simplest starting point in order to construct a bias description that is both physically meaningful and systematically improvable on large scales consists in the so-called *local-in-matter-density* (LIMD) bias in Lagrangian space. Such an approach can be regarded as the modernized version of the high-peak intuition: we just assume that the abundance of proto-tracers (that will later become galaxies) is just a local expansion of the smoothed initial overdensity of the matter field. In formulae,

$$\delta_g^L(\mathbf{q}) = b_1^L \delta_R^{(1)}(\mathbf{q}) + \frac{1}{2} b_2^L \left[\delta_R^{(1)}(\mathbf{q})\right]^2 + \ldots \tag{1.3}$$

where $\mathbf{q}$ is the Lagrangian coordinate (comoving with the proto-tracer patch), $\delta_R^{(1)}(\mathbf{q})$ is the linearly evolved matter overdensity smoothed on the proto-tracer scale R, and b_i^L are the first Lagrangian bias parameters.

Gravitational evolution mixes operators, and the late-time tracer density cannot, in general, be expressed as a function of the matter density alone. The final Eulerian position $\mathbf{x}$ is related to the Lagrangian position $\mathbf{q}$ through

$$\mathbf{x}(\tau) = \mathbf{q} + \mathbf{s}(\mathbf{q}, \tau)\,, \tag{1.4}$$

where $\mathbf{s}$ is the displacement field. If we assume that the number of tracers is conserved, we can get the following relation between the observed (Eulerian) tracer density $\delta_g(\mathbf{x}, \tau)$ and the Lagrangian one:

$$1 + \delta_g(\mathbf{x}, \tau) = \left[1 + \delta_g^L(\mathbf{q})\right] \left[1 + \delta_\rho(\mathbf{x}, \tau)\right]\,, \tag{1.5}$$

evaluated along the fluid map between $\mathbf{q}$ and $\mathbf{x}$. This relation already shows that even a simple local dependence in Lagrangian space is mapped by gravitational evolution into a more general Eulerian structure, in which galaxy formation depends on the gravitational environment along the fluid trajectory.

As we anticipated above, the modern approach to galaxy bias is to identify all the operators in Eq. (1.2) at a given order compatible with the symmetries. Under the equivalence principle, the relevant quantities must be second derivatives of the gravitational potential Φ. All the LIMD terms, i.e. powers of the matter overdensity δ_ρ^N, are related to $\nabla^2\Phi$ through the Poisson equation. The first new operator beyond density that appears is the tidal field, K_{ij}, defined as

$$K_{ij}(\mathbf{x}, \tau) \equiv \left(\frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}\right) \delta_\rho(\mathbf{x}, \tau)\,. \tag{1.6}$$

There is no symmetry reason to include δ^2 but exclude $K^2 \equiv K_{ij}K^{ij}$ from the bias expansion. At second order in the overdensity field, we can therefore write the complete galaxy bias expansion in Eulerian coordinates as

$$\delta_g(\mathbf{x}, \tau) = b_1(\tau)\, \delta_\rho(\mathbf{x}, \tau) + \frac{1}{2} b_2(\tau)\, \delta_\rho^2(\mathbf{x}, \tau) + b_{K^2}(\tau)\, K^2(\mathbf{x}, \tau) + O(\delta_\rho^3)\,. \tag{1.7}$$

There are two other types of operators that appear in the full galaxy bias expansion:

- *Higher-derivative terms*: they arise because galaxy formation depends on a region of finite size $R_\star$: expanding this finite spatial sensitivity on large scales, one obtains terms with more than two derivatives acting on a single instance of the gravitational potential Φ, such as $\nabla^2\delta_\rho$. These terms are usually suppressed by powers of $R_\star^2\nabla^2$, so they are only relevant at small scales.

- *Stochastic terms*: From the point of view of large-scale physics, tracer formation still depends on some unresolved short-scale physics which cannot be predicted *deterministically* by the large-scale density and tidal environment. This is often encoded in noise terms ε or $\varepsilon_O O$, where the ε, ε_O fields are uncorrelated with the large-scale perturbations described by the operator O and uncorrelated among themselves on large scales.

The discussion so far assumed that the initial conditions are Gaussian and adiabatic, so that the large-scale clustering of tracers can be organized in terms of local gravitational observables, together with higher-derivative and stochastic contributions. This provides the baseline perturbative description of galaxy bias on quasi-linear scales. However, one of the main virtues of large-scale structure as a cosmological probe is precisely its sensitivity to departures from Gaussian initial conditions. Primordial non-Gaussianity modifies the relation between long- and short-wavelength modes, and therefore changes both the physical interpretation and the operator content of the bias expansion. It is thus natural to regard primordial non-Gaussianity not as an alternative to the framework developed above, but as its first important generalization. In the next section, we introduce this extension and explain how the presence of primordial non-Gaussianity reshapes the large-scale clustering of biased tracers.

1.3 Primordial non-Gaussianity and scale-dependent bias

So far, the discussion has assumed Gaussian initial conditions, so that the statistics of short- and long-wavelength perturbations are independent at the initial time. Primordial non-Gaussianity changes this picture in an essential way: long and short modes become coupled already in the primordial field. Since the abundance of tracers depends on the local statistics of short-scale fluctuations, this coupling implies that tracer formation acquires a direct dependence on long-wavelength primordial perturbations. The bias expansion must therefore be generalized, and one of its most striking consequences is the emergence of a scale-dependent bias on very large scales.

The first theoretical analysis of the effect of non-Gaussianity on bias was done in [13], in the context of local primordial non-Gaussianity. This is the assumption that the primordial curvature fluctuations δ_ζ can be expanded in a power series of a Gaussian field φ_G as follows:

$$\delta_\zeta(\mathbf{x}) = \varphi_G + f_{\rm NL}\left(\varphi_G^2 - \langle\varphi_G^2\rangle\right) + g_{\rm NL}\varphi_G^3 + \dots \tag{1.8}$$

If we now split the Gaussian field into long and short modes, $\varphi_G = \varphi_\ell + \varphi_s$, we see that the short-wavelength curvature perturbation $\delta_{\zeta,s}$ at leading order in the long mode can be written as

$$\delta_{\zeta,s}(\mathbf{x}) \approx \varphi_s(\mathbf{x}) + 2 f_{\rm NL}\, \varphi_\ell(\mathbf{x})\, \varphi_s(\mathbf{x}) + \cdots\,, \tag{1.9}$$

As a consequence, the local small-scale power spectrum is modulated by the long-wavelength potential,

$$P_s(k_s|\mathbf{x}) \equiv \left\langle \delta_{\zeta,s}(\mathbf{k}_s)\, \delta_{\zeta,s}(-\mathbf{k}_s)\right\rangle_{\varphi_\ell(\mathbf{x})} \approx \left[1 + 4 f_{\rm NL}\varphi_\ell(\mathbf{x})\right] P_s(k_s)\,. \tag{1.10}$$

This mode coupling, not present for Gaussian initial conditions, enhances or suppresses the amplitude of the short-scale fluctuations according to the long-wavelength potential.

Since the abundance of tracers depends on the local small-scale power, the modulation above implies that biased tracers respond directly to the long-wavelength primordial field. The bias expansion must therefore be enlarged by operators sourced by the primordial potential itself. At leading order, this introduces a new contribution of the schematic form

$$\delta_g(\mathbf{x},\tau) = b_1(\tau)\, \delta_\rho(\mathbf{x},\tau) + b_\varphi(\tau)\, f_{\rm NL}\, \varphi_G(\mathbf{x}) + \cdots\,. \tag{1.11}$$

This is the simplest way to see that primordial non-Gaussianity modifies the Gaussian bias expansion: in addition to the late-time gravitational observables discussed in the previous section, the large-scale tracer density now depends on the primordial field that controls the local statistics of short modes.

Finally, the late-time linear matter overdensity is related to the primordial potential through a transfer function,

$$\delta_\rho^{(1)}(\mathbf{k},\tau) = \mathcal{M}(k,\tau)\, \delta_\zeta(\mathbf{k})\,. \tag{1.12}$$

One can rewrite the primordial contribution in terms of the matter field itself as

$$\delta_g(\mathbf{k},\tau) = \left[b_1(\tau) + b_\varphi(\tau)\, f_{\rm NL}\, \mathcal{M}^{-1}(k,\tau)\right] \delta_\rho^{(1)}(\mathbf{k},\tau)\,. \tag{1.13}$$

This shows that local primordial non-Gaussianity induces an effective scale-dependent linear bias coefficient

$$b_{\mathrm{eff}}(k,\tau) = b_1(\tau) + b_\varphi(\tau)\, f_{\mathrm{NL}}\, \mathcal{M}^{-1}(k,\tau)\,, \tag{1.14}$$

which, on sufficiently large scales where $\mathcal{M}(k,\tau) \propto k^2$, gives rise to the characteristic k^{-2} enhancement. This scale-dependent bias on very large scales is the best-known observational signature of local primordial non-Gaussianity in galaxy surveys, and it was first identified in N-body simulations in [14].

The Gaussian bias expansion therefore provides only the baseline large-scale description. In the presence of primordial non-Gaussianity, long-short mode coupling in the initial conditions generates new operators and new large-scale signatures in tracer clustering, whose consistent treatment requires a refined understanding of composite bias operators and their renormalization.

1.4 Renormalization and physical bias parameters

As we have seen in the previous sections, the galaxy overdensity field δ_g can be expanded in terms of local gravitational observables and, in the presence of primordial non-Gaussianity, operators sourced by the primordial field. Operators such as δ_ρ^2 are composite operators: they involve the product of fields evaluated at the same point and they are sensitive to short-distance fluctuations.

To make this short-distance sensitivity explicit, it is useful to introduce a coarse-graining scale Λ such that

$$\delta_\rho(\mathbf{x}) = \delta_{\rho,\Lambda}(\mathbf{x}) + \delta_{\rho,>\Lambda}(\mathbf{x})\,. \tag{1.15}$$

where $\delta_{\rho,\Lambda}$ is the coarse-grained long-wavelength field, while $\delta_{\rho,>\Lambda}$ contains the short modes above the smoothing scale. The bias expansion is meant to describe tracer clustering in terms of the long modes only, while the effects of the short modes should be absorbed into the bias coefficients and stochastic terms.

Under this split, the simplest nontrivial composite operator in the galaxy bias expansion becomes

$$\delta_\rho^2(\mathbf{x}) = \delta_{\rho,\Lambda}^2(\mathbf{x}) + 2\,\delta_{\rho,\Lambda}(\mathbf{x})\,\delta_{\rho,>\Lambda}(\mathbf{x}) + \delta_{\rho,>\Lambda}^2(\mathbf{x})\,. \tag{1.16}$$

In this decomposition, the first term is the long-wavelength quadratic operator one would like to keep, while the second and third terms describe, respectively, the mixing between long and short modes and a purely short-scale contribution.

If we integrate out the short modes, the mixed term vanishes at leading order in the long-wavelength expansion, while the last term gives a local contribution that depends on the cutoff Λ. In formulae,

$$\delta_\rho^2(\mathbf{x}) \;\to\; \delta_{\rho,\Lambda}^2(\mathbf{x}) + \sigma_{>\Lambda}^2 + \cdots\,, \tag{1.17}$$

where

$$\sigma_{>\Lambda}^2 \equiv \left\langle \delta_{\rho,>\Lambda}^2(\mathbf{x}) \right\rangle = \int_{|\mathbf{k}|>\Lambda} \frac{d^3k}{(2\pi)^3}\, P_\rho(k)\,, \tag{1.18}$$

and $P_\rho(k)$ is the matter power spectrum of the field δ_ρ in Fourier space.

The naive quadratic operator therefore already contains cutoff-dependent short-scale pieces, which one might try to remove simply by subtracting the constant $\sigma_{>\Lambda}^2$. However, this is not sufficient once non-linear evolution is taken into account. In that case, the short-scale variance is itself not independent of the long-wavelength background: in a region with a slowly varying long mode $\delta_{\rho,\Lambda}$, the local statistics of the short modes are shifted. Equivalently, the short-scale power spectrum responds to the long-wavelength overdensity. Schematically, one may write

$$\left\langle \delta_{\rho,>\Lambda}^2(\mathbf{x}) \right\rangle_{\delta_{\rho,\Lambda}} = \sigma_{>\Lambda}^2 + \alpha(\Lambda)\, \delta_{\rho,\Lambda}(\mathbf{x}) + \cdots\,, \tag{1.19}$$

where the coefficient $\alpha(\Lambda)$ encodes the response of the short modes to the long-wavelength background. This is the first place where operator mixing becomes visible: the response of the short-scale variance to the long-wavelength background generates a contribution with the same large-scale structure as the linear density operator. As we will see in the next chapter, in standard perturbation theory, this response is proportional to the short-mode variance; for an isotropic long-wavelength overdensity one finds

$$\left\langle \delta_{\rho,>\Lambda}^2(\mathbf{x}) \right\rangle_{\delta_{\rho,\Lambda}} = \sigma_{>\Lambda}^2 \left[1 + \frac{68}{21}\, \delta_{\rho,\Lambda}(\mathbf{x}) \right] + \cdots\,. \tag{1.20}$$

Substituting this back into the decomposition of δ_ρ^2, one obtains

$$\delta_\rho^2(\mathbf{x}) \;\to\; \delta_{\rho,\Lambda}^2(\mathbf{x}) + \sigma_{>\Lambda}^2 + \frac{68}{21}\sigma_{>\Lambda}^2\, \delta_{\rho,\Lambda}(\mathbf{x}) + \cdots\,. \tag{1.21}$$

This shows that integrating out the short modes generates not only a cutoff-dependent constant piece, but also a contribution proportional to the coarse-grained density itself. In other words, the bare quadratic operator mixes with lower-order operators. This can be achieved by defining the *renormalized operator* $[\delta_\rho^2]$, which is **independent** of the coarse-graining scale Λ:

$$[\delta_\rho^2](x) \equiv \delta_{\rho,\Lambda}^2(x) - Z_{\delta^2,1}(\Lambda) - Z_{\delta^2,\delta}(\Lambda)\,\delta_{\rho,\Lambda}(x) - \cdots\,, \tag{1.22}$$

where the ellipsis stands for any additional operators allowed by symmetry that are required at the same perturbative order. In the above formula, the *counterterm* $Z_{\delta^2,1}(\Lambda)$ removes the unphysical constant contribution, while $Z_{\delta^2,\delta}(\Lambda)$ subtracts the part of the quadratic operator that behaves like the linear density on large scales. The bare operator $\delta_{\rho,\Lambda}^2$ is therefore not yet a physical quadratic operator; only after renormalization do its coefficients acquire a direct physical meaning.

The same logic applies to all composite operators appearing in the perturbative bias expansion. Once one introduces a long/short split, short-scale contractions can in general generate lower-order long-wavelength operators with the same symmetries, and these must also be subtracted. The quadratic operator is just the simplest nontrivial example that already displays the general phenomenon of operator mixing. In more general terms, in the bare coarse-grained expansion, the cutoff label belongs on both the coefficients and the operators of the galaxy bias expansion, which can be written as

$$\delta_{g,\Lambda}(x) = \sum_O c_{O,\Lambda}\, O_\Lambda(x)\,. \tag{1.23}$$

After *renormalization*, one rewrites the same large-scale theory in terms of the renormalized operators and cutoff-independent coefficients,

$$\delta_g(x) = \sum_O b_O\,[O](x)\,. \tag{1.24}$$

In the equations above, the $c_{O,\Lambda}$ are only intermediate, cutoff-dependent bookkeeping quantities, whereas the b_O are the physical bias parameters that can be meaningfully inferred from large-scale observables.

The main constraint is that the allowed operators are fixed by the equivalence principle and by locality in the gravitational observables (with the already discussed exception of primordial non-Gaussianity). Therefore, the counterterms and operators that appear in the renormalization procedure are not arbitrary: they must be built from the same long-wavelength operator basis introduced earlier. In other words, renormalization does not change the physical content of the theory; it only reorganizes the expansion in terms of operators whose large-scale correlations are well defined. This is the main reason why achieving closure under renormalization is one of the most powerful checks to make sure that we included all the relevant operators at a given order.

We call *regularization* a specific choice or prescription for how to perform the split into short and long modes of Eq. (1.15). The main topic of the next chapter will be studying the impact of different regularization prescriptions on the counterterms of the galaxy bias expansion. In particular, this general logic is illustrated explicitly through the one-loop renormalization of the quadratic operator δ_ρ^2, whose counterterms are worked out in detail and compared across different regularization prescriptions.

1.5 Summary and outlook

In this chapter, we introduced the conceptual framework underlying the modern theory of biased tracers in large-scale structure. We first reviewed the historical emergence of bias as an effective description of the relation between visible tracers and the underlying matter field, and showed how this evolved into a systematic perturbative expansion on quasi-linear scales. In this modern perspective, the galaxy overdensity is written in terms of the most general operator basis consistent with the symmetries of the problem, with a finite set of bias parameters encoding the complicated microphysics of galaxy formation.

We then discussed how primordial non-Gaussianity modifies this Gaussian baseline. In the local case, long- and short-wavelength modes are coupled already in the initial conditions, so that the abundance of tracers becomes sensitive to the long-wavelength primordial field. This enlarges the operator content of the bias expansion and gives rise to the characteristic scale-dependent bias on large scales.

Finally, we explained why the bias expansion must be renormalized. Once the matter field is separated into long and short modes, composite operators such as δ_ρ^2 inherit cutoff-dependent contributions from short-scale fluctuations, which can reproduce lower-order long-wavelength structures. Renormalization reorganizes the theory in terms of renormalized operators and physical bias parameters whose large-scale correlations are well defined.

This provides the background necessary for the next chapter, where these ideas are developed explicitly through the one-loop renormalization of the quadratic operator δ_ρ^2 in the presence of primordial local non-Gaussianity and nonlinear gravitational evolution.

Chapter 2

REGULARIZATION SCHEME DEPENDENCE OF THE COUNTERTERMS IN THE GALAXY BIAS EXPANSION

In this chapter we explore how different regularization prescriptions affect the counterterms in the renormalization of the galaxy bias expansion. We work in the context of primordial local non-Gaussianity including non-linear gravitational evolution. We carry out the one-loop renormalization of the field δ_ρ^2 (i.e. the square of the matter overdensity field) up to third order in gravitational evolution. Three regularization schemes are considered and their impact on the values of the counterterms is studied. We explicitly verify that the coefficients of the non-boost invariant operators are regularization scheme independent.

2.1 Introduction

The galaxy bias expansion (for a review see [5]) relates the galaxy overdensity field δ_g (i.e. the relative fluctuations in the number density of galaxies) to the mass overdensity field δ_ρ. In this chapter we don't distinguish between dark matter halos and galaxies. The composite fields in this expansion can be regulated by a short distance cutoff ($1/\Lambda$). Renormalization renders the galaxy overdensity correlators cutoff independent. This has been studied both for Gaussian primordial curvature fluctuations [15–17] and for non-Gaussian primordial curvature fluctuations [13, 18].

In Section 2.2, we review the theoretical framework of local non-Gaussianity for primordial curvature fluctuations. The non-linear gravitational evolution of the matter overdensity field is discussed. We finally introduce three regularization schemes for the composite operator[1] δ_ρ^2 occurring in the galaxy bias expansion.

In Section 2.3, we re-examine the renormalization procedure for the composite operator δ_ρ^2 in the presence of primordial local non-Gaussianities up to third order in gravitational evolution. We explore the impact of three regularization prescriptions on renormalization. The first reproduces the coefficients and operators found in the literature [18]. The other two prescriptions close under renormalization using the same operators as the first one. However, in the presence of non-linear gravitational evolution, the coefficients of the boost-invariant operators for the three prescriptions

[1] In this work we use the terms field and operator interchangeably.

differ by quantities which can be written as surface terms in the UV diverging integrals.

In Appendix 2.A, after introducing a diagrammatic notation, we prove the conformal invariance of the tree-level correlators of curvature fluctuations in the context of local primordial non-Gaussianity. In Appendix 2.B, we introduce a diagrammatic notation to compute and estimate the correlators of the galaxy overdensity field and comment on an alternative way of regularizing IR divergences.

2.2 Theoretical Framework

Primordial Curvature Fluctuations and Local non-Gaussianity

Local non-Gaussianity is the hypothesis that the primordial curvature fluctuations δ_ζ are local functions of a Gaussian field φ_G of the form

$$\delta_\zeta(\mathbf{x}) = \sum_{n=1}^{\infty} f_n : \varphi_G^n(\mathbf{x}) : \ . \tag{2.1}$$

In the above equation, operators surrounded by colons are normal ordered (i.e. any contraction between themselves in correlators vanishes), φ_G is a Gaussian field with zero expectation value and the f_n's are constants. In this chapter we set $f_1 = 1$. Demanding the correlators of φ_G to be invariant under translations, rotations and scale transformations fixes the two-point function of φ_G to be

$$\langle \tilde{\varphi}_G(\mathbf{k}_1)\tilde{\varphi}_G(\mathbf{k}_2)\rangle = (2\pi)^3\delta(\mathbf{k}_1 + \mathbf{k}_2)\, P_\varphi(k_1)\,, \tag{2.2}$$

where a tilde denotes a wavevector space quantity, $P_\varphi(q) = A/q^{3-2\Delta}$ is the tree-level power spectrum of primordial curvature fluctuations (A is a constant), Δ is the scale dimension of the field φ_G, and $k_i = |\mathbf{k}_i|$. In this chapter we will set $\Delta = 0$.

Measurements of the Cosmic Microwave Background (CMB) anisotropy place bounds [19] on the local non-Gaussianity parameters, in particular (with the convention $f_1 = 1$) $A \sim 10^{-8}$, $f_2 = 3/5 f_{NL}^{\text{local}} = 3/5(0.8 \pm 5)$, $f_3 = (9/25) g_{NL}^{\text{local}}$ and $g_{NL}^{\text{local}} = (-9.0 \pm 7.7) \times 10^4$. It can be seen, by rescaling $\varphi_G \to \sqrt{A}\varphi_G$ and factoring out a common $\sqrt{A}$, that a highly non-Gaussian theory would have $f_n \sim O\left[A^{(-n+1)/2}\right]$. Hence, the above CMB bounds already indicate that our universe has nearly-Gaussian primordial curvature fluctuations.

In Appendix 2.A, after introducing a diagrammatic notation for the computation of the δ_ζ N-point functions, we prove by induction the conformal invariance of these correlators at tree-level.

Matter Density Perturbations

Even when the primordial curvature fluctuations are Gaussian, non-Gaussianities in the matter overdensity field arise at late times due to the non-linear gravitational evolution.

For simplicity, we shall assume that the all the matter in the universe is in the form of cold dark matter and behaves as a pressureless irrotational fluid. Labelling each fluid element trajectory by its initial position $\mathbf{x}_0$ (i.e. its Lagrangian coordinate), its Eulerian coordinate at conformal time τ is

$$\mathbf{x}(\mathbf{x}_0, \tau) = \mathbf{x}_0 + \mathbf{s}(\mathbf{x}_0, \tau) \tag{2.3}$$

where

$$\mathbf{s}(\mathbf{x}_0, \tau) = \int_0^\tau d\tau' \, \mathbf{v}(\mathbf{x}(\mathbf{x}_0, \tau'), \tau') \tag{2.4}$$

is called the displacement vector and $\mathbf{v}$ is the fluid element velocity.

Using standard gravitational perturbation theory [20–22] in the Newtonian approximation, the solutions to the Euler and Poisson equations in an expanding universe for the overdensity matter field $\delta_\rho(\mathbf{x}, \tau)$ can be written as

$$\delta_\rho(\mathbf{x}, \tau) = D(\tau)\delta_\rho^{(1)}(\mathbf{x}) + D^2(\tau)\delta_\rho^{(2)}(\mathbf{x}) + D^3(\tau)\delta_\rho^{(3)}(\mathbf{x}) + \dots , \tag{2.5}$$

where $D(\tau)$ is the growth factor, $D(\tau)\delta_\rho^{(1)}(\mathbf{x})$ is the solution to the linearized equations and we kept only the fastest growing modes at each order. We can express $\tilde{\delta}_\rho^{(n)}$ as

$$\begin{aligned}\tilde{\delta}_\rho^{(n)}(\mathbf{k}) = \left[\prod_{i=1}^{n} \int \frac{\mathrm{d}^3 k_i}{(2\pi)^3}\right] (2\pi)^3 \delta_{\mathrm{D}}^{(3)}\left(\mathbf{k} - \sum_{i=1}^{n} \mathbf{k}_i\right) F_n(\mathbf{k}_1, \cdots, \mathbf{k}_n) \\ \times \tilde{\delta}_\rho^{(1)}(\mathbf{k}_1) \cdots \tilde{\delta}_\rho^{(1)}(\mathbf{k}_n) \, .\end{aligned} \tag{2.6}$$

where the F_n's are called splitting functions and $F_1 = 1$. The velocity field is described by its divergence $\theta(\mathbf{x}, \tau) = \nabla \cdot \mathbf{v}(\mathbf{x}, \tau)$ which, similarly to the field $\delta_\rho(\mathbf{x}, \tau)$, can be written as a perturbative solution of the Euler and Poisson equations as

$$\theta(\mathbf{x}, \tau) = -\frac{\mathrm{d}D(\tau)}{\mathrm{d}\tau}\left(\theta^{(1)}(\mathbf{x}) + D(\tau)\theta^{(2)}(\mathbf{x}) + \dots\right) \tag{2.7}$$

where

$$\tilde{\theta}^{(n)}(\mathbf{k}) = \left[\prod_{i=1}^{n} \int \frac{\mathrm{d}^3 k_i}{(2\pi)^3}\right] (2\pi)^3 \delta_{\mathrm{D}}^{(3)}\left(\mathbf{k} - \sum_{i=1}^{n} \mathbf{k}_i\right) G_n(\mathbf{k}_1, \cdots, \mathbf{k}_n)\tilde{\delta}_\rho^{(1)}(\mathbf{k}_1) \cdots \tilde{\delta}_\rho^{(1)}(\mathbf{k}_n) \tag{2.8}$$

and $G_1 = 1$. In this chapter, we only need the explicit expressions [23] for F_2, G_2 and F_3, which are

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{1}{2}\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}, \tag{2.9}$$

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{3}{7} + \frac{1}{2}\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{4}{7}\frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}, \tag{2.10}$$

$$\begin{aligned} F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = &\frac{2k_{123}^2}{54}\left[\frac{\mathbf{k}_1 \cdot \mathbf{k}_{23}}{k_1^2 k_{23}^2} G_2(\mathbf{k}_2, \mathbf{k}_3) + (2 \text{ cyclic})\right] \\ &+\frac{7}{54}\mathbf{k}_{123} \cdot \left[\frac{\mathbf{k}_{12}}{k_{12}^2} G_2(\mathbf{k}_1, \mathbf{k}_2) + (2 \text{ cyclic})\right] \\ &+\frac{7}{54}\mathbf{k}_{123} \cdot \left[\frac{\mathbf{k}_1}{k_1^2} F_2(\mathbf{k}_2, \mathbf{k}_3) + (2 \text{ cyclic})\right], \end{aligned} \tag{2.11}$$

where $\mathbf{k}_{ijl\ldots} = \mathbf{k}_i + \mathbf{k}_j + \mathbf{k}_l + \cdots$.

The matter overdensity perturbation field at linear order $\tilde{\delta}_\rho^{(1)}$ is related to the primordial curvature fluctuation $\tilde{\delta}_\zeta$ by

$$\tilde{\delta}_\rho^{(1)}(\mathbf{k}) = M(k)\tilde{\delta}_\zeta(\mathbf{k}) \tag{2.12}$$

where for small wavevectors $M(k) \propto k^2$, and for large wavevectors (below the nonlinear scale) $M(k) \sim \sqrt{k}$ [17]. For the purposes of analytical estimates, we will use the following approximate expression for $M(k)$

$$M(k) \simeq \left(\frac{2}{5\Omega_m}\right)\left[\frac{k^2}{H_0^2}\theta(q_0 - k) + \sqrt{\frac{k}{q_0}}\frac{q_0^2}{H_0^2}\theta(k - q_0)\right]. \tag{2.13}$$

Here $H_0 \simeq 70\,\mathrm{km\,s^{-1}Mpc^{-1}}$ is the Hubble constant today, $\Omega_m \simeq 0.3$ is the fraction of matter energy density, and $q_0 \simeq 0.015\,\mathrm{Mpc}^{-1}$.

Matter overdensity correlators are computed using the Effective Field Theory of Large Scale Structure (EFT-of-LSS) [24, 25]. We work only to leading order in this theory.

Galaxy Density Perturbations and Regularization Schemes

As mentioned in the Introduction, the galaxy overdensity field δ_g can be written as a biased tracer of the underlying matter overdensity field δ_ρ

$$\delta_g(\mathbf{x}) = b_1\delta_\rho(\mathbf{x}) + \frac{b_2}{2}\delta_\rho^2(\mathbf{x}) + \ldots . \tag{2.14}$$

In the equation above we omitted operators that are outside the scope of this chapter as well as the terms that set the expectation value of δ_g to zero. The general form of the additional terms represented by the ellipses in the bias expansion above is known [5] (even for different forms of the power spectrum) and it could include other operators that do not appear as counterterms [26].

The composite operators in Eq. (2.14) need to be regularized and renormalized. In this chapter we will focus on the operator δ_ρ^2 as an illustration of the dependence of the counterterms on the regularization scheme. What we conclude could be generalized to other composite operators in the galaxy bias expansion. Generically, regularization is achieved by introducing a wavevector cutoff Λ in divergent integrals. Here we present three possible choices for the cutoff regulator.

The first regularization scheme we consider consists in cutting off the large wavevector component of the linearized solution $\delta_\rho^{(1)}$ of the gravitational evolution equations, i.e. performing the following replacement

$$\delta_\rho^{(1)}(\mathbf{q}) \to \delta_\rho^{(1)}(\mathbf{q})\theta(\Lambda - q)\,. \tag{2.15}$$

As we will discuss in Sec. 2.3 this prescription reproduces the results previously found in [18] after assuming the UV asymptotic behavior for M given in Eq. (2.13). This prescription has the advantage that, had we worked at higher order in the EFT-of-LSS, it would have regulated both the correlators of δ_ρ and of the composite operator δ_ρ^2.

Since we are working to lowest order in the EFT-of-LSS we introduce a second and a third renormalization prescriptions for δ_ρ^2 that do not explicitly cut off large wavevectors in the non-linear gravitational evolution equations.

One choice is to cut off the large wavevectors of the *full* solution δ_ρ of the gravitational equations, i.e.

$$\delta_\rho(\mathbf{q}) \to \delta_\rho(\mathbf{q})\theta(\Lambda - q)\,, \tag{2.16}$$

which implies

$$\tilde{\delta}_\rho^2(\mathbf{k}) \to \int \frac{d^3q}{(2\pi)^3}\tilde{\delta}_\rho(\mathbf{q})\tilde{\delta}_\rho(\mathbf{k}-\mathbf{q})\theta(\Lambda - q)\theta(\Lambda - |\mathbf{k}-\mathbf{q}|)\,. \tag{2.17}$$

The third prescription is obtained from the second one by expanding for $k \ll q$ one of the two thetas in the convolution integral of δ_ρ^2 and keeping only the leading term.

$$\begin{aligned}\tilde{\delta}_\rho^2(\mathbf{k}) &\to \int \frac{\mathrm{d}^3 q}{(2\pi)^3} \tilde{\delta}_\rho(\mathbf{q}) \tilde{\delta}_\rho(\mathbf{k}-\mathbf{q}) \theta(\Lambda - q) \\ &= \int \frac{\mathrm{d}^3 q}{(2\pi)^3} \tilde{\delta}_\rho(\mathbf{q}) \tilde{\delta}_\rho(\mathbf{k}-\mathbf{q}) \left[\frac{\theta(\Lambda - q) + \theta(\Lambda - |\mathbf{k}-\mathbf{q}|)}{2} \right] . \end{aligned} \tag{2.18}$$

In the square bracket of the above equation we explicitly symmetrized the regularization kernel, thus rendering the expression manifestly symmetric in the two $\tilde{\delta}_\rho$'s. This last prescription leads to a spherically symmetric integration region (hence simplifying the integrals).

We find that all three regulators give the same non-boost invariant terms[2] and that they can be removed by expressing φ_G in Lagrangian coordinates as expected from the results of [18, 27]. After assuming a UV asymptotic behavior for M (e.g. the one in Eq. (2.13)), the second and third regulators give the same coefficients for the boost invariant terms which, however, differ from the ones in the first prescription. Hence, in the following, we will only discuss the first and the third regularization schemes.

We will deal with infrared divergences that arise in the computation of correlators of the galaxy overdensity field by expanding the integrands around $q = \infty$ and retaining only the terms that contribute to the UV divergence. These terms are infrared safe.

2.3 Renormalization of the operator δ_ρ^2 at one loop

The correlators of δ_g are sensitive to the physics of large wavevectors where perturbation theory is no longer valid. To make analytic predictions using a perturbative approach, composite operators in the galaxy bias expansion need to be regularized and renormalized adding counterterms that remove the sensitivity to the physics of large wavevectors. Following [15], we only keep the fastest growing modes in the computation of the counterterms.

Note that the composite operator δ_ρ^2 can ultimately be written as a convolution of fields φ_G at different wavevectors. Therefore, we find the one-loop counterterms by contracting two such fields, introducing a UV regulator Λ (as described in Sec. 2.2), and by selecting the parts that diverge as Λ goes to infinity.

Even though all the amplitudes presented in this chapter can be obtained applying Wick's theorem, such operation can be tedious and cumbersome. Therefore, we will

[2]In this chapter we refer to operators that diverge as a single wave-vector vanishes as non-boost invariant.

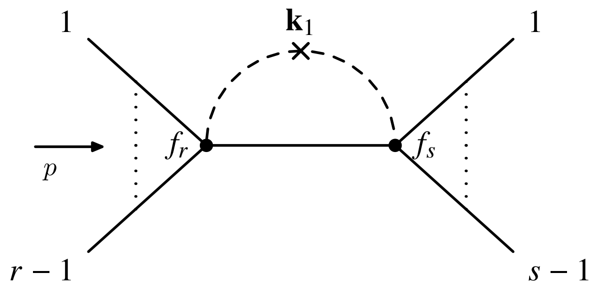


Figure 2.1: Generic one-loop divergent diagrams at first order in gravitational evolution.

use a diagrammatic formalism to graphically keep track of the different contributions in the renormalization of the operator δ_ρ^2 (see Appendix 2.B for details). Diagrams with N cross vertices represent contributions to the N-point galaxy overdensity field correlator $\langle \tilde{\delta}_g(\mathbf{k}_1) \ldots \tilde{\delta}_g(\mathbf{k}_N) \rangle$.

In this section, we discuss the one-loop renormalization of the quadratic operator δ_ρ^2 for local primordial non-Gaussianity. At the end of the section, we give the full renormalized expression of the operator δ_ρ^2 up to third order in gravitational evolution in the first and the third renormalization schemes introduced above. Using the first regularization prescription, we reproduce the results found in [18] after assuming the UV asymptotic behavior for M of Eq. (2.13).

First Order in Gravitational Evolution

In this subsection, we work at linear order in gravitational evolution, i.e. we set all the F_n's and G_n's (for $n > 1$) to zero. We compute the counterterms needed to renormalize the operator δ_ρ^2 at one loop in the presence of primordial non-Gaussian perturbations of the form of Eq. (2.1). In Fig. 2.1, we show all the (amputated) one-loop divergent diagrams with a single insertion of δ_ρ^2. These diagrams are obtained by contracting a single φ_G from each δ_ρ in δ_ρ^2.

Here, the first regularization prescription gives

$$(2-\delta_{rs}) r s f_r f_s \int \frac{d^3 q}{(2\pi)^3} M(q) M(|\mathbf{k}_1-\mathbf{q}|) P_\varphi(|\mathbf{p}+\mathbf{q}|) \theta(\Lambda-q)\theta(\Lambda-|\mathbf{k}_1-\mathbf{q}|) \quad (2.19)$$

where $\mathbf{p}$ is the total wavevector entering the f_r vertex. We evaluate this kind of integral in cylindrical coordinates where the product of the two thetas can naturally be embedded in the integration measure.

The third regularization prescription gives

$$(2-\delta_{rs}) rs f_r f_s \sigma^2(\mathbf{k}_1, \mathbf{p}; \Lambda) \tag{2.20}$$

where

$$\begin{aligned}\sigma^2(\mathbf{k}_1, \mathbf{p}; \Lambda) = \int \frac{\mathrm{d}^3 q}{(2\pi)^3} M(q)\, M(|\mathbf{k}_1 - \mathbf{q}|)\, P_\varphi(|\mathbf{p}+\mathbf{q}|) \\ \times \left[\frac{\theta(\Lambda - q) + \theta(\Lambda - |\mathbf{k}_1 - \mathbf{q}|)}{2}\right] .\end{aligned} \tag{2.21}$$

We embed the thetas in the integration measure and we expand the remaining part of the integrand around $q = \infty$, using the large wavevector asymptotic expression of Eq. (2.13) for M. We keep only the powers that give rise to UV divergent terms (i.e. all the terms up to q^{-2}). Note that terms in the integrand suppressed by powers of k/q give rise to finite contributions that are dropped. These terms do not contain IR or collinear divergences. After this procedure the divergent parts of the integral in Eq. (2.19) are

$$\sigma^2_{\text{asy}} \equiv \frac{4}{25} \frac{A q_0^3}{\Omega_m^2 H_0^4} \frac{\Lambda}{2\pi^2} . \tag{2.22}$$

Due to the spherical symmetry of the integration region of Eq. (2.21), for the third prescription an alternative procedure can be followed. Expanding the integrand for $k_1, p \ll q$ and performing the angular integral $\mathrm{d}\Omega_q$, the linearly divergent part of Eq. (2.21) is

$$\sigma^2 \equiv \sigma^2(0, 0; \Lambda) = \int \frac{\mathrm{d}^3 q}{(2\pi)^3} M^2(q) P(q) \theta(\Lambda - q) . \tag{2.23}$$

σ^2 coincides with σ^2_{asy} when using the UV asymptotic expression for M of Eq. (2.13).

We therefore introduce the following position space counterterms to make correlators involving the operator $\tilde{\delta}^2_\rho$ finite as Λ goes to infinity

$$c_{r-1,s-1}(\Lambda) : \varphi_G^{r-1}(\mathbf{x}) :: \varphi_G^{s-1}(\mathbf{x}) : \tag{2.24}$$

where

$$c_{r-1,s-1}(\Lambda) = -rs f_r f_s \sigma^2_{\text{asy}} \tag{2.25}$$

and σ^2 can be used in place of σ^2_{asy}. At one loop the two normal orderings in Eq. (2.24) can be merged to a single overall one, and consequently the counterterms take the form

$$c_n(\Lambda) : \varphi_G^n(\mathbf{x}) : = \left(\sum_{i=0}^{n} c_{i,n-i}(\Lambda)\right) : \varphi_G^n(\mathbf{x}) : \; . \tag{2.26}$$

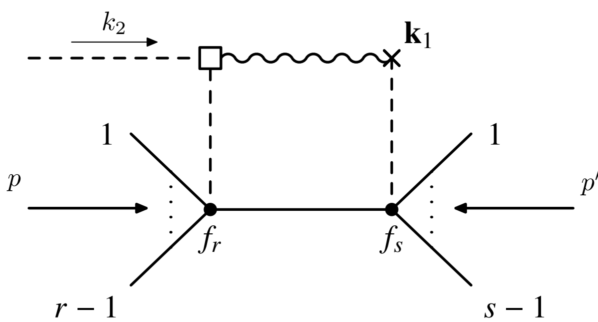


Figure 2.2: Diagrams with a single F_2 vertex used to determine the counterterm for δ_ρ^2 linear in δ_ρ.

As mentioned above the first and the third prescriptions give the same counterterms which up to second order in the field φ_G are

$$-\sigma^2_{\rm asy}\left[4f_2\,\varphi_G(\mathbf{x})+(6f_3+4f_2^2)\;:\varphi_G^2(\mathbf{x}):\right]\,. \tag{2.27}$$

Second Order in Gravitational Evolution

We now study the effect of non-linear gravitational evolution on renormalization for non-Gaussian primordial fluctuations of the local type. In this case the renormalization of δ_ρ^2 will generate counterterms linear in $\delta_\rho^{(1)}$ (which will reassemble to δ_ρ when higher order terms are included).

We begin by considering Gaussian primordial fluctuations (see Fig. 2.2 choosing $r = 1$ and $s = 1$) and we evaluate its divergent coefficient in the first and third regularization prescriptions described in Sec. 2.2.

The first prescription gives

$$4\int\frac{\mathrm{d}^3q}{(2\pi)^3}F_2(\mathbf{q},-\mathbf{k}_1)M^2(q)P_\varphi(q)\theta(\Lambda-q)\,, \tag{2.28}$$

where we dropped thetas of the type $\theta(\Lambda - k_i)$ since $k_i < \Lambda$. Expanding for small k_1/q and performing the angular integral $\mathrm{d}\Omega_q$, the linearly divergent part of the above equation is

$$\frac{68}{21}\int\frac{\mathrm{d}q}{2\pi^2}q^2M^2(q)P_\varphi(q)\theta(\Lambda-q)=\frac{68}{21}\sigma^2\simeq\frac{68}{21}\sigma^2_{\rm asy}\,, \tag{2.29}$$

where in the last step we assumed the UV asymptotic behavior of Eq.(2.13) for M. This result reproduces the counterterms already found in the literature [17].

For the third prescription, $\tilde\delta_\rho^2$ at this order is

$$\int\frac{\mathrm{d}^3q}{(2\pi)^3}\left[\tilde\delta_\rho^{(2)}(\mathbf{q})\tilde\delta_\rho^{(1)}(\mathbf{k}_1-\mathbf{q})+\tilde\delta_\rho^{(1)}(\mathbf{q})\tilde\delta_\rho^{(2)}(\mathbf{k}_1-\mathbf{q})\right]\theta(\Lambda-q)\,. \tag{2.30}$$

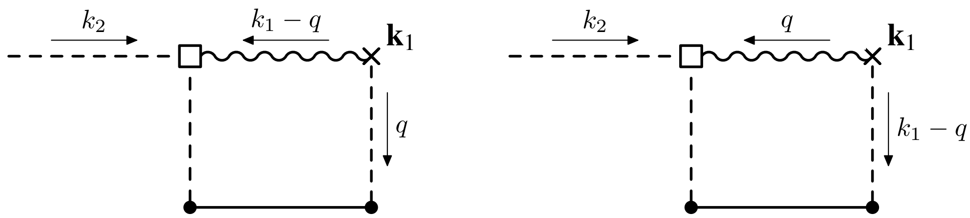


Figure 2.3: Two routings of the loop diagram with a single F_2 vertex used to determine the counterterm for δ_ρ^2 linear in δ_ρ with Gaussian primordial fluctuations.

Contracting one of the $\tilde{\varphi}_G$ fields in $\tilde{\delta}_\rho^{(2)}$ with the $\tilde{\varphi}_G$ field in $\tilde{\delta}_\rho^{(1)}$ gives an operator proportional to $\tilde{\delta}_\rho^{(1)}(\mathbf{k}_1)$ with (divergent) coefficient

$$\begin{aligned} &2\int \frac{\mathrm{d}^3 q}{(2\pi)^3}\left[F_2(-\mathbf{q},\mathbf{k}_1)M^2(q)P_\varphi(q)\right. \\ &\left.+F_2(-\mathbf{k}_1+\mathbf{q},\mathbf{k}_1)\, M^2(|\mathbf{k}_1-\mathbf{q}|)\, P_\varphi(|\mathbf{k}_1-\mathbf{q}|)\right]\theta(\Lambda - q)\,. \end{aligned} \tag{2.31}$$

In diagrammatic notation, the two terms above are represented by the (amputated) diagrams of Fig. 2.3 and correspond to two ways of *routing* the loop wavevector. We observe that the two terms in the square bracket can be made equal to each other with the change of variable $\mathbf{q} \to \mathbf{k}_1 - \mathbf{q}$, however the *integrands* are not equal because of the presence of the theta function. Thus, we expect the above integrals to differ by a surface term which will impact the explicit form of the counterterms.

Expanding for small k_1/q and performing the angular integral $\mathrm{d}\Omega_q$, the linearly divergent parts of Eq. (2.31) are

$$\begin{aligned} \int \frac{\mathrm{d}q}{2\pi^2}&\left\{\frac{34}{21}q^2M^2(q)P_\varphi(q)+\frac{34}{21}q^2M^2(q)P_\varphi(q)\right. \\ &\left.-\frac{1}{3}\frac{\mathrm{d}}{\mathrm{d}q}\left(q^3M^2(q)P_\varphi(q)\right)\right\}\theta(\Lambda-q) = \frac{68}{21}\sigma^2-\frac{1}{3}\rho^2\,. \end{aligned} \tag{2.32}$$

where

$$\rho^2 \quad = \int \frac{\mathrm{d}q}{2\pi^2}\left[\frac{\mathrm{d}}{\mathrm{d}q}\left(q^3M^2(q)P_\varphi(q)\right)\right]\theta(\Lambda-q)\,. \tag{2.33}$$

On the left hand side of Eq. (2.32), the first term corresponds to the first routing while the other two terms correspond to the second routing. Notice that ρ^2 coincides with σ^2_{asy} when using the UV asymptotic expression for M of Eq. (2.13). As anticipated, the contributions from the two routings differ by a total derivative term. Since this term is linearly divergent, it will contribute to the value of the counterterm. We observe that the first term (bulk) on the right hand side of Eq. (2.32) is the same as the counterterm for the first prescription of Eq. (2.28). In loops containing box vertices (representing non-linear gravitational evolution), the third prescription - in diagrammatic formalism - entails symmetrizing the contribution of a diagram over the two possible routings in analogy to Fig. 2.3.

Non-Gaussian primordial fluctuations generate counterterms proportional to the operators $\delta_\rho^{(1)}\varphi_G^n$. To determine their divergent coefficients we evaluate the diagrams in Fig. 2.2 in the regularization prescriptions described above. For each choice of r and s (s.t. $r+s=n$) we need to sum over two diagrams. Letting $\mathbf{p}$ and $\mathbf{p}'$ be the sum of the wavevectors entering the f_r and f_s vertices, respectively, the total amplitude is

$$\begin{aligned}2\,rsf_rf_s\int\frac{\mathrm{d}^3q}{(2\pi)^3}\,F_2(\mathbf{k}_1+\mathbf{k}_2-\mathbf{q},-\mathbf{k}_2)\,M(q)\,M(|\mathbf{k}_1+\mathbf{k}_2-\mathbf{q}|)\\ \times\left[P_\varphi(|\mathbf{p}+\mathbf{q}|)+P_\varphi(|\mathbf{p}'+\mathbf{q}|)\right]\mathcal{R}_i\,,\end{aligned} \tag{2.34}$$

where the $\mathcal{R}_i$ is the regularization kernel in scheme i and

$$\mathcal{R}_1 = \theta(\Lambda-q)\theta(\Lambda-|\mathbf{k}_1+\mathbf{k}_2-\mathbf{q}|)\,, \tag{2.35}$$

$$\mathcal{R}_3 = \frac{\theta(\Lambda-q)+\theta(\Lambda-|\mathbf{k}_1-\mathbf{q}|)}{2}\,. \tag{2.36}$$

Proceeding as before, we obtain for the divergent parts

$$2\,rsf_rf_s\left(\frac{68}{21}+\frac{\mathbf{k}_2\cdot(\mathbf{p}+\mathbf{p}')}{k_2^2}\right)\sigma_{\text{asy}}^2\,, \tag{2.37}$$

$$2\,rsf_rf_s\left[\left(\frac{68}{21}+\frac{\mathbf{k}_2\cdot(\mathbf{p}+\mathbf{p}')}{k_2^2}\right)\sigma^2-\frac{1}{3}\rho^2\right]\,. \tag{2.38}$$

We observe that within each regularization prescription presented in this chapter, the Gaussian and non-Gaussian coefficients of the k-independent parts coincide [18] up to a factor of $2rsf_rf_s$ (compare Eqs. (2.29) and (2.32) with Eqs. (2.37) and (2.38)). However, as expected [27], in the non-Gaussian case non-boost invariant terms proportional to $\mathbf{k}_2\cdot(\mathbf{p}+\mathbf{p}')/k_2^2$ appear. These shift the argument of the field φ_G from its Eulerian coordinate $\mathbf{x}$ to its initial Lagrangian position $\mathbf{x}_0$ (see Eq. (2.48)) and are the same in the three regularization schemes studied in this chapter.

Finally we notice that the two routings in the third prescription generate equal and opposite surface terms that are non-boost invariant. For example, for the counterterm

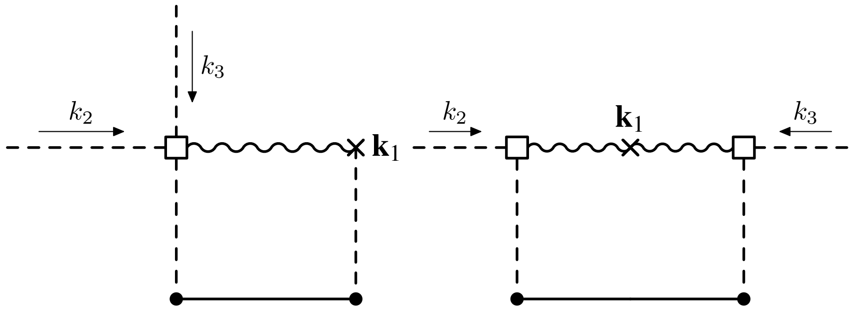


Figure 2.4: Diagrams with one F_3 and two F_2's used to determine the counterterms for δ_ρ^2 quadratic in δ_ρ with Gaussian primordial fluctuations.

proportional to $\delta_\rho^{(1)}\varphi_G$ we get

$$4f_2 \int \frac{\mathrm{d}q}{2\pi^2} \Bigg\{ \left(\frac{34}{21} + \frac{\mathbf{k}_2\cdot\mathbf{k}_3}{2k_2^2}\right) q^2 M^2(q) P_\varphi(q) - \left(\frac{\mathbf{k}_2\cdot\mathbf{k}_3}{6k_2^2} + \frac{1}{3}\right) \frac{\mathrm{d}}{\mathrm{d}q}\left(q^3 M^2(q) P_\varphi(q)\right) + \left(\frac{34}{21} + \frac{\mathbf{k}_2\cdot\mathbf{k}_3}{2k_2^2}\right) q^2 M^2(q) P_\varphi(q) + \frac{\mathbf{k}_2\cdot\mathbf{k}_3}{6k_2^2} \frac{\mathrm{d}}{\mathrm{d}q}\left(q^3 M^2(q) P_\varphi(q)\right) \Bigg\} \theta(\Lambda - q)\,. \tag{2.39}$$

where, in this case $\mathbf{p} = \mathbf{k}_3$ and $\mathbf{p}' = 0$, and the two lines correspond to the two possible routings of the loop wavevector. Again we note that the non-boost invariant surface terms proportional to $(\mathbf{k}_2\cdot\mathbf{k}_3)/k_2^2$ cancel when considering both routings.

Third Order in Gravitational Evolution

We now consider the one-loop counterterms at second order in the field $\delta_\rho^{(1)}$. We have two different diagram topologies (the Gaussian ones are shown in Fig. 2.4). The amplitudes for the two topologies at arbitrary order in primordial local non-Gaussianity are

$$12\, r s f_r f_s \int \frac{\mathrm{d}^3 q}{(2\pi)^3}\, F_3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{q}, -\mathbf{k}_2, -\mathbf{k}_3)\, M(q) \times M(|\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{q}|) \times \left[P_\varphi(|\mathbf{q}+\mathbf{p}|) + P_\varphi(|\mathbf{q}+\mathbf{p}'|)\right] \mathcal{R}_i^{(a)}\,. \tag{2.40}$$

and

$$16\,rs f_r f_s \int \frac{d^3q}{(2\pi)^3}\, F_2(\mathbf{q}+\mathbf{k}_2,-\mathbf{k}_2)\, F_2(\mathbf{k}_1+\mathbf{k}_3-\mathbf{q},-\mathbf{k}_3)\, M(|\mathbf{q}+\mathbf{k}_2|) \times M(|\mathbf{k}_1+\mathbf{k}_3-\mathbf{q}|) \times \left[P_\varphi(|\mathbf{q}+\mathbf{p}+\mathbf{k}_2|)+P_\varphi(|\mathbf{q}+\mathbf{p}'+\mathbf{k}_2|)\right]\mathcal{R}_i^{(b)}\,. \tag{2.41}$$

In the equations above,

$$\mathcal{R}_1^{(a)} = \theta(\Lambda-q)\theta(\Lambda-|\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3-\mathbf{q}|)\,, \tag{2.42}$$

$$\mathcal{R}_1^{(b)} = \theta(\Lambda-|\mathbf{q}+\mathbf{k}_2|)\theta(\Lambda-|\mathbf{k}_1+\mathbf{k}_3-\mathbf{q}|)\,, \tag{2.43}$$

and $\mathcal{R}_3^{(a)}=\mathcal{R}_3^{(b)}$ are given in Eq. (2.36).

We now give the full renormalized expression of the operator δ_ρ^2 up to third order in the gravitational evolution for local primordial non-Gaussianity. For completeness we reintroduce the subtraction of the vacuum expectation value $\langle\delta_\rho^2\rangle$.

For the first prescription we have

$$\delta_\rho^2(\mathbf{x})\big|_R = \delta_\rho^2(\mathbf{x}) - \sigma^2_{\mathrm{asy}}\mathcal{S}[\varphi_G]\left[1+\frac{68}{21}\delta_\rho(\mathbf{x})+\frac{2624}{735}\delta_\rho^2(\mathbf{x}) +\frac{254}{2205}\left(\frac{\partial_i\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)\left(\frac{\partial_i\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)\right]\,. \tag{2.44}$$

which reproduces the results of [18] after using the UV asymptotic expression σ^2_{asy} in place of σ^2 for the Gaussian cases.

When using the third prescription we obtain

$$\begin{aligned}\delta_\rho^2(\mathbf{x})\big|_R = \delta_\rho^2(\mathbf{x}) - \mathcal{S}[\varphi_G]\,\Big[\sigma^2 &+\left(\frac{68}{21}\sigma^2-\frac{1}{3}\rho^2\right)\delta_\rho(\mathbf{x})\\ &+\left(\frac{2624}{735}\sigma^2-\frac{73}{105}\rho^2+\frac{1}{30}\gamma^2\right)\delta_\rho^2(\mathbf{x})\\ &+\left(\frac{254}{2205}\sigma^2-\frac{16}{105}\rho^2+\frac{1}{15}\gamma^2\right)\left(\frac{\partial_i\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)\left(\frac{\partial_i\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)\Big]\,.\end{aligned} \tag{2.45}$$

where

$$\gamma^2 = \int \frac{dq}{2\pi^2}\left\{\frac{d}{dq}\left[q^4\frac{d}{dq}\left(M^2(q)P_\varphi(q)\right)\right]\right\}\theta(\Lambda-q). \tag{2.46}$$

In the above equations

$$\mathcal{S}[\varphi_G] = 1+4f_2\varphi_G(\mathbf{x}_0)+(6f_3+4f_2^2):\varphi_G^2(\mathbf{x}_0):+\cdots \tag{2.47}$$

where, at this order in the gravitational evolution,

$$\varphi_G(\mathbf{x}_0) = \varphi_G(\mathbf{x}) + \left(\frac{\partial_\ell}{\nabla^2}\delta_\rho(\mathbf{x})\right)\partial_\ell\varphi_G(\mathbf{x}) - \frac{1}{2}\frac{\partial_\ell}{\nabla^2}\left\{\delta_\rho^2(\mathbf{x}) - \left(\frac{\partial_i\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)^2\right\}\partial_\ell\varphi_G(\mathbf{x})$$
$$+\frac{1}{2}\left(\frac{\partial_i}{\nabla^2}\delta_\rho(\mathbf{x})\right)\left(\frac{\partial_j}{\nabla^2}\delta_\rho(\mathbf{x})\right)\partial_i\partial_j\varphi_G(\mathbf{x}) \tag{2.48}$$

and the ellipses in Eq. (2.47) can be deduced from Eq. (2.25) and Eq. (2.26). In Eq.(2.47) the one corresponds to the Gaussian case and the remaining terms arise from primordial non-Gaussianity. The relationship between the counterterms in the Gaussian and non-Gaussian cases was first noted in [18].

2.4 Concluding Remarks

Composite operators in the bias expansion for the galaxy overdensity field are defined with a large wavevector cutoff (Λ) and are renormalized to remove the dependence of galaxy overdensity correlators on the wavevector cutoff. In this chapter we explored the regularization scheme dependence of the counterterms in the Galaxy bias expansion. We showed by explicit computation how different regularization prescriptions affect the coefficients of the counterterms. As expected the coefficients of the non-boost invariant operators coincide for the regularization schemes explored in this chapter and they rearrange to shift the argument of the field φ_G from Eulerian to Lagrangian coordinates. On the other hand, the boost-invariant terms are dependent on the regularization scheme.

Appendix 2.A Conformal Invariance of δ_ζ correlators at tree-level

In this Appendix, we prove that in local non-Gaussianity all the tree-level correlators of the curvature fluctuations δ_ζ are conformally invariant.

We start by introducing a diagrammatic notation for the N-point correlators $P^{(N)}$ defined as

$$\langle\tilde{\delta}_\zeta(\mathbf{k}_1)\dots\tilde{\delta}_\zeta(\mathbf{k}_N)\rangle = (2\pi)^3\delta(\mathbf{k}_1+\cdots+\mathbf{k}_N)P^{(N)}(\mathbf{k}_1,\dots,\mathbf{k}_N)\,. \tag{2.49}$$

We separate the sum of all the tree-level contributions ($T^{(N)}$) from the sum of the loop contributions using the following notation

$$P^{(N)}(\mathbf{k}_1,\dots,\mathbf{k}_N) = T^{(N)}(\mathbf{k}_1,\dots,\mathbf{k}_N) + \text{loops}\,, \tag{2.50}$$

and we shall refer to $T^{(N)}$ as the tree-level correlator in wavevector space.

Figure 2.5: Generic chain diagram.

Using δ_ζ as in Eq. (2.1), we first consider the case where only f_1 and f_2 are non-zero. Then, it is straightforward to see that the tree-level N-point correlator is

$$T^{(N)}(\mathbf{k}_1, \ldots, \mathbf{k}_N) = 2^{N-2} A^{N-1} f_1^2 f_2^{N-2} \frac{1}{2} \sum_{\mathcal{P}(\mathbf{k}_1 \ldots \mathbf{k}_N)} \frac{1}{k_1^3} \frac{1}{|\mathbf{k}_1 + \mathbf{k}_2|^3} \times \cdots \frac{1}{|\mathbf{k}_1 + \cdots + \mathbf{k}_{N-1}|^3}\,. \tag{2.51}$$

where the sum is over all the permutations $\mathcal{P}(\mathbf{k}_1 \ldots \mathbf{k}_N)$ of the wavevectors. The diagram in Figure 2.5 represents the identity permutation contribution to $T^{(N)}$ above.

More generally, contributions to the N-point function can be represented by diagrams with N vertices (dots) connected by solid lines.

To compute the connected tree-level N-point function of $\tilde{\delta}_\zeta$, we first draw all the connected tree diagrams with N vertices, i.e. different topologies. For each diagram we arbitrarily assign the wavevectors $\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_N$ to the various vertices.

We then use the following rules to compute the contribution to $T^{(N)}$ of each diagram.

- Every vertex with n lines emerging from it contributes a factor of f_n.
- Every line between two vertices contributes a factor $\dfrac{A}{|\sum_i \mathbf{k}_i|^3}$ where the sum is over the wavevectors that precede and include either of the two vertices along the tree (these two choices are equivalent due to overall wavevector conservation).
- Every diagram has a combinatorial factor that is given by the number of ways in which the lines from the various vertices can be joined together.
- Every diagram has a symmetry factor that is given by the inverse of the number of symmetries the diagram has. For example, the diagram in Fig. 2.5 has a symmetry factor of $1/2$ because of the reflection symmetry of the chain around its center while the second diagram in Fig. 2.6 has a symmetry factor of $1/3!$.

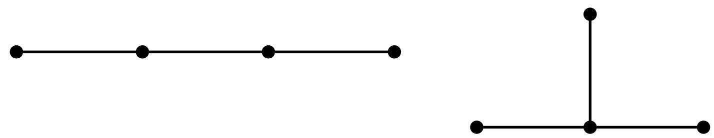


Figure 2.6: Contributions to $T^{(4)}$.

Finally, we sum over the $N!$ permutations of the wavevectors $\mathcal{P}(\mathbf{k}_1, \ldots, \mathbf{k}_N)$.

For example, the full expression for $T^{(4)}$ is given below and corresponds to the two diagrams in Fig. 2.6, where we didn't label the vertices since the sum over permutations is implied

$$\begin{aligned} T^{(4)} = 2^2 A^3 f_1^2 f_2^2 \frac{1}{2} \sum_{\mathcal{P}(\mathbf{k}_1,\ldots,\mathbf{k}_4)} \frac{1}{k_1^3 k_2^3 |\mathbf{k}_1 + \mathbf{k}_2|^3} \\ + A^3 f_1^3 f_3\, 3!\, \frac{1}{3!} \sum_{\mathcal{P}(\mathbf{k}_1,\ldots,\mathbf{k}_4)} \frac{1}{k_1^3 k_2^3 k_4^3}\,. \end{aligned} \tag{2.52}$$

It is convenient to introduce the following notation

$$\begin{aligned} T^{(N)}(\mathbf{k}_1, \ldots, \mathbf{k}_N) &= \sum_{i \in \left\{ \substack{\text{topologies} \\ \text{with } N \\ \text{vertices}} \right\}} T_i^{(N)}(\mathbf{k}_1, \ldots, \mathbf{k}_N) \\ &= \sum_{i \in \left\{ \substack{\text{topologies} \\ \text{with } N \\ \text{vertices}} \right\}} \sum_{\mathcal{P}(\mathbf{k}_1,\ldots,\mathbf{k}_N)} t_i^{(N)}(\mathbf{k}_1, \ldots, \mathbf{k}_N)\,. \end{aligned} \tag{2.53}$$

where we split each tree-level N-point function into a sum of terms from different topologies each of which is further expressed as a sum over the $N!$ permutations $[\mathcal{P}(\mathbf{k}_1, \ldots, \mathbf{k}_N)]$ of the external wavevectors.

Most inflationary models predict almost scale invariant δ_ζ correlators (with conformal weight $\Delta = 0$), some of which are also conformally invariant [28].

In Primordial local Non-Gaussianity, δ_ζ is a function of powers of φ_G only (and not its derivatives). Being that the two point function of the Gaussian field φ_G is scale invariant, we expect δ_ζ correlators to be scale invariant as well. In the following, we demonstrate that the conformal invariance of any tree-level N-point correlator of δ_ζ follows from the conformal invariance of the $N = 2$ correlator of φ_G. Moreover, we explicitly show that this holds separately for each topology and each permutation. Conformal invariance of the correlators implies that the conformal Ward identities are satisfied,

$$\left[\left(\sum_{\substack{j=1\\ j\neq\alpha}}^{N}\mathcal{K}_j\right)t_i^{(N)}(\mathbf{k}_1,\ldots,\bar{\mathbf{k}}_\alpha,\ldots,\mathbf{k}_N)\right]_{\bar{\mathbf{k}}_\alpha=-\sum_{\beta\neq\alpha}\mathbf{k}_\beta} = 0\,, \tag{2.54}$$

where

$$\mathcal{K}_j \equiv 2(\Delta-3)\nabla_j - 2(\mathbf{k}_j\cdot\nabla_j)\nabla_j + \mathbf{k}_j\nabla_j^2\,. \tag{2.55}$$

Here, $\mathcal{K}_j$ are the generators of special conformal transformations in wavevector space and $\nabla_j \equiv \partial_{\mathbf{k}_j}$. In our case Δ is equal to zero.

We choose the α'th wavevector to be the dependent one and we use the overall wavevector conservation to express that wavevector in terms of the others. Any tree-level $(N+1)-$point diagram can be constructed from an $N-$point tree-level diagram by adding a line to one of its N vertices. Labelling this vertex with the wavevector $\mathbf{k}_\alpha$ and choosing it to be the dependent one as before, we have the recursion relation

$$t_{i'}^{(N+1)}(\mathbf{k}_1,\ldots,\mathbf{k}_\alpha,\ldots,\mathbf{k}_{N+1}) = \frac{f_{m(\alpha)+1}}{f_{m(\alpha)}}t_i^{(N)}(\mathbf{k}_1,\ldots,\mathbf{k}_\alpha,\ldots,\mathbf{k}_N)\frac{A}{|\mathbf{k}_{N+1}|^3} \tag{2.56}$$

where $m(\alpha)$ is the number of legs emerging from the α'th vertex in $t_i^{(N)}$. Therefore, the conformal Ward identities for the $t_{i'}^{(N+1)}$ diagram are

$$\begin{aligned}
&\left[\left(\sum_{j=1,\,j\neq\alpha}^{N+1}\mathcal{K}_j\right)t_{i'}^{(N+1)}(\mathbf{k}_1,\ldots,\bar{\mathbf{k}}_\alpha,\ldots,\mathbf{k}_{N+1})\right]_{\bar{\mathbf{k}}_\alpha=-\sum_{\beta\neq\alpha}\mathbf{k}_\beta} \\
= \;& \frac{A}{|\mathbf{k}_{N+1}|^3}\left[\frac{f_{m(\alpha)+1}}{f_{m(\alpha)}}\left(\sum_{j=1,\,j\neq\alpha}^{N}\mathcal{K}_j\right)t_i^{(N)}(\mathbf{k}_1,\ldots,\bar{\mathbf{k}}_\alpha,\ldots,\mathbf{k}_N)\right]_{\bar{\mathbf{k}}_\alpha=-\sum_{\beta\neq\alpha}\mathbf{k}_\beta} + \\
+ \;& \mathcal{K}_{N+1}\left(\frac{A}{|\mathbf{k}_{N+1}|^3}\right)\left[\frac{f_{m(\alpha)+1}}{f_{m(\alpha)}}t_i^{(N)}(\mathbf{k}_1,\ldots,\bar{\mathbf{k}}_\alpha,\ldots,\mathbf{k}_N)\right]_{\bar{\mathbf{k}}_\alpha=-\sum_{\beta\neq\alpha}\mathbf{k}_\beta}\,.
\end{aligned} \tag{2.57}$$

Using the above equation recursively the conformal invariance of the $N-$point function follows from the conformal invariance of the two-point function. The above proof can be trivially extended to $\Delta \neq 0$.

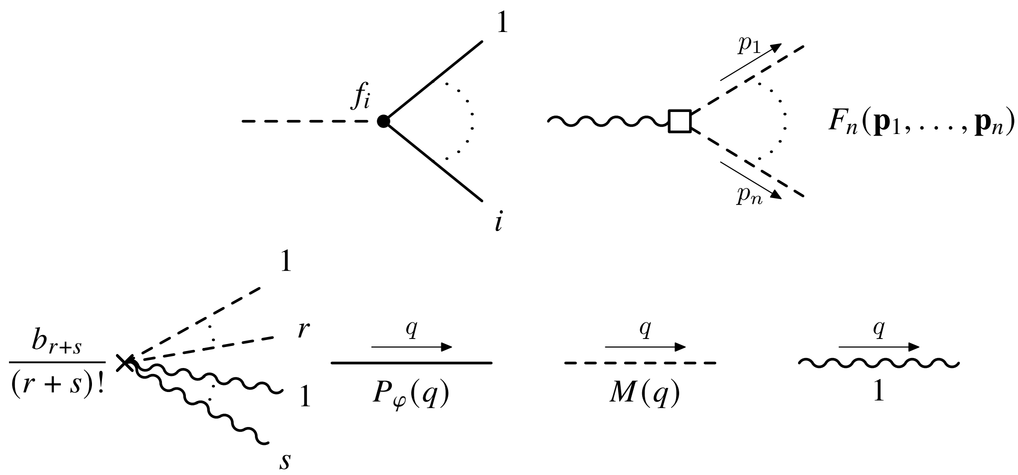


Figure 2.7: Feynman rules for the vertices and the lines used in the computation of galaxy overdensity field correlators.

Appendix 2.B Feynman Rules for the Galaxy Bias expansion

Here we describe the diagrammatic formalism we use to graphically keep track of the different contributions in the renormalization of the operator δ_ρ^2. We find the diagrammatic notation introduced here closer to the Feynman diagrams typically used in particle physics compared to what is present in the literature.

For the calculations performed in this chapter we need three kinds of vertices - the dot, the box and the cross - and three kinds of lines - the solid, the dashed and the wavy. Fig 2.7 shows the factors associated with each of the above elements. The f_i and F_n vertices describe the effect of primordial non-Gaussianities and non-linear gravitational evolution respectively. The b_i vertices represent the insertion of the operator $(\delta_\rho)^i$ in correlators of the galaxy overdensity field. Solid and dashed lines denote the power spectrum P_φ of the Gaussian field and the transfer function M introduced in Eqs. (2.2) and (2.12), respectively.

Valid diagrams representing contributions to the connected correlators of the galaxy overdensity field δ_g in wavevector space are drawn using the following rules,

- Every f_i vertex (dot) is connected to a dashed line and to i solid lines;
- Every F_n vertex (box) is connected to a wavy line and to n dashed lines. Each dashed line is connected to a solid line;
- Every b_i vertex (cross) is connected to r wavy lines and s dashed lines where $r + s = i$, $r, s \geq 0$.

Diagrams with N cross vertices represent contributions to the galaxy overdensity field correlator $\langle\tilde{\delta}_g(\mathbf{k}_1)\ldots\tilde{\delta}_g(\mathbf{k}_N)\rangle$. To translate a diagram into a formula, we label the N cross vertices (arbitrarily) with wavevectors $\mathbf{k}_1,\ldots,\mathbf{k}_N$ conventionally considered to be incoming. Then, we multiply the factors obtained using the following rules and sum over all the distinct ($N!$) labelings of the $\mathbf{k}_i$'s.

- Label each internal line with a different wavevector $\mathbf{q}_j$ and assign a factor $P_\varphi(q_j)$ to solid lines, $M(q_j)$ to dashed lines and 1 to wavy lines;
- Assign factors of f_i, $F_n(\mathbf{p}_1,\ldots,\mathbf{p}_n)$ and $b_{r+s}/(r+s)!$ to the dot, the box and the cross vertices as shown in Fig. 2.7;
- Further assign a factor of $(2\pi)^3\delta^3(\mathbf{k}_i-\mathbf{p})$ to each cross labelled with wavevector $\mathbf{k}_i$, and a factor of $(2\pi)^3\delta^3(\mathbf{p})$ to each dot and box vertex, being $\mathbf{p}$ the sum of the outgoing wavevectors at the vertex. This imposes wavevector conservation at each vertex;
- Integrate over all the internal wavevectors $\int d^3\mathbf{q}_j/(2\pi)^3$. For each loop, this procedure will leave an unconstrained wavevector $\mathbf{q}_i$ to be integrated over;
- Assign a factor that takes into account the number of possible "Wick contractions".

As anticipated in Sec.2.2 for loop diagrams regularization is needed to make the results finite. The three regularization prescriptions discussed in the main text are implemented in this diagrammatic notation as follows.

- For the first regularization scheme adopted in this chapter, we assign a $\theta(\Lambda-p)$ to each dashed line in a loop with wavevector $\mathbf{p}$ flowing into it;
- For the second regularization scheme, we assign a $\theta(\Lambda-p)$ to each line stemming from the cross vertex in the loop with wavevector $\mathbf{p}$ flowing into it;
- For the third regularization scheme (i.e. cutting off the convolution integral of δ_ρ^2) the additional Feynman rule for loops prescribes summing over two thetas as in Eq. (2.18) and dividing by a factor of two. This corresponds to the two ways of *routing* the loop wavevector $\mathbf{q}$ (e.g. Fig. 2.3).

We will now present an illustrative example to familiarize the reader with the notation.

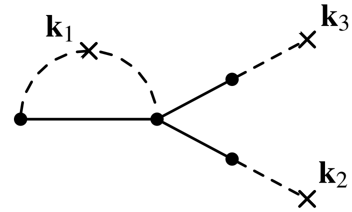


Figure 2.8: Diagram contributing to the three point function of the galaxy overdensity field.

The contribution to the bispectrum of the diagram in Fig. 2.8 in the third regularization scheme is

$$12\, b_1^2 \frac{b_2}{2} f_3 M(k_2) M(k_3) P_\varphi(k_2) P_\varphi(k_3) \int \frac{d^3q}{(2\pi)^3} M(q) M(|\mathbf{k}_1 - \mathbf{q}|) P_\varphi(q) \times \left[\frac{\theta(\Lambda - q) + \theta(\Lambda - |\mathbf{k} - \mathbf{q}|)}{2} \right]. \tag{2.58}$$

Using Eq. (2.13) the finite parts of loop integrals can easily be estimated in the region $k_i < q_0$. For example for the integral in Eq. (2.58) we have

$$\int \frac{d^3q}{(2\pi)^3} P_\varphi(q) M(q) \left[M(|\mathbf{q} - \mathbf{k}_1|) - M(q) \right] \theta(\Lambda - q) \underset{|\mathbf{k}_1| < q_0}{\sim} \frac{A}{2\pi^2} \left(\frac{2}{5\Omega_m} \right)^2 \frac{k_1^2 q_0^2}{H_0^4}. \tag{2.59}$$

In the text we addressed infrared divergences by expanding the integrands of loop diagrams and truncating the infrared divergent parts. Alternatively, infrared divergences in correlators of the galaxy overdensity field can be removed by introducing an infrared regulator m, modifying the tree-level power spectrum of the primordial curvature fluctuations as

$$P_\varphi(q) \to \frac{A}{(q^2 + m^2)^{3/2}}. \tag{2.60}$$

For example, the contribution to the galaxy overdensity bispectrum in Fig. 2.9 is

$$2 \frac{b_2}{2} b_1^2 f_2^2 M(k_2) M(k_3) P_\varphi(k_2) P_\varphi(k_3) \sigma^2(\mathbf{k}_1, \mathbf{k}_2; \Lambda, m) + \{\mathbf{k}_2 \leftrightarrow \mathbf{k}_3\} \tag{2.61}$$

where

$$\sigma^2(\mathbf{k}_1, \mathbf{k}_2; \Lambda, m) = \int \frac{d^3q}{(2\pi)^3} \frac{A M(q) M(|\mathbf{k}_1 - \mathbf{q}|)}{(|\mathbf{q} + \mathbf{k}_2|^2 + m^2)^{3/2}} \theta(\Lambda - q). \tag{2.62}$$

Most of the dependence on the infrared regulator m in the above integral is from the region of integration where the argument of M is less than q_0.

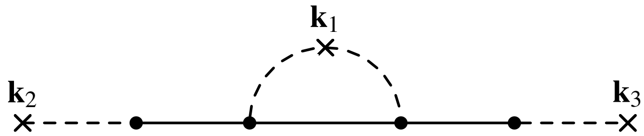


Figure 2.9: Example of an infrared divergent diagram.

We find that

$$\sigma^2(\mathbf{k}_1, \mathbf{k}_2; m; \Lambda) = \left(\frac{2}{5\Omega_m H_0}\right)^2 A \frac{k_3^2 |\mathbf{k}_1 - \mathbf{k}_2|^2}{4\pi^2} \log\left(\frac{q_0^2}{m^2}\right) + \text{IR finite as } \frac{m}{q_0} \to 0\,. \tag{2.63}$$

For a wide range of $m/H_0 \ll 1$, the correlators of the galaxy overdensity field depend very weakly on the exact value of m.

Part II

Symmetry, Naturalness, and Hidden Sectors Beyond the Standard Model

Chapter 3

FINITE NATURALNESS AND QUARK-LEPTON UNIFICATION

In this chapter, we study the implications of finite naturalness in Pati-Salam models where $SU(3)_C$ is embedded in $SU(4)$. For the minimal realization at low-scale of quark-lepton unification, which employs the inverse seesaw for neutrino masses, we find that radiative corrections to the Higgs boson mass are at least $\delta m_h^2/m_h^2 \sim O(10^4)$. The one-loop contributions to the Higgs mass are suppressed by four powers of the hypercharge gauge coupling. We find that for the vector leptoquarks the naively leading part of the two-loop corrections cancel. We assume the Dirac Yukawa couplings for neutrinos are equal to the up-type quark Yukawa couplings as predicted in the minimal theory for quark-lepton unification. Despite these findings, the two-loop corrections still dominate the finite naturalness bound. We mention a way to relax the lower bound on the vector leptoquark mass and have $\delta m_h^2/m_h^2 \sim O(10^2)$.

3.1 Introduction

The hierarchy problem (for some recent reviews see [29, 30]) has motivated many proposed extensions of the standard model (SM), including technicolor and low energy supersymmetry. This problem arises from the quadratic dependence of the Higgs mass parameter on the momentum cut-off. The absence of experimental evidence for new particles at the Large Hadron Collider (LHC) has cast doubt on the relevance of this issue. The original description of the hierarchy problem involved integrating out momentum shells [31]. Although this is a compelling physical picture, the quadratic divergences do not appear in all regulator approaches, for example in dimensional regularization.

In extensions of the SM that contain particles with masses (or fields with expectation values) much greater than the weak scale there can be a quadratic dependence of the Higgs mass on these quantities that is independent of the regularization and subtraction scheme. Demanding that these are not too large compared to the Higgs mass is the principle of finite naturalness [32]. Since the hierarchy problem is not a mathematical inconsistency, it might be a red herring associated with how we approach the theory. Nonetheless, in this chapter we take finite naturalness as a serious constraint on extensions of the SM.

An appealing extension of the SM is the SU(5) grand unified theory [33]. Here, the strongest finite naturalness constraints occur at tree level (the doublet-triplet splitting problem). There are also problematic radiative corrections to the Higgs mass from virtual super heavy gauge bosons. Since the gauge couplings are known, these radiative corrections give $\delta m_h^2/m_h^2 \sim \mathcal{O}(10^{24})$, which is clearly unacceptable if finite naturalness is to be taken seriously.

Quark-lepton unification, introduced by Pati and Salam [34, 35], is also a very attractive extension of the SM. They proposed that the standard model $\mathrm{SU}(3)_C$ is embedded in SU(4) where the leptons are interpreted as the fourth color. In this theory, baryon number is conserved and therefore the vector leptoquarks do not have to be super heavy to satisfy the strong constraints on baryon number violating processes. However, the vector leptoquarks do give rise to new flavor-violating processes that are very constrained by experimental data. For example, in the absence of large mixings, $K_L \to \mu^\pm e^\mp$ constrains the mass of the vector leptoquarks to be above 1000 TeV [36]. Using the freedom in the unknown mixings between quarks and leptons this bound can be reduced to around 100 TeV [37, 38].

In this chapter, we study the implications of finite naturalness on the minimal theory of quark-lepton unification based on the gauge group $\mathrm{SU}(4) \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_R$ [39] (see also Ref. [40]). To achieve a low-scale breaking of SU(4), we use the inverse seesaw mechanism for neutrino masses [40]. For 100 TeV vector leptoquarks, we find that $\delta m_h^2/m_h^2 \sim \mathcal{O}(10^4)$, which is problematic from the perspective of finite naturalness. In the two-loop corrections to the Higgs mass involving the vector leptoquark we show that the naively leading contributions cancel. However, we find that the remaining parts still dominate the overall radiative contribution to the Higgs mass. An analogous effect happens in the two-loop contributions of the new neutral gauge boson, in which the naively leading corrections cancel. Finally, we discuss a way to relax the bounds on the vector leptoquark mass and have $\delta m_h^2/m_h^2 \sim \mathcal{O}(10^2)$.

3.2 Minimal Theory For Quark-Lepton Unification

We consider the simplest Pati-Salam type extension of the SM, based on the gauge group [39, 40]

$$\mathcal{G}_{\mathrm{PS}} \equiv \mathrm{SU}(4) \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_R\,. \tag{3.1}$$

The quarks and leptons are unified in the following representations:

$$\begin{aligned} F_{\mathrm{QL}} &= \begin{pmatrix} u_L & \nu_L \\ d_L & e_L \end{pmatrix} \sim (4,2,0)_{\mathrm{PS}}\,, \\ F_u &= \begin{pmatrix} u_R & \nu_R \end{pmatrix} \sim (4,1,1/2)_{\mathrm{PS}}\,, \\ F_d &= \begin{pmatrix} d_R & e_R \end{pmatrix} \sim (4,1,-1/2)_{\mathrm{PS}}\,. \end{aligned} \tag{3.2}$$

Each of these comes in three copies. These representations contain the SM fermions plus (three) right-handed neutrinos. Since $\mathcal{G}_{\mathrm{PS}}$ is a product of three gauge groups, this theory has three gauge couplings, g_4, g_2, and g_R, which are determined by the SM gauge couplings g_3, g_2, and g_Y.

The gauge group $\mathcal{G}_{\mathrm{PS}}$ is connected to the SM gauge group through one breaking step, where an electrically neutral component of a non-trivial representation of $\mathrm{SU}(4)$ gets a vacuum expectation value (vev). This representation should be a singlet under $\mathrm{SU}(2)_L$, which is the same as the $\mathrm{SU}(2)_L$ in the SM. We will focus on the simplest scalar representation able to trigger such a breaking, $\chi = (\chi^\alpha, \chi^4) \sim (4,1,1/2)_{\mathrm{PS}}$, and take $\langle \chi^A \rangle = \delta^{A4} v_\chi/\sqrt{2}$. Here α is a color index ($\alpha = 1,2,3$) and A is a $\mathrm{SU}(4)$ index ($A = 1,2,3,4$). The standard model hypercharge is given by

$$Y = R + \frac{\sqrt{6}}{3} T_{15}\,, \tag{3.3}$$

where R is the $\mathrm{U}(1)_R$ charge, and T_{15} is the $\mathrm{SU}(4)$ generator

$$T_{15} = \frac{1}{2\sqrt{6}} \mathrm{diag}(1,1,1,-3)\,. \tag{3.4}$$

The quark-lepton unification angle

In this section, we summarize some aspects of minimal quark-lepton unification. Once χ gets a vev, the new massive vectors associated with the broken generators of $\mathrm{SU}(4)$ acquire mass through the covariant derivative of χ

$$D^\mu \chi \;=\; \partial^\mu \chi + i g_4 T_a V_a^\mu \chi + i \frac{g_R}{2} B_R^\mu \chi\,, \tag{3.5}$$

with T_a being the generators of $\mathrm{SU}(4)$, normalized in the standard fashion, i.e. $\mathrm{Tr}(T_a T_b) = \frac{1}{2}\delta_{ab}$. The mass of the vector leptoquarks, $X_\mu \sim (3,1,2/3)_{\mathrm{SM}}$, is given by

$$m_X^2 = \frac{1}{4} g_4^2 v_\chi^2\,. \tag{3.6}$$

The new neutral massive gauge boson, Z'_μ, is a linear combination of the $V_{15\mu}$ gauge boson associated to the broken SU(4) generator T_{15} (see Eq. (3.4)) and the $\mathrm{U}(1)_R$ gauge boson, $B_{R\mu}$. Its mass comes from the χ kinetic term,

$$\begin{aligned}\mathcal{L} &\supset (D_\mu \chi)^\dagger (D^\mu \chi) \\ &\supset \frac{v_\chi^2}{8} \begin{pmatrix} V_{15\mu} & B_{R\mu} \end{pmatrix} \begin{pmatrix} \frac{3g_4^2}{2} & -\frac{3g_R g_4}{\sqrt{6}} \\ -\frac{3g_R g_4}{\sqrt{6}} & g_R^2 \end{pmatrix} \begin{pmatrix} V_{15}^\mu \\ B_R^\mu \end{pmatrix} .\end{aligned} \tag{3.7}$$

The rotation matrix that brings the gauge bosons to their mass eigenbasis is

$$\begin{pmatrix} V_{15\mu} \\ B_{R\mu} \end{pmatrix} = \begin{pmatrix} \cos\theta_S & \sin\theta_S \\ -\sin\theta_S & \cos\theta_S \end{pmatrix} \begin{pmatrix} Z'_\mu \\ B_\mu \end{pmatrix} , \tag{3.8}$$

where the quark-lepton unification angle θ_S is given by

$$\sin\theta_S = \frac{g_R}{\sqrt{g_R^2 + \frac{3}{2} g_4^2}} , \quad \cos\theta_S = \frac{g_4}{\sqrt{\frac{2}{3} g_R^2 + g_4^2}} . \tag{3.9}$$

The hypercharge gauge coupling is $g_1 = g_R \cos\theta_S$, while the strong gauge coupling is $g_3 = g_4$. Thus, the quark-lepton unification angle can be written as a function of standard model gauge couplings and the weak angle

$$\sin^2\theta_S = \frac{2}{3} \frac{\alpha_{\rm em}}{\alpha_s} \frac{1}{\cos^2\theta_W} . \tag{3.10}$$

The mass of the gauge boson Z' is given by the trace of the mass matrix (the hypercharge gauge boson is massless in the electroweak symmetric phase),

$$m_{Z'}^2 = \frac{1}{4} \left(\frac{3}{2} g_4^2 + g_R^2 \right) v_\chi^2 . \tag{3.11}$$

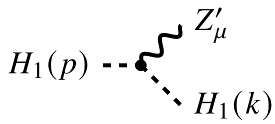
Therefore,

$$\frac{m_X^2}{m_{Z'}^2} = \frac{2}{3} \cos^2\theta_S < 1 . \tag{3.12}$$

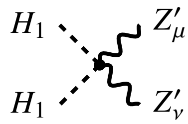
One-loop corrections to the Higgs mass

The Higgs boson is a singlet under SU(4). However, it carries $R = 1/2$ charge, $H_1 \sim (1, 2, 1/2)_{\rm PS}$, and therefore couples to the Z'_μ. From the lower bound on the vector leptoquark mass mentioned in the introduction, Eq. (3.12) implies $m_{Z'} > 130$ TeV.

The standard model Higgs couples to the Z'_μ gauge boson through the Higgs covariant derivative[1] resulting in the following Feynman rules

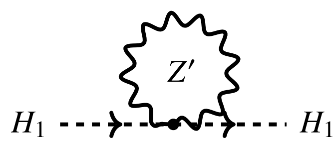


$$: \quad i\frac{g_R}{2}(p+k)_\mu \sin\theta_S\,, \tag{3.13}$$



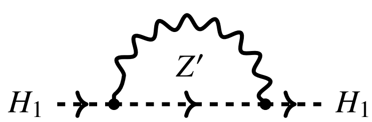
$$: \quad i2!\frac{g_R^2}{4}\sin^2\theta_S g^{\mu\nu}\,. \tag{3.14}$$

Therefore, the Z'_μ contributes to the Higgs boson mass through the following one-loop diagrams



$$\begin{aligned} i\,\delta m_h^2\big|_1 &= i\frac{g_R^2\sin^2\theta_S}{4}\,g_{\mu\nu}\int\frac{d^4k}{(2\pi)^4}\frac{-ig^{\mu\nu}}{k^2-m_{Z'}^2} \\ &= -ig_R^2\sin^2\theta_S\frac{m_{Z'}^2}{(4\pi)^2}\log\left(\frac{m_{Z'}^2}{m_h^2}\right), \end{aligned} \tag{3.15}$$

and



$$\begin{aligned} i\,\delta m_h^2\big|_2 &= i^2g_R^2\frac{\sin^2\theta_S}{4}\int\frac{d^4k}{(2\pi)^4}\frac{i(-i)k^2}{(k^2-m_{Z'}^2)k^2} \\ &= i\frac{g_R^2}{4}\sin^2\theta_S\frac{m_{Z'}^2}{(4\pi)^2}\log\left(\frac{m_{Z'}^2}{m_h^2}\right), \end{aligned} \tag{3.16}$$

where we used Feynman gauge and set the subtraction point equal to the Higgs mass. In this chapter, we neglect scheme-dependent terms that do not contain the large logarithm. Adding up both contributions, we find that the one-loop correction to the

[1]The models of Refs. [39, 40] contain two Higgs doublets that live in two different representations of the $\mathcal{G}_{\rm PS}$. However, the results in this section are not affected.

Higgs boson mass is given by

$$\left|\frac{\delta m_h^2}{m_h^2}\right| = \frac{3}{4}\frac{g_R^2 \sin^2\theta_S}{16\pi^2}\frac{m_{Z'}^2}{m_h^2}\log\left(\frac{m_{Z'}^2}{m_h^2}\right) \simeq 4\times 10^2\left(\frac{m_{Z'}}{130\text{ TeV}}\right)^2\left(\frac{\log(m_{Z'}^2/m_h^2)}{13.9}\right). \tag{3.17}$$

Note that the above result is suppressed by four powers of the hypercharge coupling since

$$g_R^2 \sin^2\theta_S = \frac{8\pi}{3}\frac{\alpha_{\text{em}}^2}{\cos^4\theta_W \alpha_s}\left(1-\frac{2}{3}\frac{\alpha_{\text{em}}}{\alpha_s\cos^2\theta_W}\right)^{-1} \simeq 6\times 10^{-3}\,. \tag{3.18}$$

Two-loop corrections to the Higgs mass

The dominant two-loop corrections to the Higgs mass contain interactions of the vector leptoquarks and the Z' gauge boson with the fermions. We will give an estimate of the size of these contributions by computing only the diagrams involving the vector leptoquarks.

Because of the vertex of the Z' gauge boson with the Higgs, there are extra two-loop diagrams contributing to the corrections involving the neutral gauge boson. We do not consider these contributions given that they are proportional to powers of the weak coupling rather than the strong coupling.

In quark-lepton unification, the Higgs Yukawa couplings predict the following mass matrix relations

$$M_d = M_e\,, \quad \text{and} \quad M_u = M_\nu^D\,. \tag{3.19}$$

The main corrections to the Higgs mass will involve the top quark and the neutrinos, and therefore we focus on the $M_u = M_\nu^D$ prediction. The other relation in Eq. (3.19) will be addressed in the next section.

To correct the $M_u = M_\nu^D$ identity, we assume the neutrinos get mass through the inverse seesaw [41, 42], adding three left-handed singlets N_L, which allows for the following interactions [40]

$$-\mathcal{L} \supset Y_1 \bar{F}_{QL}\widetilde{H}_1 F_u + Y_5 \bar{F}_u \chi N_L + \frac{1}{2}\mu N_L^T C N_L + \text{h.c.}\,, \tag{3.20}$$

where $\widetilde{H}_1 = i\sigma_2 H_1^*$. In the broken phase, the mass matrix in the basis $(\nu_L, (\nu_R)^c, N_L)$ is given by

$$\begin{pmatrix} 0 & M_\nu^D & 0 \\ (M_\nu^D)^T & 0 & M_\chi^D \\ 0 & (M_\chi^D)^T & \mu \end{pmatrix}, \tag{3.21}$$

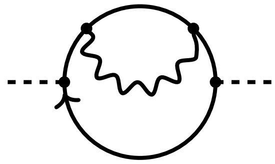
where $M_\nu^D = Y_1 v/\sqrt{2}$, $M_\chi^D = Y_5 v_\chi/\sqrt{2}$, and $v = 246$ GeV is the electroweak vev. We assume the following hierarchy $\mu \ll M_\nu^D \ll M_\chi^D$ which ensures that the light neutrinos are mostly ν_L. We work in the limit $\mu \to 0$ and $M_\nu^D \ll M_\chi^D$, where the heavy neutrino mass eigenstates are approximately Dirac and are obtained by diagonalizing M_χ^D.

In the mass eigenbasis, choosing the left-handed neutrinos to align with the left-handed up-type quarks, the relevant interactions and mass terms are

$$\begin{aligned} -\mathcal{L} \supset & -\frac{g_4}{\sqrt{2}} X_\mu \left(\bar{u}_L \gamma^\mu \nu_L + \bar{u}_R \gamma^\mu W_R \nu_R\right) \\ & + \frac{\sqrt{2}}{v} h \left(\bar{u}_L \widehat{M}_u u_R + \bar{\nu}_L \widehat{M}_u W_R \nu_R\right) \\ & + \bar{\nu}_R \widehat{M}_\chi^D N_L + \bar{u}_L \widehat{M}_u u_R + \text{h.c.}\,, \end{aligned} \tag{3.22}$$

where $\widehat{M}_u = U_R^\dagger M_u U_L = \mathrm{diag}(m_u, m_c, m_t)$, and the diagonal Dirac neutrino mass matrix $\widehat{M_\chi^D} = V_R^\dagger M_\chi^D V_L$. In the above equation, $W_R = U_R^\dagger V_R$.

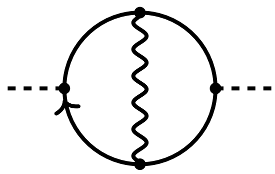
Now, we can proceed with the discussion of the two-loop contributions of the X leptoquarks to the Higgs mass. We will first take all the couplings equal to one and ignore combinatorial factors. We identify two diagram topologies involving the vector leptoquarks and the fermions that give radiative corrections to the Higgs mass. We call I_A the two-loop integral corresponding to topology A, given by



$$i\,I_{\mathrm{A}} = (-)\int \frac{d^4k}{(2\pi)^4} \int \frac{d^4q}{(2\pi)^4} \tag{3.23}$$

$$\times \frac{\mathrm{Tr}\left\{(i\not{k})\gamma_\mu(-ig^{\mu\nu})i(\not{k}-\not{q})\gamma_\nu(i\not{k})(i\not{k})P_{L,R}\right\}}{(k-q)^2(q^2-M^2)k^6}$$

while the two-loop integral associated with topology B is given by



$$i\,I_{\mathrm{B}} = (-)\int \frac{d^4k}{(2\pi)^4}\int \frac{d^4q}{(2\pi)^4} \tag{3.24}$$

$$\times\frac{\mathrm{Tr}\{(i\not{k})(i\not{k})\gamma_\mu(-ig^{\mu\nu})i(\not{k}-\not{q})i(\not{k}-\not{q})\gamma_\nu P_{L,R}\}}{(k-q)^4(q^2-M^2)k^4}\,.$$

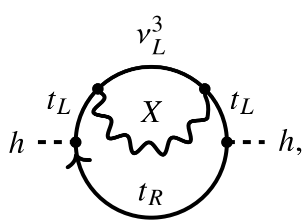
We work in the limit where the relevant mass is the gauge boson mass (M), and consider the fermions massless. Notice that we include a projector in the trace to account for the chirality characterizing the fermionic line. For the purposes of this chapter we treat γ_5 as anticommuting with all other γ matrices. We evaluate the above expressions in $d = 4 - \epsilon$ dimensions. To assess the finite naturalness of this model we need to compute the leading scheme-independent finite parts of the radiative corrections to the Higgs mass.

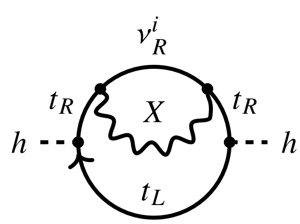
By dimensional analysis, we expect all these diagrams to be proportional to

$$M^{2-2\epsilon} \underset{\epsilon\to 0}{\simeq} M^2\left(1 - \epsilon\log M^2 + \frac{\epsilon^2}{2}\log^2 M^2\right)\,.$$

The $1/\epsilon^2$ contribution from each diagram multiplying the above generates a divergent term of the form $(1/\epsilon)\,M^2\log\!\left(M^2\right)$. Should such term survive in the sum of all the diagrams it would require non-analytic counterterms. Hence we expect no residual $1/\epsilon^2$ dependence in the overall amplitude. As a consequence, the leading finite part will depend linearly rather than quadratically on $\log M^2$. We explicitly checked that this cancellation occurs.

First, consider the diagrams





$$. \tag{3.25}$$

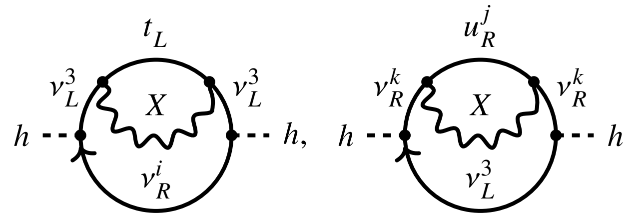
The diagram where the left-handed neutrino ν_L^3 pairs with the top quark gives

$$\delta m_h^2\big|_{X,1}^{\mathrm{A}} = 3\times 2\times g_4^2\left(\frac{2m_t^2}{v^2}\right) I_{\mathrm{A}}\,, \tag{3.26}$$

where the factor 3 comes from color, and the 2 is a combinatorial factor that arises from possible ways of contracting the fields. The diagram involving right-handed neutrinos ν_{Rj} leads to the following amplitude

$$\delta m_h^2\Big|_{X,2}^{\mathrm{A}} = 3 \times 2 \times g_4^2 \left(\frac{2m_t^2}{v^2}\right)\left(\sum_{i=1}^{3} |W_R^{3i}|^2\right) I_{\mathrm{A}} \,. \tag{3.27}$$

Similarly, the Higgs couples to the neutrinos, as the following diagrams display



. (3.28)

The diagram where the ν_L^3 couples with the vector leptoquark gives

$$\delta m_h^2\Big|_{X,3}^{\mathrm{A}} = 3 \times 2 \times g_4^2 \left(\frac{2m_t^2}{v^2}\right)\left(\sum_{i=1}^{3} |W_R^{3i}|^2\right) I_{\mathrm{A}} \,, \tag{3.29}$$

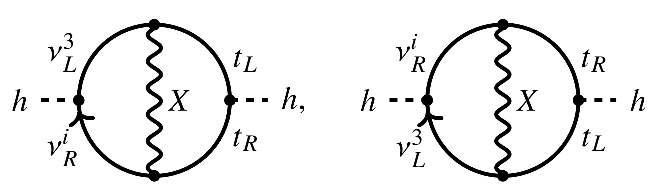
where we sum over the right-handed neutrinos ($i = 1, 2, 3$). The diagram where the right-handed neutrinos couple to the vector leptoquark also gives

$$\begin{aligned} \delta m_h^2\Big|_{X,4}^{\mathrm{A}} &= 3 \times 2 \times g_4^2 \left(\frac{2m_t^2}{v^2}\right) \qquad (3.30)\\ &\times \left(\sum_{i,j,k=1}^{3} (W_R^*)^{ij} W_R^{ik} (W_R^*)^{3k} W_R^{3j}\right) I_{\mathrm{A}} \,, \end{aligned}$$

where we sum over right-handed neutrinos and right-handed up-type quarks. Adding all contributions of topology A, *i.e.* Diag. (3.25) and (3.28),

$$\begin{aligned} \delta m_h^2\Big|_{X}^{\mathrm{A}} &= \delta m_h^2\Big|_{X,1}^{\mathrm{A}} + \delta m_h^2\Big|_{X,2}^{\mathrm{A}} + \delta m_h^2\Big|_{X,3}^{\mathrm{A}} + \delta m_h^2\Big|_{X,4}^{\mathrm{A}} \\ &= 24\, g_4^2 \left(\frac{2m_t^2}{v^2}\right) I_{\mathrm{A}} \,. \end{aligned} \tag{3.31}$$

One also expects the following diagrams,



, (3.32)

which give

$$\delta m_h^2\Big|_X^{\mathrm{B}} = 2 \times 3 \times 2 \times g_4^2 \left(\frac{2m_t^2}{v^2}\right) \left(\sum_{i=1}^{3} |W_R^{3i}|^2\right) I_{\mathrm{B}} \,, \tag{3.33}$$

where, as in the previous cases, the factor 3 comes from color, a factor 2 comes from field contractions, and a factor 2 comes from the two contributions in Diag. (3.32).

The overall contribution involving the X vector leptoquarks from both topologies is then proportional to

$$2I_{\mathrm{A}} + I_{\mathrm{B}} = \frac{1}{32\pi^4} m_X^2 \log\left(\frac{m_X^2}{m_h^2}\right) + \cdots \,. \tag{3.34}$$

As expected by general arguments and explicitly shown above, the leading scheme-independent piece scales linearly with the logarithm.

Therefore, we find

$$\begin{aligned} \delta m_h^2\Big|_{\mathrm{X}} &= \delta m_h^2\Big|_X^{\mathrm{A}} + \delta m_h^2\Big|_X^{\mathrm{B}} \\ &= i\frac{3\alpha_s}{\pi^3}\frac{m_t^2}{v^2} m_X^2 \log\left(\frac{m_X^2}{m_h^2}\right) \,. \end{aligned} \tag{3.35}$$

Hence,

$$\left|\frac{\delta m_h^2}{m_h^2}\right| = 5 \times 10^4 \left(\frac{m_X}{100\ \mathrm{TeV}}\right)^2 \left(\frac{\log\left(m_X^2/m_h^2\right)}{13.4}\right) \,. \tag{3.36}$$

Analogously to the vector leptoquark case, it can be shown that the leading contribution (proportional to g_3^2) coming from the same topologies involving the Z' scales linearly with the logarithm of the heavy neutral gauge boson mass. However, this is slightly smaller than the X-boson radiative correction. The heavier mass of the Z' boson is indeed compensated by the prefactors from the $SU(4)$ generators (see Eq. (3.4)) and the cosine of the quark-lepton unification angle.

Fermion masses

As manifest in Eq. (3.19), quark-lepton unification (with one Higgs) predicts the same mass matrix for the down-type quarks and charged leptons, which we have not addressed yet. Within the context of the $\mathcal{G}_{\mathrm{PS}}$ gauge group and demanding quark-lepton unification, there are two ways to correct the mass relation above at the renormalizable level[2]:

[2]We do not consider corrections from Planck suppressed operators as we expect the scale of quark-lepton unification to be considerably lower than M_{Pl} so that their effect should be irrelevant.

1. **Adding extra scalars**. The simplest option of this class is to add a single scalar representation in the adjoint of SU(4),

$$\Phi_{15} = (15, 2, 1/2)_{\rm PS} = \begin{pmatrix} \Phi_8 & \Phi_3 \\ \Phi_4 & 0 \end{pmatrix} + \sqrt{2} H_2 T_{15} \,, \tag{3.37}$$

which contains a colored doublet Higgs boson $\Phi_8 \sim (8, 2, 1/2)_{\rm SM}$, two scalar leptoquarks $\Phi_3 \sim (\bar{3}, 2, -1/6)_{\rm SM}$ and $\Phi_4 \sim (3, 2, 7/6)_{\rm SM}$, and a second Higgs boson $H_2 \sim (1, 2, 1/2)_{\rm SM}$. The quantum numbers of Φ_{15} allow Yukawa-type couplings with the standard model fermions

$$\begin{aligned} -\mathcal{L} \supset\, & Y_1 \bar{F}_{QL} \widetilde{H}_1 F_u + Y_2 \bar{F}_{QL} \widetilde{\Phi}_{15} F_u \\ & + Y_3 \bar{F}_{QL} H_1 F_d + Y_4 \bar{F}_{QL} \Phi_{15} F_d + \text{h.c.} \,, \end{aligned} \tag{3.38}$$

where $\widetilde{\Phi}_{15} = i\sigma_2 \Phi_{15}^*$. Because Φ_{15} is in the adjoint of SU(4), unlike the Higgs boson H_1, it distinguishes quarks and leptons and allows for a splitting in their masses, such that

$$M_u = Y_1 \frac{v_1}{\sqrt{2}} + \frac{1}{2\sqrt{3}} Y_2 \frac{v_2}{\sqrt{2}} \,, \tag{3.39}$$

$$M_\nu^D = Y_1 \frac{v_1}{\sqrt{2}} - \frac{\sqrt{3}}{2} Y_2 \frac{v_2}{\sqrt{2}} \,, \tag{3.40}$$

$$M_d = Y_3 \frac{v_1}{\sqrt{2}} + \frac{1}{2\sqrt{3}} Y_4 \frac{v_2}{\sqrt{2}} \,, \tag{3.41}$$

$$M_e = Y_3 \frac{v_1}{\sqrt{2}} - \frac{\sqrt{3}}{2} Y_4 \frac{v_2}{\sqrt{2}} \,. \tag{3.42}$$

The vevs are defined as $\langle H_1^a \rangle = \delta^{a2} v_1/\sqrt{2}$ and $\langle H_2^a \rangle = \delta^{a2} v_2/\sqrt{2}$, where $a = 1, 2$ is a SU(2) index. The scalar potential that only contains the Higgs doublets is given by

$$\begin{aligned} V = \,& m_{11}^2 H_1^\dagger H_1 + m_{22}^2 H_2^\dagger H_2 \\ & - m_{12}^2 (H_1^\dagger H_2 + H_2^\dagger H_1) + \mathcal{O}(\lambda) \,, \end{aligned} \tag{3.43}$$

where $\mathcal{O}(\lambda)$ terms arise from quartic interactions that we neglect for simplicity. The masses of the standard model Higgs, m_h, and the heavier Higgs, m_H, after applying the minimization conditions of the potential, are given by

$$m_h^2 = \mathcal{O}(\lambda v^2) \,, \quad \text{and} \quad m_H^2 = \frac{m_{12}^2 v^2}{v_1 v_2} \,, \tag{3.44}$$

where $v = \sqrt{v_1^2 + v_2^2} = 246$ GeV is the electroweak scale. Taking the determinant of the mass-squared matrix of the Higgs boson,

$$(m_{11}^2 m_{22}^2 - m_{12}^4) \simeq m_h^2 m_H^2 \,. \tag{3.45}$$

The hierarchy $m_H \gg m_h$ can be achieved when $m_{22}^2 \gg m_{12}^2 \gg m_{11}^2$, which leads to $v_2 \ll v_1$. Here we neglect the quartic terms in the scalar potential for simplicity. In the discussion below we will show that the radiative corrections to m_{22} are proportional to the mass of the heavy gauge boson.

The complete scalar potential for quark-lepton unification with χ, H_1 and Φ_{15} can be found in Refs. [40, 43–45]. It is straightforward to see that assuming small couplings for the terms enhanced by the vev of χ is enough to avoid fine-tuning problems. For example

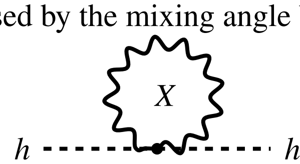
$$V \supset \lambda H_1^\dagger H_1 \chi^\dagger \chi \, , \tag{3.46}$$

does not require any cancellation if $\lambda < \alpha_s \pi m_h^2/m_X^2 \simeq 6 \times 10^{-7}(100 \text{ TeV}/m_X)^2$, which is of similar order as the electron Yukawa coupling.

Unlike in the case where there is only a Higgs doublet, singlet under SU(4), here the vector leptoquarks contribute at one-loop to the Higgs mass through the mixing between the standard model Higgs with H_2. Despite H_2 being a color singlet, it is embedded in a representation that carries SU(4) charge. The covariant derivative of an adjoint representation involves the commutator of the SU(4) vector bosons and the representation itself. The kinetic term, $\text{Tr}\{(D_\mu \Phi)^\dagger D^\mu \Phi\}$, predicts the following interaction

$$\mathcal{L} \supset \frac{2}{3} g_4^2 X_\mu^\dagger X^\mu H_2^\dagger H_2 \, . \tag{3.47}$$

Using this term in the Lagrangian, the one-loop contribution of the vector leptoquarks to the Higgs mass is suppressed by the mixing angle between the two Higgs bosons



$$\delta m_h^2 = -\frac{\alpha_s}{3\pi} \sin^2\theta \, m_X^2 \log\left(\frac{m_X^2}{m_h^2}\right) . \tag{3.48}$$

The mixing angle, θ, is defined by $h = \cos\theta \, \text{Re}(H_1^2) + \sin\theta \, \text{Re}(H_2^2)$. When $v_2 \ll v_1$, θ and thus δm_h^2 are suppressed. However, v_2 cannot be arbitrarily small since it plays the role of correcting the $M_e = M_d$ relation. For perturbative Yukawa couplings, $v_2 \gtrsim \mathcal{O}(\text{GeV})$ is required to reproduce the observed masses for the bottom quark and tau lepton. Such a hierarchy allows a H as heavy as $m_H v \sim m_{11} m_{22}(v_1/v_2)$ without requiring any tuning. It also allows to relax the one-loop correction to the Higgs mass in Eq. (3.48) to

$$\left|\frac{\delta m_h^2}{m_h^2}\right| = 20 \left(\frac{m_X}{100 \text{ TeV}}\right)^2 \left(\frac{v_2/v_1}{10^{-2}}\right)^2 \left(\frac{\log(m_X^2/m_h^2)}{13.4}\right) . \tag{3.49}$$

Notice that Eq. (3.47) implies a large radiative correction to m_{22} consistent with the hierarchy $m_{22}^2 \gg m_{12}^2 \gg m_{11}^2$ required for $v_2 \ll v_1$.

In this theory, both Y_3 and Y_4 are needed to correct the charged leptons and down-type quark masses and allow for a light (100 TeV) X vector leptoquark. In the limit where $v_2 \ll v_1$, the relation $M_u = M_\nu^D$ is preserved. We expect that the amount of tuning required in this limit is still given by the two-loop processes discussed earlier.

2. **Adding extra fermions**. Another possibility consists of adding vector-like fermions with quantum numbers such that their mixing with the standard model fermions is allowed. Thus, the mass matrix relation $M_d = M_e$ can be corrected to match the observed values of the charged leptons and down-type quarks. Notice that the above relation can be corrected, as shown for example in Ref. [46], without modifying the quark-lepton unification prediction $M_u = M_\nu^D$. Additionally, the lower bound on the vector leptoquark mass m_X, and consequently $m_{Z'}$ (see Eq. (3.12)), can be considerably relaxed, as the unitarity condition on the mixing matrices is lifted. See Ref. [46] for a recent study on this possibility. The bound on the vector leptoquark mass in this case can be as low as a few TeV, which would also relax the tuning from the two-loop contributions

$$\left|\frac{\delta m_h^2}{m_h^2}\right| = 1 \times 10^2 \left(\frac{m_X}{6\ \text{TeV}}\right)^2 \left(\frac{\log(m_X^2/m_h^2)}{7.7}\right) . \tag{3.50}$$

3.3 Concluding Remarks

The hierarchy problem, arising from the quadratic dependence of the Higgs mass parameter on the momentum cut-off, has motivated many proposed extensions of the SM. However, the absence of new particle discoveries at the LHC has raised doubts about the relevance of this issue. While quadratic divergences do not appear in all regulator approaches, certain extensions of the SM exhibit a quadratic dependence of the Higgs mass on the other masses of the theory, independent of the regularization scheme. This motivates the principle of finite naturalness, which demands that these quantities are not too much larger than the Higgs mass.

In this chapter, we explored whether Pati-Salam models with $\mathrm{SU}(3)_C$ embedded in $\mathrm{SU}(4)$ respect the principle of finite naturalness. We showed that the minimal realization of quark-lepton unification leads to radiative contributions to the Higgs boson mass squared $\delta m_h^2/m_h^2 \gtrsim \mathcal{O}(10^4)$, which is problematic regarding finite naturalness. We computed the two-loop corrections to the Higgs mass involving the vector leptoquark and showed that the naively leading contributions from these corrections cancel in the minimal theory for quark-lepton unification. Something analogous happens for the expected dominant contributions from the two-loop corrections involving the new neutral gauge boson, which cancel out. We also briefly mentioned extensions of the minimal model containing extra fermions where the level of tuning can be relaxed to $\delta m_h^2/m_h^2 \sim \mathcal{O}(10^2)$ by lowering the bound on the heavy gauge boson masses.

Chapter 4

LONG-LIVED AXIONLIKE PARTICLES FROM ELECTROMAGNETIC CASCADES

In this chapter, we study axion-like particles (ALPs) in beam dump experiments, focusing on the Search for Hidden Particles (SHiP, at CERN) experiment and the Beam Dump eXperiment (BDX, at JLab). Many existing projections for sensitivity to ALPs in beam dump experiments have focused on production from either the primary proton/electron beam, or - in the case of SHiP - the secondary (high-energy) photons produced by neutral meson decays (e.g., $\pi^0 \to \gamma\gamma$). In this chapter, we study the subsequent production of axions from the full electromagnetic shower in the target, finding order-of-magnitude enhancements in the visible decay yields across a wide range of axion masses. We update SHiP's sensitivity curve and provide new projections for BDX. Both experiments will be able to reach currently unexplored regions of ALP parameter space.

4.1 Introduction

One of the simplest mechanisms that generate light, feebly interacting particles involves the spontaneous breaking of global symmetries at a high scale. The pseudo-Nambu-Goldstone bosons that arise are naturally light, with their masses protected by a shift symmetry. Being derivatively coupled, their interactions are suppressed by a large decay constant $\sim |\boldsymbol{p}_a|/f_a$. These types of fields are often referred to as "axion-like particles" (or ALPs), and occur ubiquitously in many extensions of the Standard Model. In what follows, we will use "axion" and "ALP" interchangeably but do not restrict ourselves to models that solve the strong-CP problem [47–52].

Feebly interacting particles, such as ALPs, can be discovered using beam-dump accelerator facilities [53–55]. The search strategy involves a high-energy (typically electron or proton) beam incident on a thick target. A detector is placed sufficiently far from the target so that beam-related backgrounds are minimized, and one searches for the scattering or decay of ALPs, or other particles, produced in the beam dump.

The purpose of this chapter is to incorporate often neglected production mechanisms, in particular electromagnetic cascades, that enhance (see Fig. 4.1) the sensitivity to visibly decaying ALPs. While these effects have been previously considered[1] in the context of neutrino experiments (also proton beam dumps) [57–59] and the SLAC-E137 experiment [60, 61], they are often neglected in the analysis of other beam dumps. Our goal is to clarify when and why enhancements from electromagnetic cascades are substantial, and to quantify their impact using two representative examples: SHiP (a proton beam dump) and BDX (an electron beam dump).

A common benchmark for comparison in the ALP literature (see e.g., [53]) is an ALP coupled to photons,

$$-\mathcal{L} \supset \frac{1}{2} m_a^2 a^2 - \frac{\alpha}{4\pi} \frac{c_{a\gamma\gamma}}{f} a F \tilde{F} \; - \frac{c_{aee}}{f} (\partial_\mu a) \bar{e} \gamma^\mu \gamma_5 e \,, \tag{4.1}$$

where a is the ALP field, m_a its mass, $c_{a\gamma\gamma}$ (c_{aee}) its coupling to photons (electrons), F is the electromagnetic field strength, and $\tilde{F}$ its dual. We have included an electron coupling, c_{aee}, which is naturally generated by loops involving photons or may appear at tree-level in an ultraviolet (UV) completion. Sometimes the coupling $g_{a\gamma\gamma}$ is used in the literature; following the conventions of e.g., [65], this is related to our notation via $g_{a\gamma\gamma} = c_{a\gamma\gamma}(\alpha/\pi) f^{-1}$. As is well appreciated in the literature, high-intensity beam dump experiments can discover ALPs for masses $m_a \lesssim 2$ GeV.

The Search for Hidden Particles (SHiP) experiment is an ambitious next-generation beam dump facility [63, 64, 66]. Existing studies of SHiP [65] have found that the dominant production mechanism is Primakoff scattering on nuclei. However, only virtual photons from the primary proton beam, and real photons from first-generation meson decays, have been included in flux estimates for SHiP. As we discuss in detail below, each of these real photons produces $10^2 - 10^3$ further secondary photons, electrons, and positrons while showering in the beam stop. Furthermore, the cascade particles produce lower-energy, and therefore shorter-lived, ALPs, increasing predicted yields at small masses and couplings. We find that the inclusion of electromagnetic cascades can enhance coupling sensitivity by more than an order of magnitude (corresponding to $\sim 10^4$ in the event rate).

This enhancement is a consequence of the infrared (IR) sensitivity of the signal event rate. At small couplings, where the decay length is long, the detected event rate scales as

$$R_{\text{det}} \propto \int \mathrm{d}\omega_a \, \frac{\mathrm{d}R_a}{\mathrm{d}\omega_a} \frac{m_a}{\omega_a} \,, \tag{4.2}$$

[1] Past literature often uses a one-dimensional "forward-only approximation" that overestimates geometric acceptances [56].

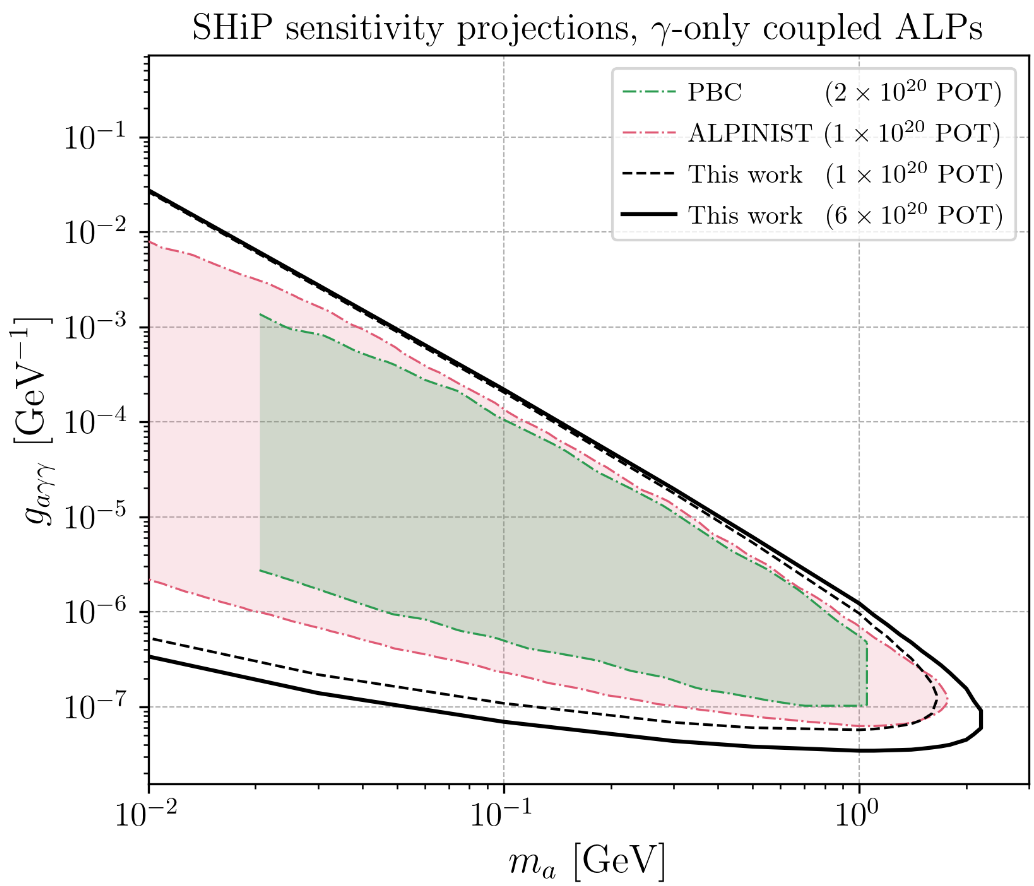


Figure 4.1: Comparison to past literature – PBC [53] (green shaded region), ALPINIST [62] (red shaded region) – of our projections for the sensitivity of SHiP (solid black line) to axion-photon couplings, with updated geometry and nominal protons-on-target (POT) [63, 64]. For reference, we also plot our sensitivity projections with a lower POT, used before in the literature (dashed black line). We defined $g_{a\gamma\gamma} \equiv f^{-1}c_{a\gamma\gamma}\alpha/\pi$, and we set $c_{aee} = 0$. Accounting for the full electromagnetic cascade substantially enhances the reach of SHiP at lower couplings and lighter masses. See Fig. 4.11 for an illustration of how this result depends on the energy threshold used at SHiP.

where m_a/ω_a accounts for the time dilation in the axion's decay length. The rate of axion production $\mathrm{d}R_a/\mathrm{d}\omega_a$ inherits the IR-dominated spectrum of bremsstrahlung photons, being cut off at some $\omega_{\min} \sim m_a$. The event rate in the long-lifetime limit is therefore dominated by the lowest energy flux of axions from the beam stop; this will be produced by the electromagnetic cascade.

The Beam-Dump eXperiment (BDX) at Jefferson Lab [67] is an electron beam dump experiment that will search for the production of light dark matter, and its elastic and inelastic scattering in a downstream detector, reaching the limit of the irreducible neutrino background. This setup will also be sensitive to a multitude of visibly-decaying dark sector states. The incident electron beam will initiate an electromagnetic cascade in the thick target in which numerous secondary particles can produce ALPs. To our knowledge, a comprehensive study of the BDX setup for axion-like particles has not been undertaken (although see [68] for an analysis with some of the relevant production mechanisms).

In what follows we briefly review the effective theory of an ALP coupled to photons and electrons in Section 4.2. We define our conventions and discuss renormalization group (RG) running below the weak scale, following [69]. In Section 4.3 we discuss the production mechanisms inside a thick target. We summarize our numerical methods in Section 4.4. In Section 4.5 we turn to the upcoming SHiP (400 GeV proton beam dump) experiment – which produces a large photon flux, primarily via π^0 production, followed by $\pi^0 \to \gamma\gamma$ – and the BDX experiment at Jefferson Lab (10.6 GeV electron beam dump). We study the extent to which electromagnetic cascades enhance sensitivity to long-lived ALPs. We find a substantial (orders of magnitude) increase in the number of axions that are produced and decay within the experiments' decay pipe. Finally, we summarize our findings and conclude with an outlook towards interesting potential future directions in Section 4.6.

4.2 Axion couplings at low scales

We define two benchmark scenarios to illustrate phenomenology dominated by photons, and electrons, respectively. We fix the photon and electron couplings at $\mu_w = 80$ GeV (just below the W-boson mass). This scale is low enough to talk sensibly about photons without considering γZ mixing, but high enough to ensure that realistic loop-induced couplings to electrons and photons are generated by renormalization group (RG) flow. We relate the couplings at μ_w to the coupling at $\mu_a = m_a$ using the RG flow described in [69]. Concretely, we take:

1. *γ-dominated:* $c_{a\gamma\gamma}(\mu_w) = 1$, $c_{aee}(\mu_w) = 0$.
2. *e-dominated:* $c_{a\gamma\gamma}(\mu_w) = 0$, $c_{aee}(\mu_w) = 1$.

Although $c_{a\gamma\gamma}$ does not flow under RG, in phenomenological applications the replacement of $c_{a\gamma\gamma} \to \tilde{c}_{a\gamma\gamma}(\mu)$ captures effects generated by loops of light particles (such as the electron). The tilded coupling is defined [69] as:

$$\tilde{c}_{a\gamma\gamma}(\mu) = c_{a\gamma\gamma} + \sum_f N_c Q_f^2 c_{aff}(\mu)\Theta(\mu - m_f) \tag{4.3}$$

with f running over all fermions in the Standard Model, Q_f the fermion electric charge, N_c its number of colors, and m_f its mass. The coupling $\tilde{c}_{a\gamma\gamma}$ naturally appears in the RG equations.

In both benchmarks, at the scale of the axion mass both couplings are present, i.e., $c_{aee}(\mu_a) \neq 0$ and $\tilde{c}_{a\gamma\gamma}(\mu_a) \neq 0$. This avoids pathologies where e.g., the computed axion bremsstrahlung rate vanishes identically.

The relevant solutions of the RG flow are given by [69],

$$\begin{aligned} c_{aee}(\mu_L) &= c_{aee}(\mu_H) - \frac{3\tilde{c}_{a\gamma\gamma}(\mu_H)}{\beta_0^{\text{QED}}} \frac{\alpha(\mu_L) - \alpha(\mu_H)}{\pi} \\ &\simeq c_{aee}(\mu_H) + \frac{3\alpha(\mu_H)}{4\pi} \log\left(\frac{\mu_H^2}{\mu_L^2}\right) \tilde{c}_{a\gamma\gamma}(\mu_H) \,, \end{aligned} \tag{4.4}$$

where μ_H is the "high scale" (where initial conditions are set) and μ_L the "low scale". In the second line, we have expanded to $O(\alpha_H^2)$. For our numerical inputs, the RG flow of $\alpha(\mu)$ is solved with the one-loop QED beta function, with the relevant fermions integrated out at $\mu = m_f$. In practice, as can be readily seen from Eq. (4.4), the end result is not sensitive to these details since the β_0^{QED} does not appear at $O(\alpha^2)$.

Next, consider the effective photon coupling $\tilde{c}_{a\gamma\gamma}$. We approximate:

$$\tilde{c}_{a\gamma\gamma} \approx c_{a\gamma\gamma} + c_{aee} \,, \tag{4.5}$$

where we neglect the running of $c_{aee} \simeq c_{aee}(\mu_w)$. This approximation is reliable for our purpose, because the two benchmarks we consider are dominated by either $c_{a\gamma\gamma} \gg c_{aff}$ or by $c_{aee} \gg c_{aff}$ for $f \neq e$. We have dropped the Heaviside function $\Theta(\mu - m_e)$ since we will always take $\mu = m_a$ and $m_a > m_e$. For the electron-dominated benchmark, our analysis matches with the conventions of [70]. For the photon-dominated benchmark, we have solved for the RG flow of c_{aee}. At a renormalization scale of $\mu_c = m_c$, we find $c_{aee}(\mu_c) = 3.6 \times 10^{-5} c_{a\gamma\gamma}$.

Before proceeding, let us comment on how small hadronic couplings must be for electromagnetic production to dominate. This can be inferred from Fig. 9 of [71]. In that work, an $O(1)$ coupling to gluons ($\frac{\alpha_s}{4\pi} c_{agg} G\tilde{G}$) was assumed. This induces a photon coupling of order $c_{a\gamma\gamma} \sim c_{agg}$, and hadronic production (from[2] $\eta \to \pi\pi a$) was found to dominate in the first interaction length of the beam dump by $\sim 10^6$.

[2] In [71], only hadronic production that properly implements chiral symmetry [72] was included. Other production modes, such as the production of axions from the primary proton beam, were not included.

Any scenario that involves couplings to the electroweak or leptonic sector will only "feed down" into quarks at one-loop, producing couplings of $c_{aqq} \sim (\alpha/4\pi)^2 \beta_0^{\mathrm{QED}} c_{a\gamma\gamma}$ and $c_{agg} \sim (\alpha/4\pi)^2 \beta_0^{\mathrm{QED}} c_{a\gamma\gamma}$. Since production rates scale as the square of the coupling, this naturally provides a 10^{-12} suppression for which electromagnetic modes easily dominate over hadronic production.

Furthermore, in what follows, we find large enhancements to the cascade production (ranging from factors of $10^2 - 10^5$ depending on the axion mass). This is not sufficient to overcome hadronic production when $c_{agg} \sim c_{a\gamma\gamma}$; however, it can be enough for scenarios where, e.g., $c_{agg} \sim 10^{-1} \times c_{a\gamma\gamma}$. This may occur if the ratio of electromagnetic and QCD anomaly coefficients of the UV model, E/N, is larger than the typical benchmark of $8/3 \approx 2.6$ [73, 74], or via a modest kinetic mixing with some hidden sector [75] among many other possible mechanisms [76–78]. A detailed study of the interplay between the shower-enhanced electromagnetic production versus hadronic production modes in specific UV completions warrants further study, but lies beyond the scope of this chapter.

4.3 Axion production mechanisms

In an electron or proton beam dump, ALPs can originate via multiple production mechanisms. As mentioned above, depending on the UV completion under consideration, either hadronic or electromagnetic production can dominate the phenomenology. Since the electromagnetic cascade is our focus, we do not discuss hadronic production modes further.

Previous projections for the SHiP experiment have neglected electromagnetic shower-initiated production in the target, focusing only on the first interaction length. In recent work [56, 79], two of us have developed a framework for the systematic inclusion of shower-induced production of light long-lived particles, implemented in the public code `PETITE`. A key advantage of this approach is its computational efficiency, thanks to the extensive usage of pre-computed `VEGAS` integrators. For this chapter, we developed `ALPETITE` [80], a dedicated tool that extends the `PETITE` framework to provide a comprehensive calculation of axion flux from electromagnetic cascades. Our approach is twofold. First, for production channels that have a direct analogue between dark photon and axion models (e.g., Bremsstrahlung), `ALPETITE` re-weights the simulated dark photon events generated by `PETITE` to derive the corresponding axion flux. Second, for processes unique to axion physics that lack a vector-particle equivalent (such as Primakoff production), we implemented and

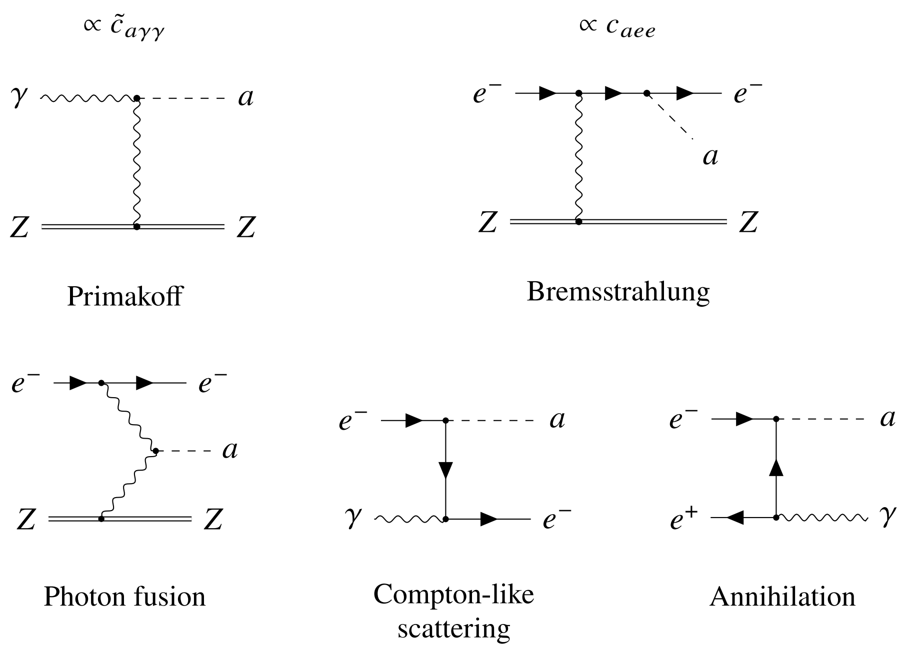


Figure 4.2: Electromagnetic production mechanisms for axion-like particles in a beam dump. Primakoff production and photon fusion proceed through the effective photon coupling $\tilde{c}_{a\gamma\gamma}$, while bremsstrahlung, Compton-like scattering, and resonant annihilation are controlled by the electron coupling c_{aee}.

integrated dedicated Monte Carlo samplers. These samplers are not run in isolation; instead, they take the SM particle shower information generated by `PETITE` as their direct input. This hybrid strategy allows `ALPETITE` to combine the speed of re-weighting with the completeness of a full simulation, ensuring an accurate and efficient result.

We now describe the relevant electromagnetic processes that dominate the ALP production rate in an electromagnetic cascade.

Primakoff upscattering: Photons can scatter on nuclei to produce ALPs of mass m_a via the $c_{a\gamma\gamma}$ coupling,

$$\gamma Z \to a Z \,, \tag{4.6}$$

where Z denotes a nucleus with charge Z. Treating the nucleus as a static charge density $\rho(\mathbf{x})$ that sources a Coulomb field, the differential cross-section is

$$\begin{aligned} \frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} =& \left(\frac{1}{f}\right)^2 \frac{Z^2\alpha^3 c_{a\gamma\gamma}^2}{32\pi^2 Q^4 \omega^2} \times \\ & \left(2Q^2(2\omega^2 - m_a^2) - m_a^4 - Q^4\right) |F_A(Q^2)|^2 \,, \end{aligned} \tag{4.7}$$

where

$$F_A(Q^2) = \int \mathrm{d}^3x\, \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}} \rho(\mathbf{x}) \,, \tag{4.8}$$

and ω is the energy of the incoming photon. Because the energy transfer to the nucleus, $\Delta\omega \simeq \mathbf{q}^2/2M_A$, is negligible, the three-momentum and four-momentum transfer are nearly identical

$$Q^2 = -q^2 \simeq \mathbf{q}^2 \,. \tag{4.9}$$

Axions may also be produced via incoherent upscattering. This corresponds to inelastic final states for the nuclear system (e.g., nucleon knockout). For consistency with the treatment of dark vector production, we incorporate incoherent upscattering using a replacement rule for the form factor,

$$|F_A(Q^2)|^2 \longrightarrow G_{2,\mathrm{el}}(Q^2) + G_{2,\mathrm{inel}}(Q^2)/Z \,, \tag{4.10}$$

where the two contributions are given by [81]

$$G_{2,\mathrm{el}}(Q^2) = \left(\frac{a^2Q^2}{1+a^2Q^2}\right)^2 \left(\frac{1}{1+Q^2/d^2}\right)^2 , \tag{4.11}$$

$$G_{2,\mathrm{inel}}(Q^2) = \left(\frac{a'^2Q^2}{1+a'^2Q^2}\right)^2 \left(\frac{1+\frac{Q^2}{4m_p^2}(\mu_p^2-1)}{(1+Q^2/\Lambda_H^2)^4}\right) . \tag{4.12}$$

The atomic parameters are given by

$$a = \frac{111}{Z^{1/3}m_e} \quad \text{and} \quad a' = \frac{773}{Z^{2/3}m_e} \,, \tag{4.13}$$

while the hadronic parameters are

$$\Lambda_H = 0.84 \text{ GeV} \quad \text{and} \quad d = 0.405\, A^{-1/3} \text{ GeV} \,. \tag{4.14}$$

The proton's magnetic moment is $\mu_p = 2.79$, and m_p is the mass of the proton. It should be noted that the second bracketed term in Eq. (4.12) is *not* squared, see Footnote 4 of [82].

In past studies of SHiP, only the coherent Primakoff channel has been included. It is commonplace in dark photon searches to also include the inelastic production discussed above (see e.g., [81], or Appendix B of either [82] or [56]). This production channel can become important at larger axion masses and we therefore include it in our analysis. In App. 4.A, we show the (modest) gain in downstream flux from the inelastic contribution for SHiP with and without electromagnetic showers.

In addition to production from real photon beams, ALPs can also be produced when an electron beam impinges on a target via photon fusion,

$$e^- Z \xrightarrow{\gamma\gamma} a\, e^- Z \tag{4.15}$$

This process is calculated using the Equivalent Photon Approximation (EPA), also known as the Weizsäcker-Williams method [83, 84]. In this framework, the total cross-section is obtained by convolving the on-shell Primakoff cross-section with the equivalent photon spectrum, expressed as a function of the photon energy fraction $x = \omega/E_e$:

$$\sigma(eZ \to eZa) = \int_{x_{\min}}^{1} \mathrm{d}x\, \frac{\mathrm{d}N_\gamma}{\mathrm{d}x} \sigma_{\gamma Z \to aZ}(xE_e)\,, \tag{4.16}$$

where E_e is the incident electron energy and $x_{\min}$ is an arbitrary IR cutoff. The "doubly-logarithmic" version of the EPA flux is given by [85]

$$\frac{\mathrm{d}N_\gamma}{\mathrm{d}x} = \frac{\alpha}{\pi}\left[\frac{1+(1-x)^2}{x}\right] \log \frac{Q^2_{\max}}{Q^2_{\min}}\,. \tag{4.17}$$

Here, $Q^2_{\min} = m_e^2 x^2/(1-x)$ is the kinematically allowed minimum photon virtuality, and $Q^2_{\max}$ is a cutoff scale related to the target's structure. The photon flux is peaked at low x, providing a direct mechanism for ALP production. This channel is particularly important for fixed-target experiments utilizing primary electron beams, e.g. BDX, providing a direct mechanism for ALP production that competes and – as we will see in the following – exceeds the primary electron bremsstrahlung.

Axion bremsstrahlung: Given a non-zero coupling to electrons, c_{aee}, one may also consider axion bremsstrahlung,

$$e^- Z \to a\, e^- Z\,. \tag{4.18}$$

This process has a matrix element structure that is similar to dark photon, A'_μ, bremsstrahlung. It is instructive to study the ratio between the square matrix elements for the two processes. In the static limit (i.e., the nucleus is taken as infinitely heavy), both processes are allowed for $E_e > m$, where m is either the vector or axion mass. In the ultra-relativistic limit ($E_e \gg m_e$), we find that

$$\frac{\langle|\mathcal{M}_a|^2\rangle}{\langle|\mathcal{M}_V|^2\rangle} = \frac{1}{2}\left(\frac{\epsilon_a}{\epsilon_V}\right)^2 + O\left(\frac{Q^2}{m^2}\right) \tag{4.19}$$

where $\epsilon_a \equiv 2m_e c_{aee}/ef$, and the interaction lagrangian for a kinetically mixed dark photon is given by

$$\mathcal{L}_V = \epsilon_V e A'_\mu \bar{\psi}\gamma^\mu\psi\,. \tag{4.20}$$

where ψ is the electron field.

Just like for vector bremsstrahlung [60, 81], when $m_a \gg m_e$ there is a kinematic preference for producing energetic, forward-peaked bosons, particularly when the electron energy is much larger than the axion mass. The resulting differential cross-section is difficult to sample efficiently. To address this, we use the `PETITE` package [56] which employs a histogram approximation to the differential cross-section, that is used for fast importance sampling of the exact one. The approximation is provided by the `VEGAS` integrator trained using the dark photon matrix element [86]. We use `PETITE` to sample the V phase space and then reweigh by the full ratio of matrix elements squared (see App. 4.B for details); this yields Monte Carlo (MC) samples that are distributed according to the correct ALP's phase space.

For electron beams, bremsstrahlung and photon fusion compete as the primary production mechanism. Technically, the amplitudes for the two mechanisms should be summed coherently since the initial and final states are the same. However, as we have shown above, the two processes have little phase space overlap: photon fusion is enhanced at small axion energies due to the soft divergence of EPA flux, while bremsstrahlung is peaked for high-energy forward ALPs. Furthermore, as we will see below, subsequent shower-induced reactions entirely dominate the sensitivity projections. Therefore, we sum the two processes independently when computing the primary production yields.

Resonant annihilation: Positrons produced in the electromagnetic shower can annihilate resonantly on atomic electrons,

$$e^+e^- \to a(\gamma)\ . \tag{4.21}$$

Note that we have allowed for the emission of a photon to bring the incident electron onto the resonant peak, i.e., radiative return.

In `PETITE`, the analogous resonant annihilation process $e^+e^- \to V(\gamma)$ involving a dark vector in the final state is implemented using a framework that combines a QED parton distribution function for initial state radiation with a model for atomic binding effects. The treatment of atomic binding begins with the tree-level cross-section for a positron of momentum $\boldsymbol{k}$ annihilating an electron in a fixed atomic orbital, which is given by

$$\sigma^{(0)}(k) = \frac{1}{4m_e\sqrt{\omega^2 - m_e^2}} \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{2E_e} \frac{1}{2E_V} \times (2\pi)\delta(\omega + m_e - \varepsilon_A - E_V)|\psi_A(p)|^2 \langle|\mathsf{M}_V|^2\rangle \tag{4.22}$$

where $|\psi_A(p)|^2$ is the bound-state electron wave-function, ε_A is the binding energy, E_V is the energy of the final-state vector and $\langle|\mathsf{M}_V|^2\rangle$ is the matrix element calculated for free-electron states. For many-electron atoms, PETITE approximates the total momentum distribution by modeling it as a sum of $1s$ hydrogenic orbitals. This choice enables an analytic calculation of the cross-section, which is then convolved with the leading-order $e^- \to e^-\gamma$ splitting function to account for initial state radiation. This method provides a computationally efficient framework for MC sampling while capturing the essential physics of resonance broadening due to the momentum of bound atomic electrons. The dark vector sample is then simply converted into an axion by reweighing it by the free-electron amplitude for the production of a pseudoscalar of the same mass (see App. 4.B for details).

The cross-section is highly peaked near the resonant energy

$$E_{\text{res}} \simeq \frac{m_a^2}{2m_e} , \tag{4.23}$$

and can dominate the axion flux. This means that resonant annihilation typically leads to flux that is dominated by $E \approx E_{\text{res}}$. Depending on the mass of the axion, this flux can have modest or very large boosts. We direct the reader to [79, 87] and references therein for a detailed discussion of the formalism for resonant production (see also [82, 88–91]).

Compton-like scattering: High energy photons can scatter off atomic electrons to produce an axion and an electron in the final state,

$$\gamma e^- \to a\, e^- . \tag{4.24}$$

When considering a free electron at rest, this process requires a minimum axion energy:

$$E_{\text{min}} \simeq \frac{m_a^2}{2m_e} . \tag{4.25}$$

In PETITE, dark Compton scattering ($\gamma e^- \to e^- V$) is treated analogously to the resonant annihilation channel. The production cross-section is computed by convolving the QED splitting function $P_{\gamma\to e}(x)$ with the cross-section for resonant annihilation of the intermediate positron on a bound atomic electron, utilizing the same atomic framework as the direct annihilation channel described above [79]. The integration over the momentum fraction, x, of the intermediate positron averages over the narrow resonance condition. This results in a broad energy spectrum for the final-state vector, in contrast to the sharply peaked spectrum from direct e^+e^- annihilation, with

$$\sigma(k) = \langle |\mathsf{M}_V|^2 \rangle \times \frac{2}{3m_e\Lambda}\frac{1}{k^2} \times \frac{\beta}{4}\mathcal{J}(a,b), \tag{4.26}$$

where $\boldsymbol{k}$ is the momentum of the incident photon, $\Lambda = Z_{\text{eff}}\,\alpha\, m_e$,

$$\beta = \frac{2\alpha}{\pi}\left(\log\left(\frac{s}{m_e^2}\right) - 1\right),$$

and $\mathcal{J}(a,b)$ is defined in Appendix A of [79].

The sampled dark vector events are then re-weighted to model axion production by applying the ratio of the squared matrix elements, as described in App. 4.B. As the Compton-like channel provides a subdominant contribution relative to resonant annihilation across the parameter space of interest, we sum their respective yields for brevity's sake.

Our list of production mechanisms is not exhaustive, but we believe that it captures all dominant production mechanisms over the parameter space we consider. For example, one can check that electromagnetically mediated meson decays such as $\pi^0 \to \gamma\gamma a$ or $\eta \to \gamma\gamma a$ will not compete with Primakoff production. Similarly, virtual Primakoff production (from a nearly-on-shell γ^* emitted by a fast proton) is subdominant relative to on-shell Primakoff production [65] (as we have explicitly verified).

4.4 Monte Carlo Methods

The expected number of detectable ALP events in a beam dump experiment depends on several sequential processes: ALP production in the target, survival during transport, decay within the observable decay region, geometric acceptance, and energy thresholds. Conceptually, if these processes were independent with uniform probabilities, the event rate could be written as:

$$N_{\text{events}} \sim N_{\text{prim}} \times P_{\text{prod}} \times P_{\text{decay}} \times P_{\text{geo}} \times P_{\text{Ecut}}\,, \tag{4.27}$$

where N_{prim} is the number of particles-on-target (POT), P_{prod} is the probability of producing an ALP, P_{decay} is the decay probability within the decay pipe and/or detector volume, P_{geo} is the geometric acceptance, and P_{Ecut} represents energy threshold requirements. However, this simplified picture breaks down in realistic simulations where each ALP candidate has different kinematics, production mechanisms, and acceptance criteria. Given the extremely small probabilities involved and the need to account for event-by-event variations, we instead calculate the event rate using a weighted MC approach. The simulation begins by generating a large sample of SM showers in the target using `PETITE`, initiated by the incident particle beam. From this shower, we generate a sample of ALP candidates, applying a series of weights corresponding to each physical process.

Production Probability This is the number of ALPs produced per primary particle. We treat every particle in the shower as a potential source and compute weighted events. For each shower particle, we generate ALP candidates using the appropriate differential cross-sections (see Section 4.3). Each candidate i is then assigned a production weight, $w_{\text{prod},i}$, equal to the total probability of its production:

$$w_{\text{prod},i} \propto \frac{\sigma_a}{\sigma_{\text{SM}}} w_{\text{prim},i} \,, \tag{4.28}$$

where σ_a is the (total) cross-section for generating an axion, σ_{SM} is the cross-section of any other SM process, and $w_{\text{prim},i}$ is the MC weight assigned to the parent/primary particle. For the specific details on how the weight is assigned for each of the production mechanisms described above, see Appendix 4.B.

Decay Probability This is the probability that an ALP decays within the decay volume of length L, situated at a distance d from the target. This is applied as a weight to each candidate i:

$$w_{\text{decay},i} = \mathrm{e}^{-d/\lambda_i}\left(1 - \mathrm{e}^{-L/\lambda_i}\right), \tag{4.29}$$

where $\lambda_i = (|\boldsymbol{p}_i|/m_a)c\tau$ is the lab-frame decay length of the candidate.

In our study, we first consider the di-photon decay mediated by the (effective) axion-photon coupling $\tilde{c}_{a\gamma\gamma}$, whose partial width is given by

$$\Gamma_{a\to\gamma\gamma} = \frac{\tilde{c}_{a\gamma\gamma}^2 \alpha^2}{64\pi^3} \frac{m_a^3}{f^2} \,, \tag{4.30}$$

where $\tilde{c}_{a\gamma\gamma}$ is given in Eq. (4.5). When electron couplings are present at tree-level, the partial width for $a \to e^+e^-$ can instead dominate,

$$\Gamma_{a\to e^+e^-} = \frac{c_{aee}^2}{2\pi} \frac{m_e^2 m_a}{f^2} \sqrt{1 - \frac{4m_e^2}{m_a^2}}. \tag{4.31}$$

We define the total lifetime τ in the particle's rest frame as:

$$\tau \equiv \hbar \left(\Gamma_{a\to\gamma\gamma} + \Gamma_{a\to e^+e^-}\right)^{-1} . \tag{4.32}$$

In the long-lived particle (LLP) limit ($\lambda_i \gg L, d$), the decay weight simplifies to $w_{\text{decay},i} \approx L/\lambda_i \propto 1/f^2$.

Geometric Acceptance This is the probability that the ALP's decay products intersect the detector. For the sake of simplicity, we consider only the ALP's trajectory. The geometry is characterized by the length of the decay volume, L, its distance from the target, d, and the cross-sectional area of the detector A. Due to relativistic kinematics, the ALP flux is typically collimated in the forward direction. Let θ_i be the angle with respect to the beam direction at which the i − th axion is produced. Let $\theta_{\text{acc}} \approx \sqrt{A}/(2L + 2d)$ be the angular acceptance of the detector. In our MC, the geometric weight is given by a binary filter applied to each candidate i:

$$w_{\text{geo},i} = \begin{cases} 1 & \theta_i \le \theta_{\text{acc}} \, , \\ 0 & \text{otherwise} \, . \end{cases} \tag{4.33}$$

A more sophisticated treatment of the geometric acceptance of the decay products can be performed (see Appendix B of [79] for a detailed analysis of SHiP's geometric efficiency). However, this was shown to be relevant for low-energy downstream fluxes, which – in our region of interest – are mostly reduced by the energy acceptance threshold described below.

Energy Acceptance A decay is only considered a signal event if it can be successfully reconstructed. In practice, the experimental requirement is a minimum visible electromagnetic energy (e.g. from decay photons and electron-positron pairs) in the detector; we implement this as a cut on the parent ALP energy E_{cut}. This acceptance criterion is applied as a step-function weight:

$$w_{\text{Ecut},i} = \begin{cases} 1 & E_i \ge E_{\text{cut}} \\ 0 & \text{otherwise} \end{cases} , \tag{4.34}$$

where E_i is the energy of the ALP candidate. Following the discussion in [79, 92], we set $E_{\text{cut}} = 200$ MeV for SHiP (lower than the nominal 1 GeV). For BDX, we set $E_{\text{cut}} = 300$ MeV in agreement with the observation made in [93]. We illustrate the effect of different energy cuts on the sensitivity curves in App. 4.C.

Parameter	**SHiP**	**BDX**
Beam	400 GeV protons	10.6 GeV electrons
Target	Molybdenum	Aluminum
Integrated Exposure (particles-on-target)	6×10^{20} POT	1×10^{22} EOT
Distance to decay d	33.5 m	20 m
Decay pipe/fiducial volume L	50 m	2.95 m
Transverse acceptance area A	4 m $\times$ 6 m	0.50 m $\times$ 0.55 m
Energy acceptance threshold E_{cut}	200 MeV	300 MeV

Figure 4.3: Benchmark experimental parameters used in our SHiP and BDX projections. For BDX there is no decay pipe; detectable decays must occur within the calorimeter fiducial volume.

In conclusion, the total number of events per POT is the sum of the final weights over all generated candidates:

$$\frac{N_{\text{events}}}{N_{\text{prim}}} = \sum_i \left(w_{\text{prod},i} \times w_{\text{decay},i} \times w_{\text{geo},i} \times w_{\text{Ecut},i} \right). \tag{4.35}$$

We note that in the LLP limit, $w_{\text{decay}} \propto 1/f^2$. Since $w_{\text{prod}} \propto 1/f^2$, the total event rate in the LLP limit scales as $N_{\text{events}} \propto 1/f^4$. Therefore, in the LLP limit the *shape* of the axion energy distribution can be rescaled by a factor f^4 to make it coupling-independent. In contrast, when the decay length is comparable to experimental length scales, the event weights become exponentially sensitive to f (c.f. Eq. (4.29)).

4.5 Experimental projections

In what follows, we focus on two representative experiments, SHiP and BDX, though our methodology applies broadly to other beam dump experiments. We have developed the publicly available code `ALPETITE`, which converts electromagnetic cascades and dark vector fluxes from `PETITE` simulations into ALP event predictions for arbitrary beam dump configurations. The modular design allows users to easily modify experimental parameters (geometry, POT, acceptance thresholds) and theoretical inputs (couplings, masses) without the need of regenerating Monte Carlo samples.

The SHiP experiment [63, 64] will use a 400 GeV proton beam incident on a molybdenum target with the goal of collecting 6×10^{20} POT. An axion is detectable if it decays visibly in some region of space after the beam dump. The version of the experiment that was approved in March 2024 is more compact than the original design. For long-lived particle signals considered here, the relevant detector subsystem is the hidden sector decay spectrometer (HSDS). The new decay pipe starts at $d = 33.5$ m from the target and is $L = 50$ m long, followed by a charged particle tracker, timing detector, electromagnetic and hadronic calorimeters. The tracker has an aperture of $4 \text{ m} \times 6$ m, which determines the angular acceptance of potential decays. Following [62], we simulated the N_γ primary on-shell photons arising from π^0, η, η' di-photon decays by using `PYTHIA 8.2` [94] with the `SoftQCD:all` flag and a pomeron flux parametrization `SigmaDiffractive:PomFlux(5)`, for $N_p = 10^5$ pp collisions at a center-of-mass energy of 27.4 GeV. We randomly selected a subsample of the above list with energy above a certain threshold and assigned an initial MC weight to each primary photon equal to

$$w_{\text{prim}} = N_\gamma(E_\gamma > 50 \text{ MeV})/N_p \times (1 - \mathrm{e}^{-1}) \approx 5.34\,, \tag{4.36}$$

where we added the exponential factor as a conservative choice to consider only the γ flux coming from the first interaction length. The weight approximately matches the number of expected photons per POT (~ 5) for SHiP in [95], Table I. For transparency, we show in App. 4.E (Fig. 4.15) the energy and transverse-momentum spectra of the sampled primary-photon list used to seed the SHiP cascade, together with a diagnostic comparison to the full parent photon population. Finally, we generated SM showers for the primary photons using `PETITE` and produced axion samples of various masses, assigning MC weights as described in Section 4.4.

Based on the current proposal [67], the BDX experiment at Jefferson Lab will use a 10.6 GeV continuous-wave electron beam incident on the Hall-A aluminum beam dump with 10^{22} electrons on target (EOT). The detector will be housed in a new underground facility located $d \sim 20$ m downstream of the dump, behind ~ 6 m of shielding. Its active volume will be a plastic-scintillator electromagnetic calorimeter, with fiducial dimensions $\sim 50 \text{ cm} \times 55 \text{ cm} \times 295$ cm, corresponding to a frontal acceptance area of about $\sim 0.275 \text{ m}^2$. Unlike SHiP, BDX will have no decay pipe, so detectable ALP decays must occur within the detector's fiducial volume. Consequently, we use $\theta_{\text{acc}}^{\text{BDX}} \equiv \sqrt{A}/2d \sim 0.013$. The benchmark experimental parameters used in our projections are summarized in Fig. 4.3.

Downstream fluxes

We begin our study with an analysis of the predicted ALP coupling-independent event rates in the LLP limit for the two experimental setups described above.

As an illustration, in Figs. 4.4 and 4.5 we present the individual contributions from each process described in Section 4.3 for $m_a = 100$ MeV ALP in the γ-dominated benchmark (SHiP) and the e-dominated benchmark (BDX). The spectra

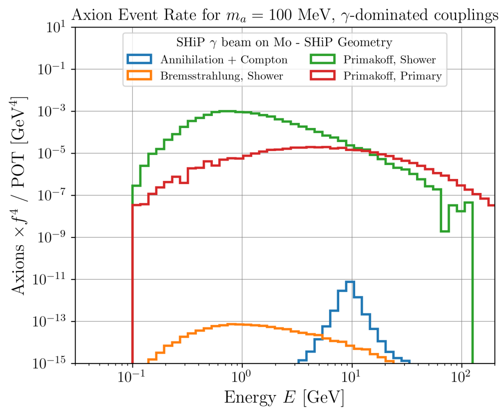


Figure 4.4: Event rate of axions per protons-on-target (POT), weighted by f^4, for an ALP with mass $m_a = 100$ MeV and photon-dominated couplings. The setup simulates SHiP photons emerging from meson decays impinging on a molybdenum target, with the corresponding SHiP experimental geometry. The flux is calculated for ALPs directed towards the detector. The individual production channels are: Primakoff conversion from shower photons (green); bremsstrahlung from shower electrons and positrons (orange); e^+e^- annihilation and Compton-like scattering of shower $e^\pm$ (blue); Primakoff conversion from the primary photon beam from neutral meson decays (red).

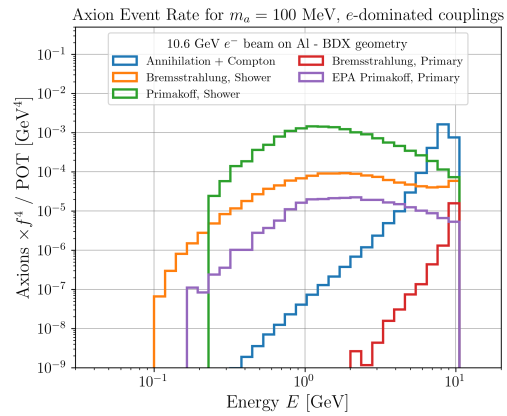


Figure 4.5: Event rate of axions produced per electrons-on-target (EOT), weighted by f^4, for an ALP with mass $m_a = 100$ MeV and electron-dominated couplings. The setup simulates a 10.6 GeV electron beam on an aluminum target, with a geometry representative of the BDX experiment. The flux is restricted to ALPs with trajectories pointing towards the forward detector. The contributions from different production mechanisms are shown separately: Primakoff conversion from shower photons (green); bremsstrahlung from shower electrons and positron (orange); e^+e^- annihilation and Compton-like scattering from shower particles (blue); bremsstrahlung (red) and photon fusion (purple) from the primary electron beam.

are histograms of the expected event rate as a function of energy. The quantity on the y-axis, "Axions $\times f^4$/ POT", is the weighted event rate per POT for each energy bin, with the coupling dependence factored out. This is constructed by summing the total weights of all MC candidates that fall within that energy bin and setting $1/f = 1$, such that the event rate depends only on the ALP mass and experimental setup.

Examining the plots, several salient features emerge. First, Primakoff production dominates over electron bremsstrahlung across most of the parameter space in both the electron- and photon-dominated benchmarks. In the electron-dominated scenario, this occurs because the RG-induced photon coupling, Eq. (4.5), is of comparable size with respect to the electron one. For $m_a = 100$ MeV, the peak

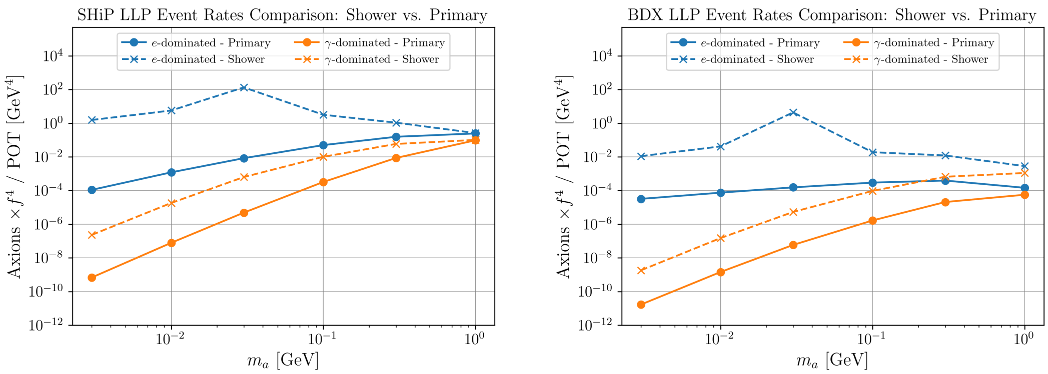


Figure 4.6: Comparison between the primary-only and the shower-induced coupling-independent weighted event rates in the LLP limit for different axion masses for SHiP (**left**) and BDX (**right**). As above, the axion count has been weighted by decay (in the LLP limit), geometry, and energy acceptance. In both panels, as the mass is increased, the ability of softer shower particles to produce axions is diminished, and production rates from the showers becomes smaller (in the case of SHiP) than the primary-only estimates for masses greater than 1 GeV.

Primakoff flux arising from the electromagnetic shower exceeds the primary-beam peak by $O(10^2)$; consequently, we expect at least a factor of 3 improvement in sensitivity to $1/f$ from this contribution alone in the LLP limit. Notably, even in the electron-dominated case, electron bremsstrahlung remains subdominant to Primakoff production once $m_a \gtrsim 100$ MeV. Resonant annihilation of positrons on atomic electrons yields competitive ALP fluxes at lower masses, with a spectrum peaked near $E_{\rm res} \simeq m_a^2/(2m_e)$. There are two production mechanisms in an electron beam dump experiment (e.g. BDX) from the primary e^- beam: bremsstrahlung, mediated by the axion-electron coupling (c_{aee}), and photon fusion, mediated by the (effective) axion-photon coupling ($\tilde{c}_{a\gamma\gamma}$). The dominance of one process over the other is model-dependent. The ratio of their raw integrated fluxes, Φ, without experimental cuts, is well approximated by the following expression, which we have confirmed via MC simulation:

$$\frac{\Phi_{\gamma-\text{fusion}}}{\Phi_{\text{brem}}} \sim \frac{\tilde{c}^2_{a\gamma\gamma}}{c^2_{aee}}\, \alpha\, \frac{m_a^2}{2m_e^2} \log^{-1}\left(\frac{m_a^2}{2m_e^2}\right). \tag{4.37}$$

For instance, in a scenario with equal couplings ($c_{aee} = \tilde{c}_{a\gamma\gamma}$, e.g. the e-dominated scenario of Section 4.2), photon fusion surpasses bremsstrahlung to become the leading *primary* process for axion masses $m_a \geq 20$ MeV.

In Fig. 4.6, we show the integrated ALP event rate in the LLP limit for both primary and shower-only particles, at several ALP masses. As in the histograms above, the yield is weighted by the decay probability, the detector energy threshold, and the angular acceptance. We render the result coupling-independent by factoring out the decay constant f and normalizing to the particles on target. A flux enhancement of $\mathcal{O}(10^4)$ corresponds to an $\mathcal{O}(10)$ gain in sensitivity at small coupling. For both experiments and both benchmarks, the flux gain is pronounced – especially at lower masses, where the broader phase space enables substantially more ALP production.

For SHiP, electromagnetic cascades enhance the flux by up to $\mathcal{O}(10^3)$ in the photon-dominated scenario and up to $\mathcal{O}(10^4)$ in the electron-dominated scenario at the lowest simulated mass (3 MeV). A comparably large enhancement also appears at higher masses (30 MeV) in the electron-dominated benchmark, primarily due to resonant annihilation.

For BDX, the flux enhancement is at least $\mathcal{O}(10^2)$ across all masses for both e-dominated and γ-dominated couplings, with the same resonant feature at m_a = 30 MeV in the electron-dominated case.

Sensitivity projections

In Figs. 4.7 and 4.8, we show how the inclusion of EM showers impacts the projected sensitivity of the experiments, accounting for a substantial increase in sensitivity in both electron-dominated and photon-dominated benchmarks. The showers effectively multiply the number of secondary particles (photons, electrons, and positrons) within the target, opening up or enhancing various ALP production channels. We estimate the sensitivity of each experiment using Eq. (4.35), and drawing $N_{\text{events}} = 5$ contours. The dependence on the assumed signal yield threshold (5, 10, and 100 events) is shown in App. 4.D. In `ALPETITE`, the function `compute_and_save_sensitivities` can very easily compute sensitivity curves for any number of events desired and/or particles-on-target.

Many searches for Beyond the Standard Model (BSM) physics at SHiP are anticipated to be nearly background-free [63]. The search for long-lived ALPs originating from the electromagnetic cascade is expected to share this characteristic. Indeed, preliminary analyses show no significant accumulation of background events at low track momenta [96]. Nevertheless, the precise background rate remains subject to final determination.

A core design principle of BDX is the achievement of a nearly background-free environment. This is accomplished mainly by placing the detector far downstream of the dump, behind significant earth shielding, to suppress beam-related backgrounds.

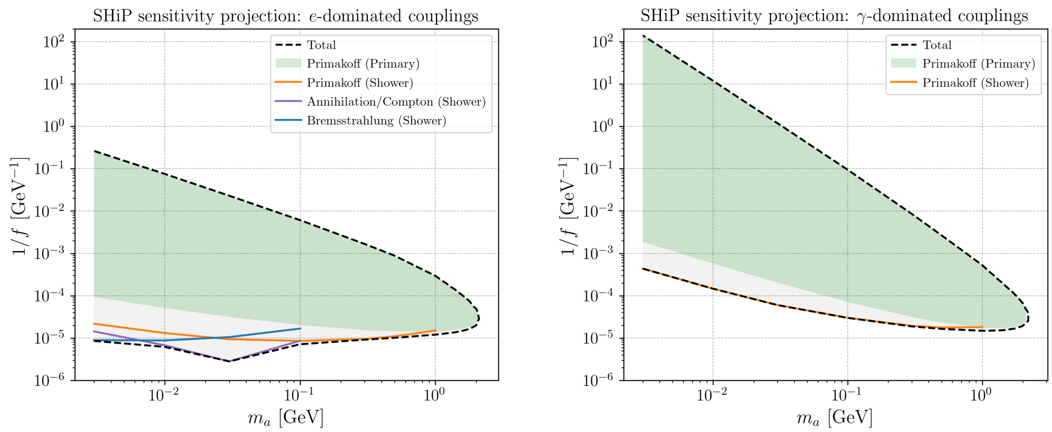


Figure 4.7: Projected sensitivity for 5 signal events for the SHiP experiment for the electron-dominated (**left**) and photon-dominated (**right**) ALP coupling scenarios, assuming an integrated 6×10^{20} POT and an energy acceptance threshold $E_{\rm cut} = 200$ MeV. The green shaded region shows the sensitivity for ALPs generated by primary photons arising from neutral meson decays. The solid lines show the individual contributions from shower-induced processes: Primakoff (orange), bremsstrahlung (blue), and resonant annihilation/Compton scattering (purple). The total sensitivity, combining primary and shower contributions, is shown by the dashed black line, with the gray shaded area indicating the full parameter space covered. For the photon-dominated case, we do not show the bremsstrahlung and resonant annihilation/Compton scattering lines since they are sub-dominant with respect to all the other processes.

Consequently, searches for light dark matter are projected to be exceptionally clean. Extensive Monte Carlo studies have validated this approach, demonstrating that beam-related backgrounds and cosmogenic rays are reduced to negligible levels by the shielding and active veto systems, respectively [67]. While these simulations are robust, the final background budget is subject to verification, and initial in-situ measurements from the BDX-MINI prototype have confirmed the effectiveness of the background mitigation strategies [97].

For the SHiP experiment (Fig. 4.7), which uses a proton beam, the primary source of high-energy SM particles capable of producing ALPs are the neutral mesons (π^0, η, η') that decay into photons. In the photon-dominated scenario (right panel), the sensitivity is driven by the Primakoff process. As already pointed out in previous studies [62, 65], the primary photons from meson decays provide a strong baseline sensitivity (green region). However, as these photons initiate EM showers, they generate a much larger flux of lower-energy secondary photons, which in turn produce ALPs. This shower contribution (orange line) substantially improves the sensitivity,

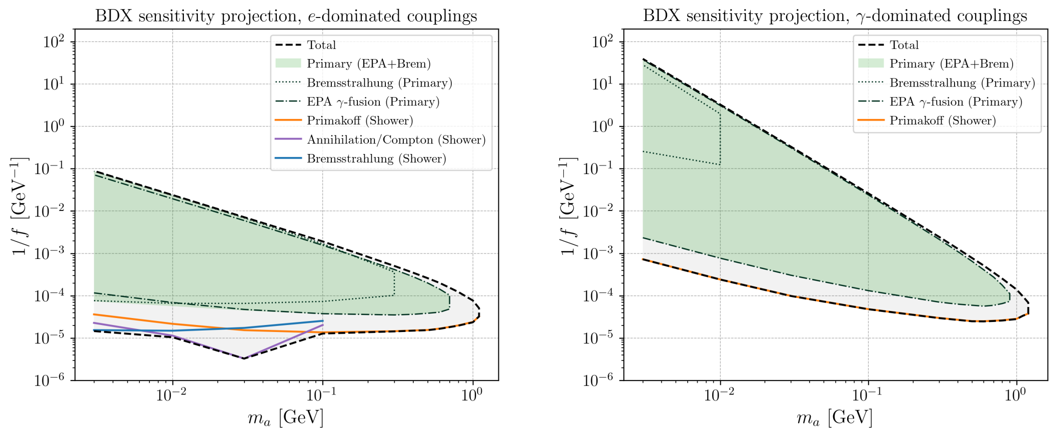


Figure 4.8: Projected sensitivity for 5 signal events for the BDX experiment for the electron-dominated (**left**) and photon-dominated (**right**) ALP coupling scenarios, assuming an integrated 10^{22} EOT and an energy acceptance threshold $E_{\rm cut} = 300\,\text{MeV}$. The green shaded region represents the contribution from ALPs generated by the initial electron beam, and it is obtained by summing the contributions from the bremsstrahlung (dotted green line) and the EPA γ-fusion process (dot-dashed green line). The solid lines show the contributions from the subsequent electromagnetic shower: Primakoff (orange), bremsstrahlung (blue), and annihilation/Compton (purple). The total sensitivity is shown by the dashed black line, which encompasses the gray shaded region. For the photon-dominated scenario, we do not show the bremsstrahlung and resonant annihilation/Compton scattering lines since they are sub-dominant with respect to the Primakoff channel.

particularly for lower ALP masses ($m_a \lesssim 100\,\text{MeV}$), where the phase space for production by lower-energy photons is larger.

The effect is even more dramatic in the electron-dominated scenario (left panel). Here, electron-coupling-dependent channels - like bremsstrahlung or resonant annihilation - are not available to primary photons. The sensitivity from primary photons alone (green region) cannot take advantage of the very efficient resonant annihilation channel (electron bremsstrahlung also plays an important role at lower masses). Once showers are included, the primary photons generate a cascade of secondary electrons and positrons. These secondaries then become a powerful source of ALPs via bremsstrahlung (blue line) and resonant annihilation/Compton (purple line). These shower-induced channels completely dominate the sensitivity for $m_a \leq 100$ MeV, extending the reach in $1/f$ by more than an order of magnitude.

Because the shower particles have smaller energies, the resulting axion event yields are more sensitive to energy cuts compared to those coming from the highest energy primary particles. Nevertheless, in App. 4.C we find that significant enhancements in sensitivity are obtained even when one requires $E_{\rm cut} \sim$ GeV.

For BDX (Fig. 4.8), the primary particles are electrons. In the electron-dominated scenario (left panel), the primary electrons can already produce ALPs via either bremsstrahlung or photon fusion (c.f. Eq. (4.37) for an estimate of the relative importance of the two processes). The shower simulation enhances this by tracking the multiplication of electrons and positrons as they traverse the dump. Consequently, the shower-induced bremsstrahlung (blue line) and annihilation/Compton (purple line) contributions provide the dominant sensitivity for lighter masses. Notably, the Primakoff contribution from on-shell photons dominates for $m_a \geq 100$ MeV, and already overcomes the (shower) bremsstrahlung contribution for $m_a \geq 20$ MeV, in analogy to the primary case (c.f. Eq. (4.37) and discussion below).

In the photon-dominated scenario for BDX (right panel of Fig. 4.8), the primary sensitivity curve is essentially dominated by the photon fusion process, across all the masses. Nevertheless, the electromagnetic cascade from the primary electrons generates a large number of photons, enhancing the ALP yield from the Primakoff process (orange line). As a result, the total sensitivity is increased by the shower-induced contribution, improving the reach across the entire mass range and at high masses.

In Fig. 4.9 we compare our projected sensitivities to other constraints in the literature. To ease comparison, we choose the photophilic scenario in which $c_{aee} = 0$ and one considers only tree-level production involving the $c_{a\gamma\gamma}$ vertex. As we saw above, this scenario is essentially equivalent to the γ-dominated one in Section 4.2; it naturally arises when the ALPs are coupled only to photons in the UV (e.g., at the weak scale) and the induced electron coupling at lower scales is so small that it can be neglected. For consistency with the results quoted in the literature, we recast all the bounds in terms of the dimensionful coupling $g_{a\gamma\gamma} \equiv f^{-1}(c_{a\gamma\gamma}\alpha/\pi)$, where we use the on-shell coupling $\alpha^{-1} = 137.036$, and take $c_{a\gamma\gamma} = 1$ in converting between f and $g_{a\gamma\gamma}$ (other choices of $c_{a\gamma\gamma}$ are obtained by trivial re-scalings).

As emphasized in Fig. 4.1, the sensitivity derived herein for SHiP is stronger, and therefore covers more parameter space than past literature [53, 62]. The sensitivity curve for BDX derived here is, to the best of our knowledge, new. We find that the BDX collaboration will be able to probe currently untouched parameter space, reaching up to $m_a \gtrsim 1$ GeV.

At lower masses (below roughly $m_a \lesssim 200$ MeV) the regime in which we have derived

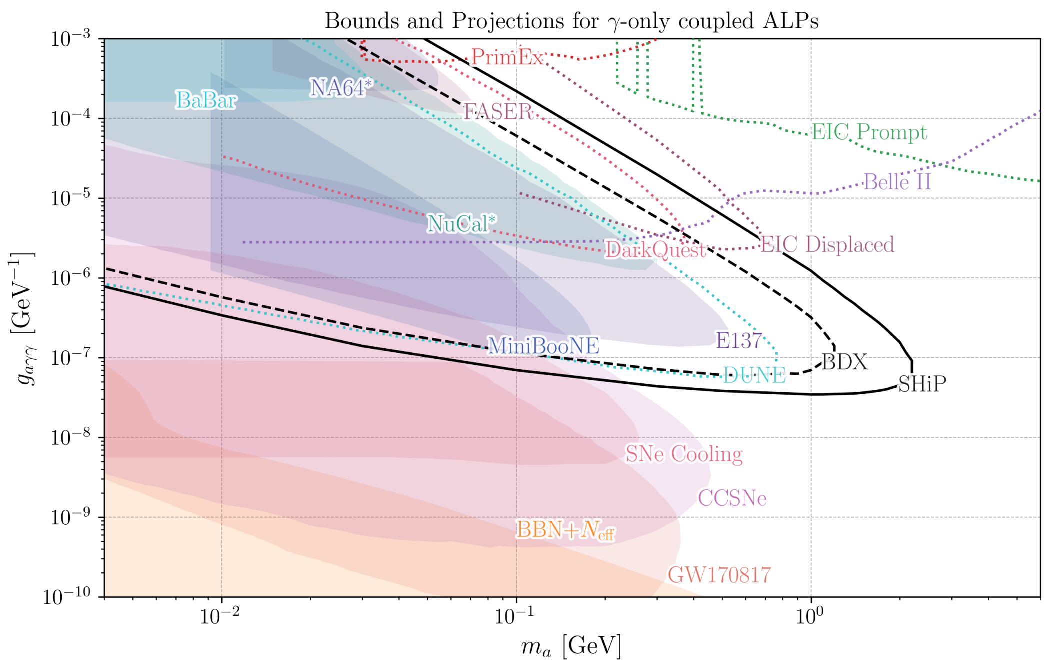


Figure 4.9: Comparison between our sensitivity projections for SHiP (black solid line) and BDX (black dashed line) with other relevant bounds (colored shaded regions) and projections (colored dotted lines) in the literature. We use $g_{a\gamma\gamma} \equiv f^{-1} c_{a\gamma\gamma}\alpha/\pi$, and we set $c_{aee} = 0$ (the same conventions as Fig. 4.1). Starred bounds were computed without taking into account any enhancement from secondary production. Sources: PrimEx (projection only, current bound not shown) [98], BaBar [99], NA64 [100], FASER (57.7 fb^{-1}) [101], EIC Prompt (10 fb^{-1} line) and EIC Displaced (S_a = 100, [10, 100] cm line) [102], Belle II [103], NuCal [62, 104, 105], DarkQuest (10^{18} POT) [71], MiniBooNE Beam Dump [59, 106], E137 [99], DUNE (GAr, 1 yr line) [57], SNe Cooling and CCSNe (1 B line) [107], BBN+$N_{\rm eff}$ [108], GW170817 fireball bound [109]. LUXE-NPOD Phase-I [110] projection curve is not shown since it is very close to DarkQuest.

enhanced sensitivity is largely covered by astrophysical constraints (in particular those stemming from supernovae). Nevertheless, the projections derived herein are complementary, and perhaps most importantly, can be obtained in a laboratory setting. Moreover, for $m_a \gtrsim 70$ MeV all of SHiP, BDX, and DUNE [57] are capable of probing axion-photon couplings $g_{a\gamma\gamma} \lesssim 2 \times 10^{-7}$ GeV^{-1} corresponding to axion decay constants above the weak scale $f \sim 10$ TeV.

Among the existing limits in the literature, some have incorporated electromagnetic cascades and others have not. For example, the E137 collaboration performed a Monte Carlo simulation of the electromagnetic shower in their aluminum and steel target station [61]. To the best of our understanding, their simulation did not include the incoherent Primakoff-like production discussed above; however, the quoted limit ends at a low enough mass such that coherent production dominates over incoherent production and we expect their constraint to be reliable.[3] Re-analyses of CHARM [62, 111] and NuCal [62, 104] have not included showers in the modelling of axion production in the target, and so we expect those experiments to have somewhat stronger reach than that shown in Fig. 4.9. The effects of the electromagnetic cascade are less dramatic when recasting old analyses due to their higher energy cuts. For example, at E137 the observed signal in the downstream detector was required to have energy greater than 1 GeV [61]. This effectively truncates the shower despite a growing photon multiplicity at lower energies (see Fig. 14 of [61]).

In summary, we expect modest improvements to the sensitivity derived from old experimental data through the inclusion of electromagnetic showers, but the gains are somewhat limited by the typically large energy cuts applied in these old analyses. This lesson is important for upcoming experiments. As discussed in [79], the SHiP experiment may be able to substantially extend its sensitivity by lowering its analysis thresholds. If BDX can achieve sub-300 MeV thresholds, this would also benefit their search for axion-like particles.

4.6 Conclusions

Axionlike particles are a well-motivated lamppost for new physics at very high scales. Being pseudo-Nambu-Goldstone bosons, they are naturally expected to be light, and when their decay constant is set by the high scale of new physics, they are further expected to be weakly coupled. When ALPs couple to hadrons (as is necessary for solutions of the strong-CP problem), they are efficiently searched for using direct production from meson decays or hadronic collisions. If, however, ALPs couple predominantly to leptons and/or electroweak gauge bosons in the ultraviolet, then leptonic and electromagnetic probes are best suited for their discovery.

In this chapter, we have investigated both proton and electron beam dump experiments in detail. We find that the electromagnetic cascade will essentially always dominate the production of "hadrophobic" ALPs. Furthermore, enhancements relative to naive estimates based on the first radiation length are so large that, in fact, electromagnetic cascades can compete with hadronic production even if gluon couplings are only somewhat smaller than photonic couplings (say by a factor of five). Such a hierarchy

[3] Only the long-lived limit was analyzed in [61], whereas the effect of attenuation due to decays was added later in [99].

can easily arise from a modest number of color-neutral fermions in the ultraviolet carrying electroweak quantum numbers.

Our main conclusions are as follows: when electron couplings are absent at tree-level, all electromagnetic phenomenology is dominated by the photon coupling. This statement holds essentially independently of the axion mass. By way of contrast, when electron couplings are present at tree-level they induce photon couplings which are capable of competing with electron-initiated processes. This crossover occurs around $m_a \simeq 100$ MeV with Primakoff production dominating above this mass and resonant annihilation dominating below it.

In summary, our results demonstrate that electromagnetic cascades are not merely a correction, but an essential component for accurately estimating the sensitivity of beam dump experiments to ALPs, in both electron- and photon-dominated scenarios. This was well understood by the E137 collaboration [61], but it is often missing from sensitivity projections for upcoming experiments. This mirrors similar conclusions in the context of dark photon searches where resonant annihilation and the electromagnetic cascade more generally can dramatically enhance sensitivity at the weakest couplings accessible to the experiments [56, 79, 82, 88–90].

Methodologically, our simulation framework goes beyond the simple one-dimensional approximation to shower development often used in the literature. By tracking the full electromagnetic cascade through `PETITE` - including transverse dynamics, multiple scattering, and realistic secondary production - we accurately predict angular distributions and thereby determine detector angular acceptances in realistic geometries.

While we have focused on SHiP and BDX, this lesson also applies to recasts of old experiments. For example, the sensitivity of NuCal [104] and CHARM [111] has likely been underestimated in the literature [62, 99, 112]. It would therefore be interesting to revisit the limits from these past experiments, including the missing production mechanisms discussed above.

Appendix 4.A Inelastic Primakoff-like scattering

In this appendix, we assess the impact of including inelastic Primakoff-like scattering – encoded by the inelastic nuclear form factor $G_{2,\text{inel}}$ from Eq. (4.12) – on the predicted ALP flux at SHiP. This production channel is often neglected in the literature but can provide a non-negligible contribution to the total rate. Figure 4.10 quantifies this effect by showing the ratio of the projected ALP flux calculated with the full form

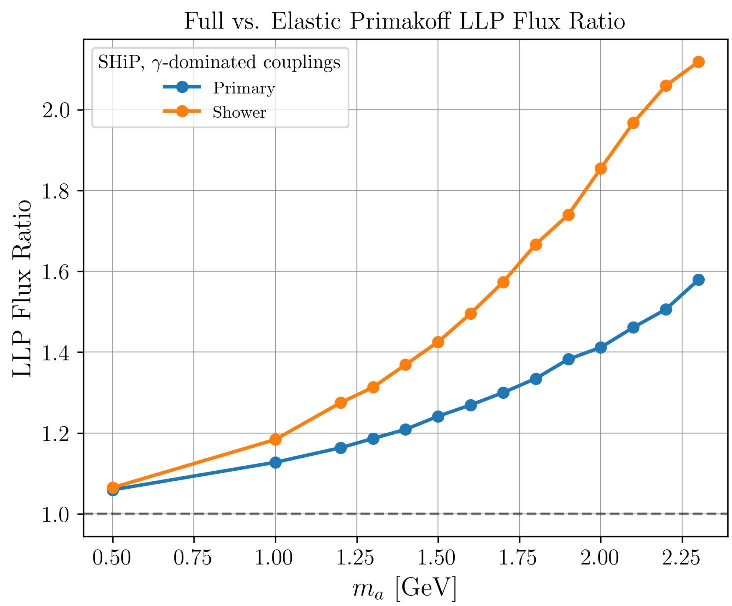


Figure 4.10: Ratio of the projected downstream ALP flux (in the LLP limit) at SHiP calculated with and without the inelastic Primakoff form factor ($G_{2,\mathrm{inel}}$), for γ-dominated couplings. The enhancement is shown separately for ALPs produced by primary photons from neutral meson decays (blue) and secondary photons from electromagnetic showers (orange). The inclusion of inelastic scattering provides an $O(1)$ enhancement to the flux, which becomes increasingly significant at higher ALP masses.

factor to that calculated with the coherent term alone ($G_{2,\mathrm{inel}} = 0$).

For both primary photons from meson decays and secondary photons from electromagnetic cascades, the inelastic channel provides an $O(1)$ enhancement. This effect is most pronounced at higher ALP masses, as the larger momentum transfer required for production preferentially probes the inelastic regime where the nucleus breaks up. Because this enhancement depends fundamentally on the effective ALP–photon coupling, a comparable gain is expected in the e-dominated scenario. While a factor-of-a-few increase in flux leads to only a modest shift in the final sensitivity contours, it is a crucial correction for accurately predicting event rates, particularly for heavier ALPs and for experiments using lighter target nuclei.

Appendix 4.B Monte Carlo production weights

In the following, we describe how we computed the production weights w_{prod} for each process described in Sec. 4.3. For initial particles with primary weights w_{prim} (e.g., SHiP primary photons), the production weight is simply multiplied by the value of the primary weight given.

Primakoff Primakoff production is controlled by nuclear and atomic (squared) form factors $G_{2,\mathrm{el}}(Q^2), G_{2,\mathrm{inel}}(Q^2) \leq 1$, defined above in Eqs. (4.11) and (4.12). It is therefore convenient to compute the relevant cross-section where the sum of these two factors is set to unity and subsequently apply a "screening weight" i.e.,

$$w_{\mathrm{prod},i} = \frac{\sigma_P}{\sigma_{\mathrm{SM}}}\left[G_{2,\mathrm{el}}(Q^2) + \frac{G_{2,\mathrm{inel}}(Q^2)}{Z}\right], \tag{4.38}$$

where σ_P is computed for an infinitely heavy, point-like, and unscreened nucleus of charge Z. We will refer to this as the "Born-Primakoff cross-section," in formulae

$$\begin{aligned}\sigma_P =& -\left(\frac{1}{f}\right)^2 \frac{Z^2\alpha^3|\tilde{c}_{a\gamma\gamma}|^2}{32\pi^2\omega^2}\int_{Q_-^2}^{Q_+^2} \mathrm{d}Q^2 \times \\ & \left(Q^4 + 2Q^2(m_a^2 - 2\omega^2) + m_a^4\right)/Q^4 \\ \equiv& \frac{\alpha g_{a\gamma\gamma}^2 Z^2}{8}\left[\left(1+\beta^2\right)\log\left(\frac{1+\beta}{1-\beta}\right) - 2\beta\right],\end{aligned} \tag{4.39}$$

where $g_{a\gamma\gamma} \equiv \alpha\tilde{c}_{a\gamma\gamma}/\pi f$, $\beta \equiv |\boldsymbol{p}_a|/\omega = \sqrt{1 - m_a^2/\omega^2}$ and the kinematic boundaries are $Q_\pm^2 = (\omega \pm |\boldsymbol{p}_a|)^2$. The Q^2 of the axion is sampled according the Born-Primakoff differential cross-section (given in Eq. (4.7)). This ensures that the expected value for w_{prod} corresponds to the full screened cross section. In formulae:

$$\langle w_{\mathrm{prod},i}\rangle = \frac{\sigma_P}{\sigma_{\mathrm{SM}}}\frac{\int \mathrm{d}Q^2 \frac{\mathrm{d}\sigma_P}{\mathrm{d}Q^2}F(Q^2)}{\int \mathrm{d}Q^2 \frac{\mathrm{d}\sigma_P}{\mathrm{d}Q^2}} \equiv \frac{\sigma_P^{\mathrm{screened}}}{\sigma_{\mathrm{SM}}}, \tag{4.40}$$

where $F(Q^2) \equiv G_{2,\mathrm{el}}(Q^2) + G_{2,\mathrm{inel}}(Q^2)/Z$.

For an incident electron beam, the photon fusion process $e^-Z \to e^-Za$ is simulated using the Equivalent Photon Approximation (EPA). From each primary electron of energy E_e, a fixed number of n_{samples} virtual photons are drawn to represent the EPA flux. The energy fraction of each photon, $x \equiv \omega/E_e$, is sampled from the probability distribution proportional to the EPA spectrum, $\mathrm{d}N_\gamma/\mathrm{d}x$, defined in Eq. (4.17). This sampled photon is then fed to the Primakoff production pipeline described above. The primary weight assigned to each photon is the total integrated photon flux per electron, divided by the number of samples:

$$w_{\mathrm{EPA}} = \frac{1}{n_{\mathrm{samples}}}\int_{x_{\min}}^{x_{\max}} \mathrm{d}x\,(\mathrm{d}N_\gamma/\mathrm{d}x) \approx \frac{0.42}{n_{\mathrm{samples}}}, \tag{4.41}$$

where we used $E_e = 10.6$ GeV, $x_{\min} = 50$ MeV$/E_e$, and $x_{\max} = 0.99$ for BDX simulation. In addition, the virtuality of the photon, Q^2, is sampled log-uniformly between its kinematic limits, $Q^2_{\rm min} \equiv m_e^2 x^2/(1-x)$ and $Q^2_{\rm max} \equiv x^2 E_e^2$. While the on-shell approximation is used for the Primakoff cross-section calculation, the sampled virtuality is kinematically relevant as it dictates the scattering angle of the equivalent photon and is crucial for applying angular acceptance criteria. The final MC weight for the axion produced by photon fusion is the product of the two, $w_{\rm EPA} \times w_{{\rm prod},i}$.

Dark Bremsstrahlung The kinematics of axion production by bremsstrahlung is identical to dark vector production; moreover, the squared matrix elements are similar (c.f., Eq. (4.19)). This means that we can simply rescale the weight w_V of dark vectors sampled by `PETITE` by the ratio of differential cross-sections:

$$w_{{\rm prod},i} = \frac{{\rm d}^3\sigma_a}{{\rm d}^3\sigma_V}\, w_V \ . \tag{4.42}$$

The differential cross-section used in `PETITE` is given by Eq. (10) in [113] (which makes use of results from [114, 115]),

$$\begin{aligned}\frac{{\rm d}^3\sigma}{{\rm d}x\,{\rm d}\cos\theta\,{\rm d}t} = \frac{\epsilon_V^2\alpha^3|\boldsymbol{k}|E_0}{|\boldsymbol{p}|\,|\boldsymbol{k}-\boldsymbol{p}|}\,\frac{G_2^{\rm tot}(t)}{t^2}\\ \times\left(\frac{1}{2\pi}\int_0^{2\pi}{\rm d}\varphi_q\,\frac{|A_V|^2}{8M^2}\right)\ ,\end{aligned} \tag{4.43}$$

and it is derived by the following interaction Lagrangian

$$\mathcal{L}_V = \epsilon_V e A'_\mu \bar\psi \gamma^\mu \psi \ . \tag{4.44}$$

In more detail, $\frac{1}{2\pi}\int_0^{2\pi}{\rm d}\varphi_q\,|A_V|^2$ is given by

$$A_V^{(0)} + Y\cdot A_V^{(1)} + \frac{1}{W^{1/2}}A_V^{(-1)} + \frac{Y}{W^{3/2}}A_V^{(-2)}\ , \tag{4.45}$$

where the different contributions are defined in Eqs. (15)-(18) of [113] and the equations above. The differential cross-section for the axion is given by making the substitution $A_V^{(i)} \to A_a^{(i)}$ in the φ_q integral, as follows

$$A_V^{(1)} \to A_a^{(1)} = \frac{4M^2}{\tilde{u}}\,, \tag{4.46}$$

$$\begin{aligned} A_V^{(0)} \to A_a^{(0)} = \frac{4M}{\tilde{u}^2}\Big[&-4E_0^2 M m_a^2 (x-1)^2 + 2E_0 t\big(m_a^2 - x(m_a^2+\tilde{u})\big) \\ &+ M\big(m_a^2 t + 2\tilde{u}^2\big)\Big]\,, \end{aligned} \tag{4.47}$$

$$\begin{aligned} A_V^{(-1)} \to A_a^{(-1)} = \frac{4M}{\tilde{u}}\Big[&E_0^2\big(8M m_a^2 (x-1) - 4Mtx^2\big) \\ &- 2E_0 t\big(m_a^2(x-2) + \tilde{u}x\big) + M\big(2m_a^2 t + \tilde{u}^2\big)\Big]\,, \end{aligned} \tag{4.48}$$

$$A_V^{(-2)} \to A_a^{(-2)} = 4M\,\big[m_a^2(-4E_0^2 M + 2E_0 t + Mt)\big]\,, \tag{4.49}$$

where the effective interaction lagrangian we are using is

$$\mathcal{L}_a = \epsilon_a e a \bar{\psi}\gamma_5\psi, \tag{4.50}$$

with $\epsilon_a \equiv 2m_e c_{aee}/ef$, being c_{aee}/f the coupling in Eq. (4.1). Recall that the derivative (Eq. (4.1)) and the pseudoscalar (Eq. (4.50)) interactions are equivalent at leading order in a/f [116].

Annihilation and Compton In these cases, since both processes have cross-sections proportional to the spin-averaged square amplitude $\langle |\mathsf{M}_V|^2\rangle$, the Monte Carlo weights of the dark vectors w_V are just rescaled by the axion cross-sections $\langle |\mathsf{M}_a|^2\rangle$, in formulae

$$w_{\text{prod},i} = \frac{\langle |\mathsf{M}_a|^2\rangle}{\langle |\mathsf{M}_V|^2\rangle} w_V = \frac{\epsilon_a^2}{\epsilon_V^2}\frac{m^2}{2(m^2+2m_e^2)} w_V, \tag{4.51}$$

where $m = m_V = m_a$, and we used the same interaction lagrangians in Eqs. (4.44) and (4.50).

Appendix 4.C Energy thresholds at SHiP

In Figs. 4.11 and 4.12, we analyze the impact of changing the energy threshold acceptance at SHiP as well as the effect of the geometric acceptance on the observed flux. As discussed in [79], the SHiP collaboration can optimize sensitivity to cascade produced long-lived particles by modifying their analysis strategy. In particular lower-energy thresholds are important to profit from the resonant annihilation of high energy positrons with atomic electrons.

In Fig. 4.11, we show different sensitivity curves at SHiP for different energy cuts, in the electron-dominated scenario (left panel) and photon-dominated one (right panel).

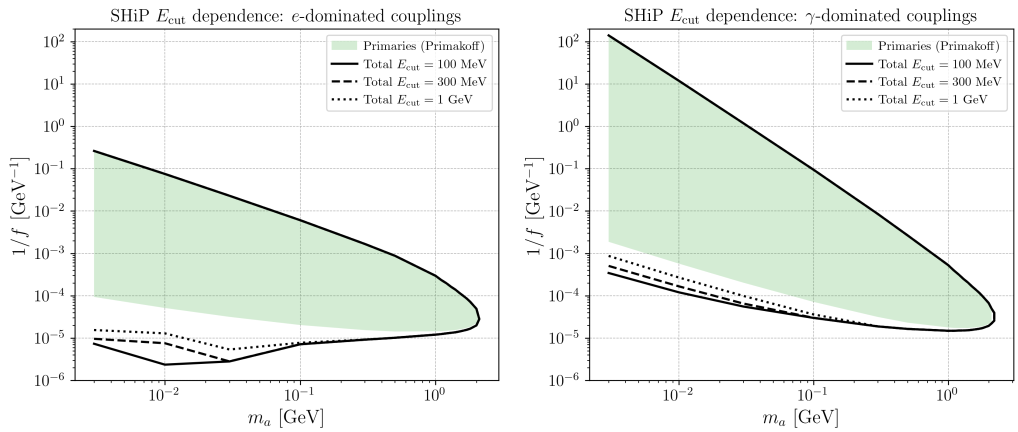


Figure 4.11: Sensitivity curves of SHiP for 5 signal event in the electron-dominated (**left panel**) and photon-dominated (**right panel**) benchmarks with different energy cuts applied. The green shaded region shows the sensitivity for ALPs generated by primary photons arising from neutral meson decays. The solid, dashed, and dotted black line, are indicating the full parameter space covered including the shower contributions for $E_{\rm cut} = 100,\ 300,\ 1000$ MeV respectively. We observe (almost) no dependence on the energy cut for the primary reach.

In the electron-dominated benchmark, the main effect is to impact the resonant annihilation sensitivity, which is one of the dominant active channels at lower masses. Indeed, the peak of the resonant flux lies at

$$E_{\rm res} \simeq \frac{m_a^2}{2m_e} \simeq 100 \text{ MeV} \left(\frac{m_a}{10 \text{ MeV}}\right)^2 . \tag{4.52}$$

For energy cuts $E_{\rm cut} \leq 300$ MeV, the sensitivity lower bounds for $m_a \geq 30$ MeV are not affected, and for lower masses, the shower gain at small couplings is still of $\mathcal{O}(10)$, even for thresholds as large as $E_{\rm cut} \sim 1$ GeV. In the photon-dominated scenario, Primakoff is the leading production channel. The effect of the different energy cuts is to suppress the flux (and therefore the gain in sensitivity) for smaller masses.

Fig. 4.12 illustrates the progressive application of experimental selections to the ALPs flux in the LLP limit produced via Primakoff conversion for both primary photons and electromagnetic showers. The first panel shows the raw production spectrum (rescaled by f^2) for various ALP masses $m_a = 3 - 1000$ MeV, exhibiting a pronounced enhancement at low masses due to the abundance of soft photons in the shower spectrum that efficiently convert when $m_a \ll E_\gamma$. The second panel includes the decay probability weight in the LLP limit $w_{\rm decay} \approx L/\lambda_i \propto 1/f^2$, which suppresses the high-energy tail where boosted ALPs escape the fiducial volume

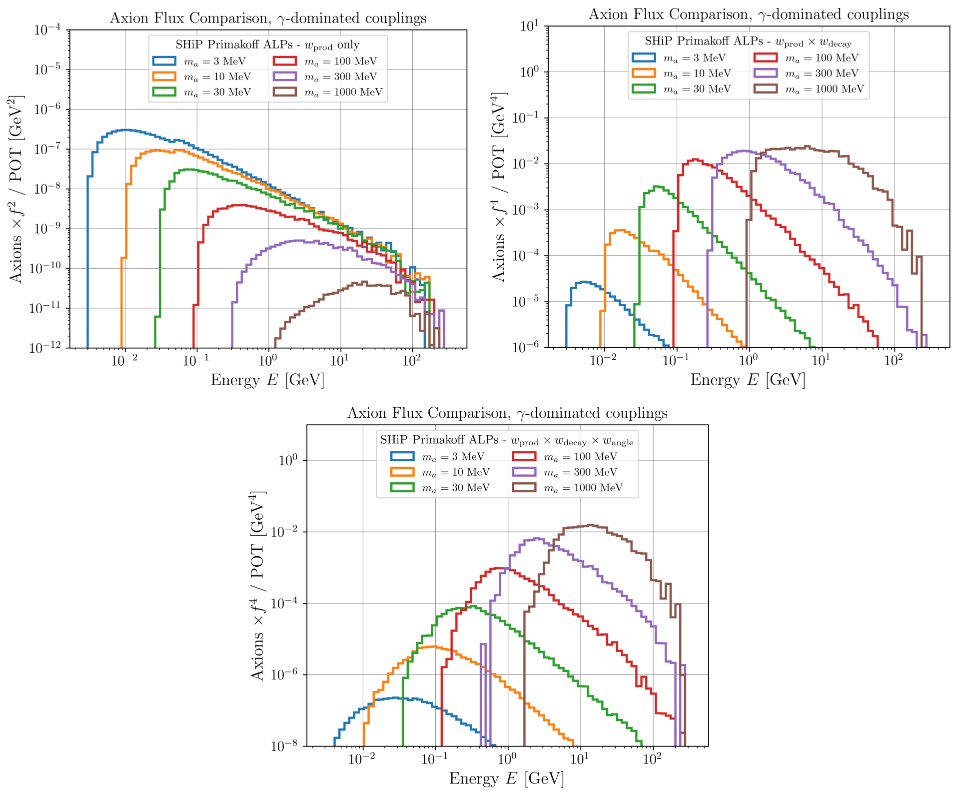


Figure 4.12: Evolution of the ALP flux through successive application of weights for $m_a = \{3, 10, 30, 100, 300, 1000\}$ MeV; note that we work in the LLP limit, and scale out $1/f^4$ such that weights can be greater than unity. **Left panel:** Raw production flux, i.e. with w_{prod} only, from Primakoff conversion (from both primaries photons arising from neutral meson decay and secondary photons arising from electromagnetic showers) showing enhanced production for lighter ALPs due to the abundance of soft photons in the showers. **Middle panel:** Decay-weighted flux in the LLP limit ($w_{\text{prod}} \times w_{\text{decay}}$), with high-energy suppression from ALPs escaping the decay pipe; note that w_{decay} is greater than unity because $1/f$ has been factored out. **Right panel:** Downstream predicted flux including the angular acceptance of the detector ($w_{\text{prod}} \times w_{\text{decay}} \times w_{\text{geom}}$), with no energy cut applied. The interplay between geometric acceptance and decay kinematics creates a characteristic double-cutoff structure that shifts to higher energies for heavier ALPs. All fluxes are normalized per proton on target (POT) and made coupling-independent by rescaling by the appropriate powers of f.

before decaying. The third panel incorporates the geometric acceptance weight w_{geom}, introducing a sharp low-energy cutoff at $E_a \sim m_a/\theta_{\max}$ where $\theta_{\max}$ is the angular acceptance of the decay volume. As ALPs' three-momentum goes to zero, they become increasingly transverse and fail the angular selection. The resulting peaked distribution represents the kinematically accessible phase space bounded by geometric acceptance at low energies and decay length considerations at high energies,

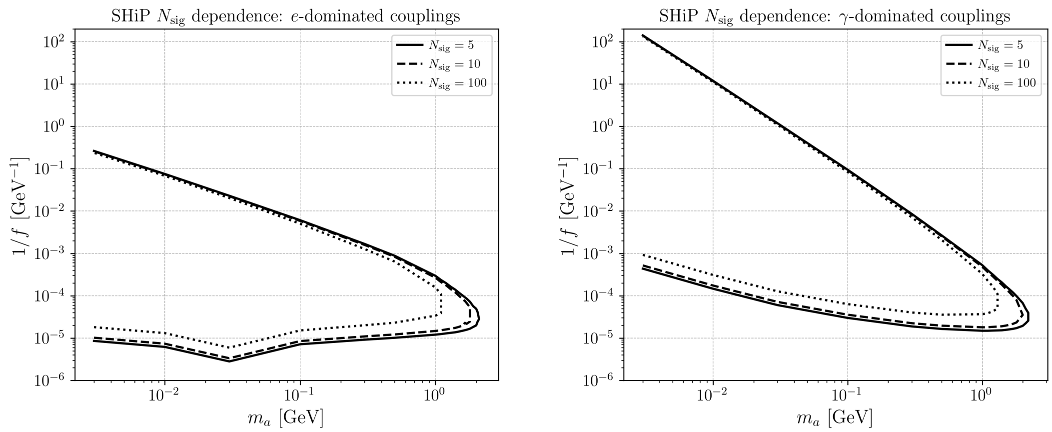


Figure 4.13: $N_{\rm sig}$ dependence of the projected SHiP sensitivity for the electron-dominated (**left panel**) and photon-dominated (**right panel**) ALP coupling scenarios. We show contours corresponding to $N_{\rm sig} = 5$ (solid), $N_{\rm sig} = 10$ (dashed), and $N_{\rm sig} = 100$ (dotted) signal events, assuming an integrated 6×10^{20} POT and an energy acceptance threshold $E_{\rm cut} = 200$ MeV.

with the peak position shifting to higher energies for heavier ALPs as expected. For SHiP geometry, $\theta_{\rm max} \sim 0.1$ which implies that $E_{\rm peak} \sim 10\, m_a$. Therefore, in order to fully profit from the shower enhancement for a fixed axion mass m_a, the optimal cut for maintaining the signal is $E_{\rm cut} \lesssim 10\, m_a$. A detailed study of signal vs. background is required to understand the optimal cuts for a full experimental analysis.

Appendix 4.D $N_{\rm sig}$ dependence

Throughout the main text we present baseline projections assuming background-free operation and a discovery threshold of $N_{\rm sig} = 5$ signal events. Since this choice is somewhat conventional, in this appendix we show how the projected reach changes as the required signal yield is varied. Figs. 4.13 and 4.14 display the projected sensitivity contours for $N_{\rm sig} = 5$, 10, and 100 events for SHiP and BDX, respectively, for both ALP coupling scenarios considered. As expected, the large-coupling boundary is only weakly affected by the choice of $N_{\rm sig}$, since it is primarily set by the requirement that the ALP survive to the detector (lifetime/geometry), while the small-coupling region is rate-limited and shifts more noticeably as $N_{\rm sig}$ is varied.

Appendix 4.E Primary photon spectra used to seed the SHiP cascade

In the SHiP projection, the electromagnetic cascade is seeded by primary on-shell photons produced predominantly in neutral-meson decays (e.g. $\pi^0 \to \gamma\gamma$) in the target. To generate the cascade efficiently, we draw a random subsample of size $N_{\rm samp}$ from the full generated photon population of size $N_{\rm parent}$ and assign them

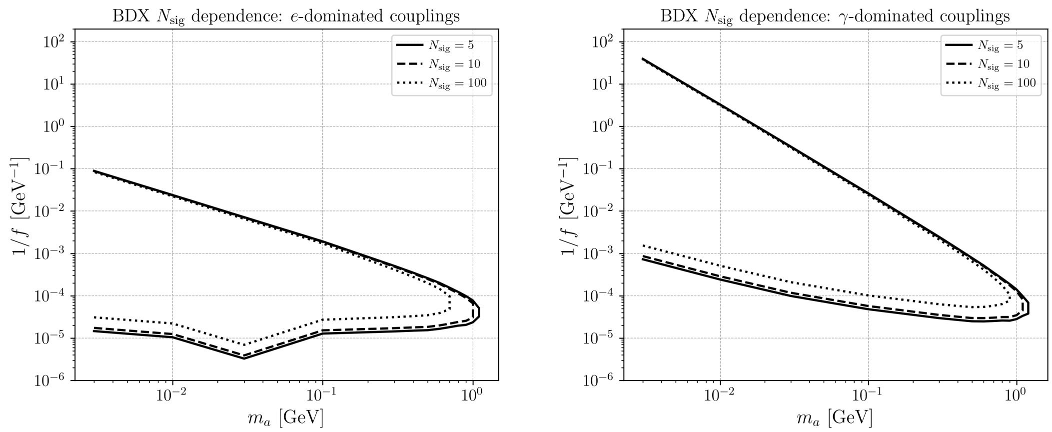


Figure 4.14: N_{sig} dependence of the projected BDX sensitivity for the electron-dominated (**left panel**) and photon-dominated (**right panel**) ALP coupling scenarios. We show contours corresponding to $N_{\text{sig}} = 5$ (solid), $N_{\text{sig}} = 10$ (dashed), and $N_{\text{sig}} = 100$ (dotted) signal events, assuming an integrated 10^{22} EOT and an energy acceptance threshold $E_{\text{cut}} = 300$ MeV.

Monte Carlo weights (cfr. Eq. (4.36) and related discussion) such that the subsample is a statistically representative proxy for the primary distribution.

Fig. 4.15 shows the resulting photon energy and transverse-momentum spectra for a sampled list used to seed one of the cascade simulations, together with a diagnostic comparison to the parent population. The histograms show the observed subsample counts, while the curves show the expected counts obtained by scaling the parent distributions by $N_{\text{samp}}/N_{\text{parent}}$. The good agreement verifies that the sampled photon list used as the cascade input reproduces the parent distributions.

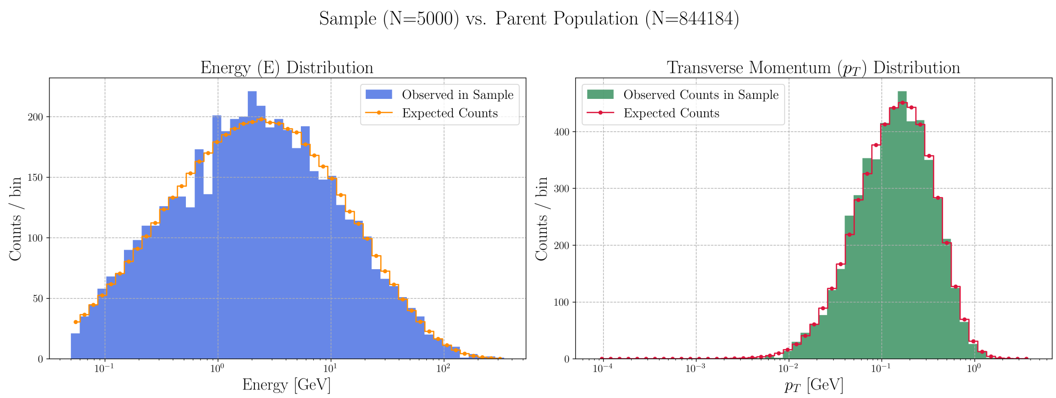


Figure 4.15: Primary-photon spectra used to seed the SHiP cascade simulation. The photon list shown is a random subsample of $N_{\rm samp} = 5 \times 10^3$ photons drawn from a parent population of size $N_{\rm parent}$, which constitute all the photons obtained from the 10^5 pp `PYTHIA`'s collision simulations. **Left panel**: photon energy spectrum. **Right panel**: photon transverse-momentum spectrum. Histograms show the observed subsample counts, while the curves show the expected counts obtained by scaling the parent-population distributions to the subsample size, $N_{\rm samp}/N_{\rm parent}$. Both panels use logarithmic binning on the horizontal axis. Photons with $E \leq 50$ MeV have been removed from the sampled list.

Chapter 5

LEPTOGENESIS IN AUTOMATIC NELSON-BARR MODELS

In this chapter, we numerically show that automatic Nelson-Barr models with new chiral fermions can simultaneously solve the strong CP problem and generate the observed baryon asymmetry via high-scale leptogenesis. In these models, all CP violation arises from a single spontaneous symmetry-breaking scale, linking the origin of quark and lepton CP phases. Using conservative assumptions and minimal dynamics, we identify a viable parameter window where successful leptogenesis occurs without spoiling the quality of the strong CP solution. Models with vector-like fermions face tension within this leptogenesis scenario. A key prediction is a correlation between the baryon asymmetry and the induced QCD vacuum angle shift. Remarkably, we find that the majority of the available parameter space is within reach of current and future nucleon EDM experiments.

5.1 Introduction

In spite of the enormous success of the Standard Model (SM) of particle physics in explaining nature, there is experimental evidence that cannot be accommodated within the model and calls for an extension either in field content, gauge symmetries, or both. Among those, measurements of the Cosmic Microwave Background (CMB) [19] and Big Bang Nucleosynthesis (BBN) [117, 118] manifest a primordial baryon asymmetry

$$\eta_B \equiv \frac{n_B^0 - n_{\bar{B}}^0}{n_\gamma^0} = (6.12 \pm 0.04) \times 10^{-10}, \tag{5.1}$$

where n_B^0 ($n_{\bar{B}}^0$) and n_γ^0 are the (anti-)baryon and photon number density today, respectively. A mechanism is required to dynamically generate this amount of baryon asymmetry, which demands ingredients beyond the SM [119]. It is noteworthy that the same new physics may also play a role in generating neutrino masses. At least two generations of active neutrinos with nondegenerate mass, in particular between 0.04 and 0.12 eV [19, 120], are a requisite to address the neutrino oscillation data [121–123]. These oscillations reflect lepton flavor violation, parametrized by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. Analogously, in the quark sector, the weak mixing is described by the CabibboLep-Kobayashi-Maskawa

(CKM) matrix, which contains an imaginary phase that violates a combination of charge-parity (CP) and parity (P) symmetries and, in general, contributes to the base-independent $\bar{\theta}_{\rm QCD}$ entering in the nucleon electric dipole moment (EDM),

$$\bar{\theta}_{\rm QCD} = \theta_{\rm QCD} + \arg\{{\rm Det}(\mathcal{M}_u \mathcal{M}_d)\} < 10^{-10}. \tag{5.2}$$

Here, $\mathcal{M}_u$ and $\mathcal{M}_d$ are the mass matrices of the up-type and down-type quarks, respectively, in the weak basis, and $\theta_{\rm QCD}$ is the QCD vacuum angle, sourced by the topological term $\mathcal{L}_{\rm QCD} \supset \alpha_s \theta_{\rm QCD} G^a_{\mu\nu} \tilde{G}^{\mu\nu a}/8\pi$. The constraint on the right-hand side comes from null measurements of the neutron EDM [124–126]. Eq. (5.2) defines the so-called strong CP problem[1].

Among the different beyond the SM mechanisms proposed to solve the aforementioned issue, two well-known candidates stand out for their simplicity: the Peccei-Quinn (PQ) mechanism [128, 129] and the Nelson-Barr (NB) mechanism [130–132]. In the PQ mechanism a pseudo-Goldstone boson, called the axion [47–52], associated to an anomalous global abelian symmetry (referred as PQ symmetry) dynamically sets $\bar{\theta}_{\rm QCD} = 0$ when it relaxes to the minimum of its potential, generated non-perturbatively by QCD instantons. Nelson-Barr models, on the contrary, start by assuming that CP is an exact symmetry of the Lagrangian, spontaneously broken by the true vacuum of the theory. Both mechanisms involve new scalar degrees of freedom and, in the case of the KSVZ axion [49, 50], new colored fermions as well. The QCD axion generally suffers from a severe quality problem [133]. Since gravitational effects are not expected to respect global symmetries, in order to preserve an acceptably small $\bar{\theta}_{\rm QCD}$, gauge invariant non-renormalizable operators violating the PQ symmetry of dimension equal to or higher than eleven must be forbidden by the symmetries and matter content of the ultraviolet (UV) theory [134–136][2]. Nelson-Barr models are easier to protect against gravitational effects and their quality is straightforwardly preserved in theories where the NB mechanism is an automatic consequence of the gauge symmetry and matter content [143, 144]. We refer to this kind of models as *automatic NB models*. Last but not least, PQ models provide automatically a cold dark matter candidate [145–147] while NB

[1] A cancellation between bare parameters in the Lagrangian may not be seen as a problem given that these parameters are nonphysical [127]. However, we expect the smallness of $\bar{\theta}_{\rm QCD}$ to be a consequence of a physical mechanism.

[2] This does not apply when the PQ symmetry emerges from a higher dimensional gauge field as in the string axiverse [137–139]. In modern language, this is understood in terms of the generalized global symmetries they possess [140–142].

models, in principle, do not[3]. Nevertheless, while PQ models are uncorrelated with the generation of a baryon asymmetry[4], baryogenesis is, to some extent, parasitic to Nelson-Barr models, as we aim to reinforce in this chapter.

A necessary condition for baryogenesis is the existence of a CP violating phase [119], which must be large enough to account for the observed asymmetry. In models where CP is spontaneously broken, the Lagrangian is real and the only contribution to a non-zero CP phase comes from the vacuum of the theory. It is therefore crucial to examine the dynamical generation of a baryon asymmetry in the context of models with spontaneous CP violation (SCPV) [152].

In NB models, the scale of SCPV is constrained from above by the need to preserve the quality of the mechanism, and therefore cannot be arbitrarily large. However, it cannot be arbitrarily low either, as inflation (or any alternative mechanism solving the horizon problem) must sufficiently dilute the domain walls associated with the spontaneous breaking of a gauged space-time symmetry [153]. Consequently, baryogenesis must occur within a specific energy window.

In this chapter, we study the viability of generating the observed baryon asymmetry in the context of automatic Nelson-Barr models. In these models, the CP phase from the vacuum plays a two-fold role: it generates the CKM phase [154] and the baryon asymmetry via leptogenesis. We assume that the NB mechanism operates similarly in both the quark and lepton sectors. This assumption is further motivated by expected new physics in the lepton sector – given the necessity of a mass generation mechanism for neutrinos – as well as recent data from the T2K [155] and NOνA [156, 157] experiments, which suggest a nonzero CP phase in the PMNS matrix[5].

Several studies in the literature have explored the generation of a baryon asymmetry within the framework of Nelson-Barr models. Branco and Parada extended the minimal implementation of the NB mechanism [159] to incorporate the lepton sector by using a discrete $\mathbb{Z}_4$ symmetry [160]. However, as noted in Ref. [143], such simplified models face tension with the thermal leptogenesis scenario. To address this, Ref. [161] relaxed the quality bound on the SCPV scale by introducing, on top of the previous $\mathbb{Z}_4$, an extra approximate $\mathbb{Z}_4$. Ref. [162] explored a $\mathbb{Z}_4 \times \mathbb{Z}_4'$,

[3] Ref. [148], for instance, shows how ultralight dark matter can arise in NB models with a shallow scalar potential.

[4] Attempts were made by Kuzmin, Tkachev, and Shaposhnikov [149], leading to negative conclusions. These were followed up, for example, by Servant [150], considering supercooled phase transitions, or by Co and Harigaya [151], considering the kinetic energy of the rotation of the axion.

[5] The joint fit of T2K and NOνA shows a mild preference for the inverted ordering, in which case a null CP phase is excluded at 3σ [158].

supplemented by additional scalars beyond the vector-like fermions necessary for the mechanism, thereby enabling a significantly lower reheat temperature. Instead, we focus on automatic Nelson-Barr models, where the required texture of the fermion mass matrices naturally arises as a consequence of a gauge symmetry.

The chapter is organized as follows. In Sec. 5.2 we describe the NB mechanism in the context of the minimal (simplified) model that can host it. In Sec. 5.3, the quality of the mechanism is discussed and different upper bounds on the SCPV scale are derived, depending on the nature of the theory. In Sec. 5.4, we extend the Nelson-Barr mechanism to the lepton sector and study the viability of the model in terms of the quality of the mechanism and the ability to generate the observed baryon asymmetry. In Sec. 5.5, we perform a numerical scan to explore the feasible regions of parameter space and the correlations between the relevant parameters. In Sec. 5.6, we emphasize a connection between leptogenesis and $\bar{\theta}_{\rm QCD}$ that can offer a pathway to experimentally probe these theories in the near future. We draw our conclusions in Sec. 5.7. Supplementary materials are provided in the appendices.

5.2 The Nelson-Barr mechanism in a Nutshell

SCPV allows for an alternative solution to the strong CP-problem without invoking a pseudogoldstone boson at low energies - the axion. The key idea of SCPV models is to assume that CP is an exact (gauge) symmetry of nature: this implies that there exists a basis in which the UV Lagrangian can be written in terms of real-valued parameters and, consequently, $\bar{\theta}_{\rm QCD} = 0$ at high energies. In order to reproduce the observed order-one phase of the CKM matrix, complex scalars beyond the SM Higgs are then added to the theory whose complex vacuum expectation value (vev) breaks dynamically CP. A further mechanism - such as a discrete symmetry - has to be introduced, as it was first noted by Nelson [130] and generalized by Barr [131], to enforce that $\arg\{\mathrm{Det}(\mathcal{M}_u \mathcal{M}_d)\}$ in Eq. (5.2) vanishes, thus preventing these complex phases to spoil $\bar{\theta}_{\rm QCD} = 0$ at tree-level.

To review how the Nelson-Barr mechanism works, we adopt the simplified model proposed by Bento, Branco and Parada (BBP) in Ref. [159]. Within this model, NB criteria are achieved by considering an extended version of the SM, where the only additional fields are a vector-like down-type quark, $D_{L,R} \sim (3, 1, -1/3)$, and a complex scalar singlet X, both odd under a further $\mathbb{Z}_2$ symmetry, under which the SM fields are even. The most generic renormalizable Lagrangian for the down-type quark mass terms can be written as

$$-\mathcal{L} \supset \mathsf{Y}_d^{ij}\,\bar{Q}_L^i H d_R^j + (\lambda_d^i X + \lambda_d^{\prime i} X^*)\bar{D}_L d_R^i + m_D \bar{D}_L D_R + \text{h.c.}, \tag{5.3}$$

where $i, j = 1, 2, 3$ are family indices; $H \sim (1, 2, 1/2)$, $Q_L \sim (3, 2, 1/6)$, and $d_R \sim (3, 1, -1/3)$ are, respectively, the SM Higgs doublet, left-handed and right-handed quarks; Y_d and $\lambda_d \neq \lambda'_d$ are Yukawa couplings encoded in a matrix and a vector, respectively, and m_D is the allowed Dirac mass for the additional vector-like quark. Assuming CP-invariance, we can always find a basis in which all the above couplings are real.

For a range of parameters in the Higgs potential, the new scalar field can acquire a complex vev of the form $\langle X\rangle = v_{\rm CP}\, e^{\mathrm{i}\alpha}/\sqrt{2}$, which breaks spontaneously CP and defines the scale of SCPV, $v_{\rm CP}$. Proceeding further, below the electroweak scale, the Higgs acquires a vev[6], $\langle H\rangle = v_H/\sqrt{2}$, leading to the following mass matrix for the down-type quarks:

$$\mathcal{M}_d^{AB} = \begin{pmatrix} \mathsf{M}_d^{ij} & 0 \\ \mu_i & m_D \end{pmatrix} . \tag{5.4}$$

Above, $\mathcal{M}_d^{AB}$ is written in block form, with the indices $A, B = 1, \ldots, 4$; $i, j = 1, 2, 3$, and

$$\begin{aligned} \mathsf{M}_d^{ij} &= \quad \mathsf{Y}_d^{ij} v_H/\sqrt{2}, \\ \mu_i &= \quad (\lambda_{d,i}\, e^{\mathrm{i}\alpha} + \lambda'_{d,i} e^{-\mathrm{i}\alpha}) v_{\rm CP}/\sqrt{2}\,. \end{aligned} \tag{5.5}$$

The crucial observation is that, due to the structure of the matrix, $\arg\{\mathrm{Det}\,\mathcal{M}_d\} = 0$, implying $\bar{\theta}_{\rm QCD} = 0$ at tree-level. Notice that, if $\lambda_d = \lambda'_d$, all the entries in the mass matrix would become real, preventing the phase from surviving and getting propagated at low energies.

The mass matrix $\mathcal{M}_d$ can be diagonalized using a bi-unitary transformation such that $U_L^\dagger \mathcal{M}_d U_R = \mathrm{diag}(\widehat{\mathsf{M}}_d, \bar{m}_D)$, where $\widehat{\mathsf{M}}_d$ is the diagonal 3×3 matrix of the SM down-quark masses and $\bar{m}_D$ is the physical mass of the new vector-like quark, given by $\bar{m}_D^2 \simeq m_D^2 + \sum_{i=1}^3 |\mu_i|^2$. This relation holds when the mass of the new vector-like down quark is much heavier than the standard model down-quarks, already guaranteed by the gap between the bottom mass and bounds from direct searches on new colored vector-like quarks [163]. Inverting $U_L^\dagger \mathcal{M}_d \mathcal{M}_d^\dagger U_L = \mathrm{diag}(\widehat{\mathsf{M}}_d^2, \bar{m}_D^2)$, we

[6] The complex phase associated to the SM Higgs vacuum can be rotated away by the residual electroweak symmetry in the scalar potential.

can write

$$(\mathsf{V}\widehat{\mathsf{M}}_d^2\mathsf{V}^\dagger)_{ij} = (\mathsf{M}_d)_{ik}\left(\delta_{k\ell} + \frac{\mu_k^*\mu_\ell}{\bar{m}_D^2}\right)\left(\mathsf{M}_d^T\right)_{\ell j} + O([\mathsf{M}_d]^4[\mu]^2/\bar{m}_D^4)\,. \tag{5.6}$$

where V is the 3×3 upper-left block of U_L. V is nothing but the contribution from the down-quark sector to the CKM matrix, i.e. $\mathsf{V} \equiv \mathsf{V}_{\text{CKM}}$ if the up-quark mass matrix is diagonal in the weak basis. Here and thereafter, we use the notation $[a] \equiv O(|a_i|)$ and $[A] \equiv O(|A_{ij}|)$ for any vector a or matrix A.

We observe that the 3×3 effective mass matrix for the SM quarks in the right-hand side of the above equation is approximately diagonalized by the matrix V, which contains an order-one phase if $[\mu] \sim \bar{m}_D$. V is unitary up to $O([\mathsf{M}_d]^2[\mu]^2/\bar{m}_D^4) < 10^{-6}$. Furthermore, according to the analysis done in Ref. [164], $2 \lesssim [\mu]/m_D \ll 10^3$ is required in order to simultaneously reproduce the CKM phase and the SM quark masses.

5.3 Quality of Nelson-Barr Solution

One criticism of the Nelson-Barr mechanism is its sensitivity to the physics in the ultraviolet [165] (which is known in the literature as quality issues), namely the contribution of higher-dimensional operators and radiative corrections to $\bar{\theta}_{\text{QCD}}$ that can spoil the null tree-level result. We discuss them in what follows.

Higher dimensional operators

Higher-dimensional operators, ultimately suppressed by the Planck scale, are expected to induce shifts in the QCD vacuum angle. The effect on $\bar{\theta}_{\text{QCD}}$ from corrections to the tree-level mass matrix of Eq. (5.4), $\delta\mathcal{M}_d$, can be parametrized as[7]

$$\begin{aligned}\Delta\bar{\theta}_{\text{QCD}} &= \text{Im}\{\text{Tr}\{\mathcal{M}_d^{-1}\delta\mathcal{M}_d\}\} + O(\text{Tr}\{\mathcal{M}_d^{-1}\delta\mathcal{M}_d\}^2)\\ &\simeq \frac{[\text{Im}\{\delta\mathcal{M}_d^{ij}\}]}{[\mathsf{M}_d]} + \frac{[\text{Im}\{\delta\mathcal{M}_d^{44}\}]}{m_D} + \frac{[\mu]}{m_D}\frac{[\text{Im}\{e^{\mathrm{i}\alpha}\delta\mathcal{M}_d^{i4}\}]}{[\mathsf{M}_d]}\,,\end{aligned} \tag{5.7}$$

where we assumed $\mu_i = e^{\mathrm{i}\alpha}[\mu]$. It is worth noting that $\delta\mathcal{M}_d^{i4}$ could be real and still contribute to $\Delta\bar{\theta}_{\text{QCD}}$ as indicated above, while $\delta\mathcal{M}_d^{4i}$ does not contribute at leading order. The terms $\delta\mathcal{M}_d^{AB}$ in Eq. (5.7) contain phases coming from the vev $\langle X\rangle$. Since we are concerned about order of magnitude estimates of the quality bounds, we will take $[\text{Im}\{\delta\mathcal{M}^{AB}\}] \sim [\delta\mathcal{M}^{AB}]$: this is motivated by knowing that $O(1)$ phases are needed in order to reproduce the measured Jarlskog determinant [154], by assuming no strong phase cancellations.

[7] See, for example, the appendix in Ref. [2] for a derivation.

In the following, we start discussing the BBP simplified model [159] (see Sec. 5.2). As we will show, the low quality of this solution motivates us to consider automatic Nelson-Barr models, where the role of the $\mathbb{Z}_2$ symmetry is taken on by an extra (spontaneously broken) gauge symmetry in the UV [2, 143].

Imposed $\mathbb{Z}_2$ symmetry. Consider the Lagrangian in Eq. (5.3), which represents the minimal implementation of the NB mechanism. In this context, the following dimension-five operators are allowed (among others),

$$\frac{1}{\Lambda}\bar{Q}_L H D_R(\xi_1 X + \xi_2 X^*) + \frac{1}{\Lambda}\bar{D}_L D_R\big(\xi_3 X^2 + \xi_4 (X^*)^2\big) + \text{h.c.}, \tag{5.8}$$

where Λ is an ultraviolet scale. In the broken phase, the above operators contribute to the down-quark mass matrix, potentially compromising the tree-level prediction $\bar{\theta}_{\rm QCD} = 0$. Using Eq. (5.7), the shift in $\bar{\theta}_{\rm QCD}$ can be parametrized as

$$\Delta\bar{\theta}_{\rm QCD} \simeq [\xi]\frac{v_{\rm CP}}{\Lambda}\left(\frac{[\mu]}{m_D}\frac{1}{[\mathsf{Y}_d]} + \frac{v_{\rm CP}}{m_D}\right), \tag{5.9}$$

where $[\mathsf{Y}_d] = [\mathsf{M}_d]/v_H$, and $[\xi]$ quantifies the magnitude of the Wilson coefficients in Eq. (5.8). We take $[\xi] = 1$ in the following estimates. As we argued in Sec. 5.2, to generate a quark mass spectrum consistent with experiment, $m_D \sim [\mu]$ [164]. Note that, as long as $m_D/v_{\rm CP} > [\mathsf{Y}_d]$, the dominant contribution to $\bar{\theta}_{\rm QCD}$ comes from the operators proportional to $\xi_{1,2}$ in Eq. (5.8). In the following, we will make the conservative choice[8] of taking $[\mathsf{Y}_d] = \sqrt{2}\, m_d/v_H = 3\times 10^{-5}$. Then, $\Delta\bar{\theta}_{\rm QCD} < 10^{-10}$ implies that

$$v_{\rm CP} \lesssim 40~\text{TeV}\left(\frac{1}{[\xi]}\right)\left(\frac{\Delta\bar{\theta}_{\rm QCD}}{10^{-10}}\right)\left(\frac{m_D/[\mu]}{1}\right)\left(\frac{\Lambda}{1.2\times 10^{19}~\text{GeV}}\right), \tag{5.10}$$

where we have considered the ultimate scenario where higher dimensional operators are suppressed by inverse powers of the Planck scale.

The spontaneous breaking of the CP symmetry generates domain walls [153] that are not observed. One of the most popular mechanisms to dilute them is inflation. However, such a low upper bound is in tension with its simplest implementations (see Ref. [143] for a detailed discussion). One could consider particular scenarios where CP is not restored at high temperatures, which would prevent the formation of dangerous topological defects at the cost of a reduced parameter space in the scalar

[8] One could, for example, adopt the assumption of minimal flavor violation [166, 167]. A spurion analysis would identify $[\xi] = [\mathsf{Y}_\mathsf{d}]$, which would substantially relax the upper bound shown in Eq. (5.10). We refer the reader to Ref. [143] for more details.

potential [168, 169]. Alternatively, higher CP breaking scales compatible with the quality of the mechanism can be achieved if the role of the $\mathbb{Z}_2$ symmetry is replaced by a gauge symmetry in the UV. We discuss this scenario in the upcoming subsection.

Automatic Nelson-Barr. The tree-level zeroes in $\mathcal{M}_d$, see Eq. (5.4), play a crucial role in the implementation of the Nelson-Barr mechanism. Theories where these zeroes are an accident of the gauge symmetry group and matter content are referred to as *automatic Nelson-Barr models*.

For simplicity, we extend the SM gauge group with an abelian U(1)[9]. The zeroes arise when $Q_{D_R} \neq Q_{d_R}$, being Q_f the gauge charge of the f field under the new force. At least two new scalar fields are needed to generate a non-zero CP phase (see discussion in Refs. [144, 170]): we will use the subindex $a = 1, 2$ in X_a to distinguish between them. The charge requirement $Q_{X_a} = Q_{D_L} - Q_{d_R}$ must hold for the *bridge* term needed to implement the NB mechanism, $\bar{D}_L d_R X_a$, to be allowed. Since we work in a simplified model, anomaly cancellation in a UV completion may require additional charge constraints and/or extra fermions.

Generically, unless $|Q_{X_a}| = |Q_{D_R} - Q_{d_R}|$, the non-renormalizable operators leading to the aggressive bound in Eq. (5.10) are forbidden. However, if $Q_{D_L} = Q_{D_R}$, dimension-five operators of the kind

$$\frac{\xi_{ab}}{\Lambda} \bar{D}_L D_R X_a^* X_b \,, \tag{5.11}$$

analogous to the operators weighted by $\xi_{3,4}$ in Eq. (5.8), still contribute to $\bar{\theta}_{\rm QCD}$. They lead to the following upper bound on the scale of CP violation

$$v_{\rm CP} \lesssim 10^9 \text{ GeV} \left(\frac{1}{[\xi]}\right) \left(\frac{\Delta\bar{\theta}_{\rm QCD}}{10^{-10}}\right) \left(\frac{m_D/v_{\rm CP}}{1}\right) \left(\frac{\Lambda}{1.2\times 10^{19}\text{ GeV}}\right), \tag{5.12}$$

where we have taken the optimistic choice $m_D \sim v_{\rm CP}$, and Λ to be the Planck scale.

The bound above can be further relaxed if the SM vector-like new quarks are chiral under the new gauge symmetry ($Q_{D_L} \neq Q_{D_R}$). In this case, m_D can be associated to the breaking scale of the new gauge group, v_S. We assume $v_S \gtrsim v_{\rm CP}$. The leading corrections to the tree-level down-type quark mass matrix from non-renormalizable operators in this scenario occur at dimension-six,

$$\frac{1}{\Lambda^2} X_a^* X_{b\neq a} \left(\xi_{1,ab} \bar{D}_L D_R S + \xi_{2,ab} \bar{Q}_L H d_R\right) + \frac{\xi_{3,a}}{\Lambda^2} \bar{Q}_L H D_R X_a^* S + \text{h.c.}, \tag{5.13}$$

[9]Non-abelian extensions could also enforce the required texture.

leading to

$$
\begin{aligned}
[\delta \mathcal{M}^{44}] &= \xi_{1,ab}\, v_{\rm CP}^2\, v_S/\Lambda^2\,,\\
[\delta \mathcal{M}^{ij}] &= \xi_{2,ab}\, v_{\rm CP}^2\, v_H/\Lambda^2\,,\\
[\delta \mathcal{M}^{i4}] &= \xi_{3,a}\ \ v_{\rm CP}\, v_S\, v_H/\Lambda^2\,.
\end{aligned}
$$

Here, S is the scalar responsible for the breaking of the new gauge group. Remarkably, the last operator in Eq. (5.13) is always allowed in the presence of the bridge term, $\bar{D}_L d_R X_a$, due to the chiral nature of the new quarks under the new gauge symmetry. The above contributions generate a shift to $\bar{\theta}_{\rm QCD}$, which according to Eq. (5.7) is given by

$$
\Delta\bar{\theta}_{\rm QCD} \simeq [\xi]\frac{v_{\rm CP}^2}{\Lambda^2}\left[\frac{1}{[\mathsf{Y}_d]}\left(1+\frac{[\mu]}{m_D}\frac{v_S}{v_{\rm CP}}\right)+\frac{v_S}{m_D}\right]. \tag{5.14}
$$

Since $v_S \gtrsim v_{\rm CP}$ and $[\mu] \sim m_D$, the second term between the round brackets is expected to be larger or equal than one. In order to derive an upper bound on the CP violating scale, let us assume that $v_{\rm CP} \sim v_S$, and $[\mu]/v_{\rm CP} < [\mathsf{Y}_d]$, which leads to

$$
v_{\rm CP} \lesssim 7\times 10^{11}\ {\rm GeV}\left(\frac{\Delta\bar{\theta}_{\rm QCD}}{10^{-10}}\right)^{\frac{1}{2}}\left(\frac{1}{[\xi]}\right)^{\frac{1}{2}}\left(\frac{v_{\rm CP}/v_S}{1}\right)^{\frac{1}{2}}\left(\frac{\Lambda}{1.2\times 10^{19}\ {\rm GeV}}\right). \tag{5.15}
$$

In general, the second term in the square brackets in Eq. (5.14) may dominate the shift in the angle, imposing a tighter bound. As we will show at the end of the next subsection, the finite naturalness criterion [32, 144] requires the new down-quarks to be on the threshold of direct searches, $\bar{m}_D \sim 2$ TeV (see Ref. [163] and references therein, and the discussion in the next subsection). The upper bound in this case reads

$$
v_{\rm CP} \lesssim 3\times 10^{10}\ {\rm GeV}\left(\frac{\Delta\bar{\theta}_{\rm QCD}}{10^{-10}}\right)^{\frac{1}{3}}\left(\frac{1}{[\xi]}\right)^{\frac{1}{3}}\left(\frac{\Lambda}{1.2\times 10^{19}\ {\rm GeV}}\right)^{\frac{2}{3}}\left(\frac{m_D}{2\ {\rm TeV}}\right)^{\frac{1}{3}}. \tag{5.16}
$$

Radiative corrections

Quantum corrections start threatening the mechanism at one loop, via the diagrams displayed in the upper panel of Fig. 5.1. These induce a shift in $\bar{\theta}_{\rm QCD}$ given by

$$
\Delta\bar{\theta}_{\rm QCD} \sim \frac{\lambda_4}{16\pi^2}\frac{\bar{m}_D^2}{m_X^2}\ln\left(\frac{m_X^2}{m_b^2}\right), \tag{5.17}
$$

where λ_4 is the quartic coupling weighting the $H^\dagger H X_a X_b^*$ $(a \neq b)$ interaction. Here, we used $v_{\rm CP}$ as the scale for the mass of the X fields, and the mass of the bottom

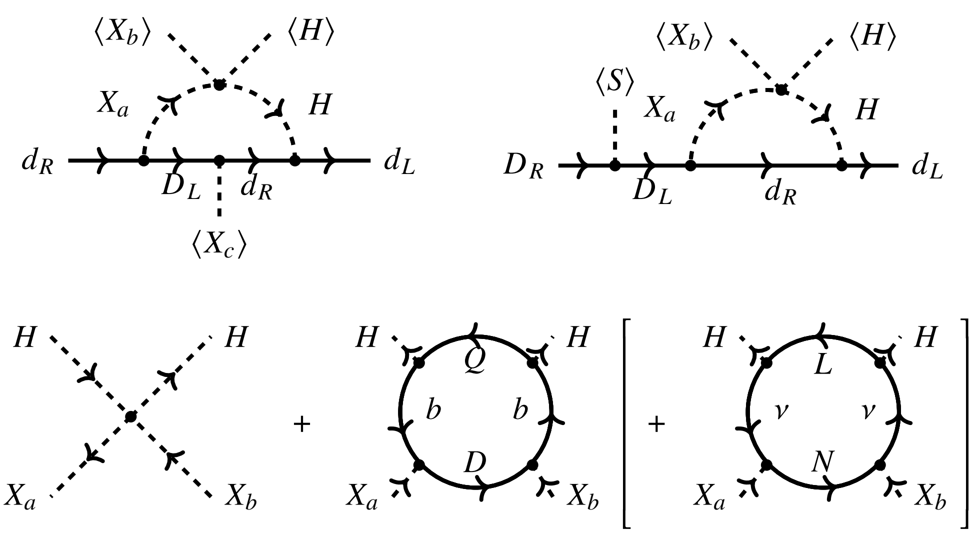


Figure 5.1: Upper panel: Radiative corrections at 1-loop contributing to $\bar{\theta}_{\rm QCD}$ by inducing complex quark masses. Lower panel: Tree-level and 1-loop corrections to the quartic coupling $H^\dagger H X_a X_b^*$. In square brackets we add the would-be contribution from extending the Nelson-Barr mechanism to the neutral lepton sector (including right-handed neutrinos).

quark as a subtraction scale. $\Delta\bar{\theta}_{\rm QCD} < 10^{-10}$ requires

$$\lambda_4 \lesssim 2\times 10^{-8}\left(\frac{v_{\rm CP}^2}{\bar{m}_D^2}\right)\frac{1}{\ln\!\left(v_{\rm CP}^2/m_b^2\right)}. \tag{5.18}$$

After SCPV, λ_4 contributes to the Higgs mass, so that one expects that, at tree-level, $\lambda_4|_{\rm tree} < 10^{-12}(10^8\ {\rm GeV}/v_{\rm CP})^2$. However, λ_4 cannot be arbitrarily small, as it will eventually be dominated by radiative corrections. The heavy down quarks contribute radiatively to the quartic coupling (see the bottom panel of Fig. 5.1), inducing a shift of [144]

$$\delta\lambda_4|_{\rm 1-loop} \simeq \frac{1}{16\pi^2}\frac{m_b^2}{v_H^2}\frac{\bar{m}_D^2}{v_{\rm CP}^2}\ln\left(\frac{\bar{m}_D^2}{m_b^2}\right). \tag{5.19}$$

The finite naturalness criterion demands $\delta\lambda_4|_{\rm 1-loop}v_{\rm CP}^2 < m_H^2$, being m_H the SM Higgs mass. As a consequence, $\bar{m}_D \sim 2$ TeV as we already mentioned in the previous subsection. Eq. (5.18) is still satisfied by the radiative quartic as long as $\bar{m}_D < 0.1\, v_{\rm CP}$. One could read the latter condition as a lower bound on the CP violating scale, $v_{\rm CP} > 20$ TeV.

5.4 Leptogenesis

In the previous sections we focused on the implementation of the NB mechanism in the quark sector of the SM. In general, one would expect that the UV completion hosting the NB mechanism acts *democratically* on all fermions. In appendix 5.A, we show that leptons are naturally involved in the automatization of the NB mechanism. Moreover, recent data from the T2K [155] and NOνA [156, 157] experiments hint towards a non-zero phase in the PMNS matrix, which motivates extending the NB mechanism to the lepton sector. In the following, we study whether the observed baryon asymmetry can be generated in the context of automatic Nelson-Barr models. We discuss the implications of successful leptogenesis in the available parameter space, considering that both quarks and leptons share the source of CP violation.

A simplified model

In the context of automatic Nelson-Barr models, the lagrangian in Eq. (5.3) is promoted to

$$-\mathcal{L} \supset \bar{Q}_L \Upsilon_d H d_R + \bar{D}_L(\lambda_{d1} X_1 + \lambda_{d2} X_2) d_R + \lambda_D S \bar{D}_L D_R + \text{h.c.}. \tag{5.20}$$

Above, the fields are charged under a new U(1) such that the zeroes in the down quark mass are guaranteed, which occurs when the new charges of the SM right-handed down quarks and the new D_R are different, i.e. $Q_{d_R} \neq Q_{D_R}$. We note that the condition $Q_X = Q_{D_L} - Q_{d_R}$ must also be required for the CKM matrix to inherit a complex phase. Furthermore, in order to relax the quality bound as discussed in Sec. 5.3, we demand the new quarks to be chiral under the new U(1); the charge condition $Q_S = Q_{D_L} - Q_{D_R}$ allows them to get a mass proportional to v_S, the spontaneously symmetry-breaking scale of this new force.

In the broken phase, $\langle X_a \rangle = e^{i\alpha_a} v_{X_a}/\sqrt{2}$, $\langle S \rangle = v_S/\sqrt{2}$ and $\langle H \rangle = (0, v_H/\sqrt{2})^T$. The down-quark mass matrix is shown in Eq. (5.4), with $m_D = \lambda_D v_S/\sqrt{2}$, and μ_i now given by

$$\mu_i = (\lambda_{d1} e^{i\alpha_1} v_{X_1} + \lambda_{d2} e^{i\alpha_2} v_{X_2})/\sqrt{2}. \tag{5.21}$$

Without loss of generality, we will assume that $v_{X_1} \sim v_{X_2} \sim v_{\text{CP}}$. As in the BBP model presented in Sec. 5.2, here $\bar{\theta} = \arg\{\det\{\mathcal{M}\}_d\} = 0$ at tree-level. In Sec. 5.3, we extensively discussed the interplay between renormalizable operators and/or radiative corrections and the induced shifts in $\bar{\theta}$, examining under which conditions those are consistent with the current experimental bounds on the neutron electric dipole moment (nEDM).

In Appendix 5.A, we show how UV completions of the Lagrangian in Eq. (5.20) naturally involve leptons charged under the new force, and we discuss in detail some concrete charge assignments.

Mirroring the quark sector, in a bottom-up approach to the problem, we now identify the general requirements for the lepton sector. We therefore extend the SM with three right-handed neutrino singlets, ν_R, and a single pair of neutral leptons, N_L and N_R, vector-like under the SM. The relevant Lagrangian for the neutral leptons is given by

$$\begin{aligned}-\mathcal{L} \supset\, & \bar{L}_L \mathsf{Y}_\nu \tilde{H} \nu_R + \bar{L}_L \lambda_\nu \tilde{H} N_R + \bar{N}_L \mu \nu_R + m_N \bar{N}_L N_R \\ & + \tfrac{1}{2}\overline{\nu_R^c}\,\mathsf{M}_R \nu_R + \tfrac{1}{2} m_L \overline{N_L^c} N_L + \tfrac{1}{2} m_R \overline{N_R^c} N_R + \tfrac{1}{2}\overline{N_R^c} m_{RR} \nu_R + \text{h.c.},\end{aligned} \tag{5.22}$$

where $\tilde{H} = i\sigma_2 H^*$. Assuming the same features of the quark sector, we require the charges of the fields under the new $\mathrm{U}(1)$ gauge group to satisfy the following relations:

(i) $Q_H = Q_{\nu_R} - Q_{L_L}$, so that SM neutrinos get a Dirac mass term with the three ν_R;

(ii) $Q_X = Q_{N_L} - Q_{\nu_R}$, to transfer the phase to the SM (complex PMNS matrix);

(iii) $Q_{N_R} \neq Q_{\nu_R}$, to enforce the zeroes of the Nelson-Barr texture;

(iv) $Q_{N_L} \neq Q_{N_R}$, to avoid a vector-like mass term for N_L and N_R.

Condition (ii) involves the scalar fields X responsible for inducing a phase in the CKM matrix (quark sector). Regarding condition (iv), we will assume that the scalar responsible for the $\mathrm{U}(1)$ breaking, S, singlet under the SM, has charge $|Q_S| = |Q_{N_L} - Q_{N_R}|$ under the new symmetry, so that N_L and N_R get mass dynamically.

Under (i)-(iv), in general, there are no Majorana mass terms allowed, and the previous Lagrangian, analogously to the quark sector, simply reads

$$-\mathcal{L} \supset \bar{L}_L \mathsf{Y}_\nu \tilde{H} \nu_R + \bar{N}_L (\lambda_{\nu_1} X_1 + \lambda_{\nu_2} X_2) \nu_R + \lambda_N S \bar{N}_L N_R + \text{h.c.}. \tag{5.23}$$

However, there are particular charge assignments leading to surviving Majorana masses. In the following, we assume that the ν_Rs get a dynamical Majorana mass from the spontaneous symmetry breaking of the new gauge group, while the rest of the Majorana mass terms are forbidden[10]. Coupling the SCPV fields X_a as above represents a conservative assumption in the sense that it does not introduce other complex parameters in the Lagrangian except the ones arising from the bridge term $(\bar{N}_L X_a \nu_R)$. In fact, within this choice, $\mathsf{M}_R = \mathsf{Y}_R v_S$ is a real, symmetric matrix. The relevant Lagrangian for our study is then given by,

$$\begin{aligned} -\mathcal{L} \supset\ & \bar{L}_L \mathsf{Y}_\nu \tilde{H} \nu_R + \bar{N}_L(\lambda_{\nu 1} X_1 + \lambda_{\nu 2} X_2)\nu_R + \lambda_N S \bar{N}_L N_R + \tfrac{1}{2} S\, \overline{\nu_R^c}\, \mathsf{Y}_R \nu_R + \text{h.c.}, \\ \xrightarrow{\langle S\rangle, \langle X_a\rangle}\ & \bar{L}_L \mathsf{Y}_\nu \tilde{H} \nu_R + \bar{N}_L \mu\, \nu_R + m_N \bar{N}_L N_R + \tfrac{1}{2} \overline{\nu_R^c}\, \mathsf{M}_R \nu_R + \text{h.c.}, \end{aligned} \tag{5.24}$$

where $m_N = \lambda_N\, v_S \in \mathbb{R}$, and $\mu = \lambda_{\nu 1}\langle X_1\rangle + \lambda_{\nu 2}\langle X_2\rangle \in \mathbb{C}$. From the perturbativity of the Yukawa couplings, the following inequalities hold $[\mathsf{M}_R], m_N \lesssim v_S$, and $[\mu] \lesssim v_{\text{CP}} \leq v_S$. Eq. (5.24) involves two heavy scales v_S and v_{CP}. Should condition (iv) be relaxed, an additional scale m_N, independent of v_S, would emerge.

One concrete example of a UV completion leading to the Lagrangians in Eqs. (5.20) and (5.24) is the model discussed in Ref. [144], with $|Q_S| = 2|Q_{\nu_R}|$. In Appendix 5.A, we discuss this particular example, as well as other explicit examples of UV theories that satisfy the conditions (i)-(iv) in both the quark and lepton sectors.

The simplified model described above contains the ingredients for a type-I see-saw mechanism [171–174] and calls for thermal leptogenesis. In the following, we will address whether the observed baryon asymmetry can arise within this framework, consistently with the quality constraints discussed in Sec. 5.3.

High-scale leptogenesis

High-scale leptogenesis via the decay of heavy Majorana neutrinos has been extensively studied in the literature. For comprehensive reviews on this topic, see for example Refs. [175, 176]. In the vanilla leptogenesis scenario [177], the lightest heavy neutrino, N_1, is responsible for leptogenesis. The Yukawa couplings parametrizing the interactions of the heavy neutrinos with the SM lepton doublets and the Higgs boson, in general c-numbers, can satisfy the required ingredients to generate a lepton asymmetry [119, 177]: (i) they transfer the total lepton number violation coming from the Majorana mass term to the SM, (ii) they can source the needed CP violation

[10] Under particular choices of charges, other Majorana mass terms could also be allowed. For example, on top of $\mathsf{M}_R \neq 0$, $m_L \neq 0$ in Eq. (5.22) if $Q_S = -2Q_{\nu_R}$ and $Q_X = -2Q_{N_L}$, while m_R, $m_{RR} \neq 0$ if $Q_S = 2Q_{\nu_R}$ and $Q_X = -2Q_{N_R}$. In scenarios where the only Majorana mass is m_L, m_R or m_{RR}, i.e. $|Q_S| \neq 2Q_{\nu_R}$, the rank of the heavy neutral lepton mass matrix is reduced, resulting in an increase in the number of light neutrinos. Further constraints need to be considered in such models to avoid spoiling the observed number of relativistic degrees of freedom, N_{eff} [19].

via the interference between 1-loop and tree-level decay of the heavy neutrinos, (iii) their smallness can help the departure from thermal equilibrium. Subsequently, the lepton asymmetry is transferred to the baryon sector via the sphaleron processes, which violate baryon (B) and lepton (L) numbers in three units while preserving $B-L$ [178]. The mass of the lightest heavy neutrino, M_{N_1}, is related to the active neutrino masses via the type-I seesaw mechanism. Since there is a minimum amount of CP asymmetry required to generate the observed baryon asymmetry, see Eq. (5.1), the upper bound on the active neutrino masses translates into a lower bound on the lightest heavy neutrino mass, $M_{N_1} \gtrsim 10^8$ GeV, commonly referred to as the Davidson-Ibarra (DI) bound [179].

The Nelson-Barr model in Eq. (5.24), however, has two major differences with respect to the vanilla scenario studied in the literature:

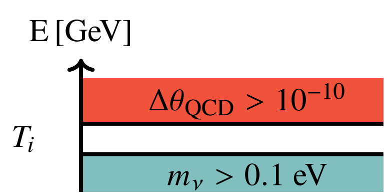
(i) The starting inputs are real Yukawa couplings and Majorana masses; the only CP phase comes from the vacuum of the theory, $\langle X_a \rangle$. Notice that such a phase is shared between the quark and the lepton sectors, and is inherited at low energies by both the CKM and the PMNS matrices.

(ii) There are in total five heavy neutral leptons, whose mass parameters come, up to perturbative Yukawa couplings, from two scales: the CP breaking scale, v_{CP}, and the new gauge symmetry breaking scale, $v_S \gtrsim v_{CP}$.

Difference (i) implies that leptogenesis must occur below the SCPV scale. Yet, v_{CP} cannot be arbitrarily high, as it is bounded from above to preserve the quality of the NB mechanism, as we discussed in Sec. 5.3. Whatever mechanism is responsible for diluting the topological defects arising from the breaking of CP [153] should occur below v_{CP} and above the scale of leptogenesis; otherwise, the lepton asymmetry would also be erased.

We then require the first temperature of the universe, T_{RH}, to be above the lowest starting temperature for successful leptogenesis, dubbed T_i. Hence, we expect

$T_i \le T_{RH} \le v_{CP}$, or, schematically,



(5.25)

This hierarchy delineates the window of scales where leptogenesis within automatic Nelson-Barr models is viable. The upper boundary of this window is set by the quality of the mechanism (the red region is disfavored by unacceptably large shifts to $\bar{\theta}_{\rm QCD}$ from higher dimensional operators). This reflects on different upper bounds on $v_{\rm CP}$, depending on the specific NB model considered. The lower boundary arises from a minimal allowed temperature for successful leptogenesis, T_i, which is typically of the order $T_i \sim M_{N_1}$. As we discuss in the upcoming subsections, M_{N_1} is bounded from below by the Davidson-Ibarra bound as a consequence of a maximal allowed value for the heaviest active neutrino (the turquoise region is ruled out by active neutrino masses that exceed observational limits).

Our aim in this chapter is to study the interplay between the two boundaries and show whether such a window exists and how it maps onto the parameter space of the theory.

Majorana neutrino masses

In the broken phase, Eq. (5.24) generates the following Majorana masses for the eight neutral leptons of the theory,

$$-\mathcal{L} \supset \frac{1}{2}\, f_N^T C\, \mathsf{M}^{8\times 8}\, f_N + \text{h.c.}, \qquad \text{with} \quad \mathsf{M}^{8\times 8} = \begin{pmatrix} 0^{3\times 3} & \mathsf{M}_D^{3\times 5} \\ (\mathsf{M}_D^T)^{5\times 3} & \mathsf{M}_H^{5\times 5} \end{pmatrix}, \tag{5.26}$$

where $f_N^T = (\nu_{L_1}^c, \nu_{L_2}^c, \nu_{L_3}^c, N_L^c, \nu_{R_1}, \nu_{R_2}, \nu_{R_3}, N_R)$, and

$$\mathsf{M}_D = \begin{pmatrix} 0 & & 0 \\ 0 & \mathsf{m}_D^{3\times 3} & 0 \\ 0 & & 0 \end{pmatrix}, \quad \text{and} \quad \mathsf{M}_H = \begin{pmatrix} 0 & \mu_1 & \mu_2 & \mu_3 & m_N \\ \mu_1 & m_{R_1} & 0 & 0 & 0 \\ \mu_2 & 0 & m_{R_2} & 0 & 0 \\ \mu_3 & 0 & 0 & m_{R_3} & 0 \\ m_N & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.27}$$

Above, $\mathsf{m}_D = (\mathsf{Y}_\nu \mathsf{O}_R) v_H/\sqrt{2}$, where O_R is the orthogonal matrix that diagonalizes M_R after rotating the ν_Rs, that is, $\mathsf{O}_R^T \mathsf{M}_R \mathsf{O}_R = \text{diag}(m_{R_1}, m_{R_2}, m_{R_3})$. Without loss of generality, we choose O_R such that $m_{R_1} \leq m_{R_2} \leq m_{R_3}$. We also note that the μ_i masses in M_H correspond to those in Eq. (5.4), up to a reshuffling induced by O_R.

Applying the seesaw block diagonalization [180], the three light and five heavy masses in the physical basis are given by

$$\begin{aligned} \widehat{\mathsf{M}}_L &= -\mathsf{U}_L^T \mathsf{M}_D \mathsf{M}_H^{-1} \mathsf{M}_D^T \mathsf{U}_L + \mathcal{O}([\mathsf{m}_D]^3/[\mathsf{M}_R]^2), \\ \widehat{\mathsf{M}}_H &= \mathsf{U}_H^T \mathsf{M}_H \mathsf{U}_H + \mathcal{O}([\mathsf{m}_D]^2/[\mathsf{M}_R]), \end{aligned} \tag{5.28}$$

where $\mathsf{U}_{L,H}$ are the unitary matrices that diagonalize, *à la Takagi* [181, 182], the light and heavy mass matrices.

We note that, considering the structure of the mass matrices outlined above, the masses of the active neutrinos depend only on the light Dirac masses m_D^{ij}, and m_{R_i}. Consequently, μ_i and m_N are not affected by the heaviest active neutrino mass bound, $m_\nu \lesssim 0.1$ eV [19, 183, 184]. A significant implication is that, unlike the vanilla leptogenesis scenario widely studied in the literature, M_{N_1} cannot be generically identified as the lightest eigenvalue of M_R, m_{R_1}. We find that, for a mass matrix texture as given in Eq. 5.26, the lightest heavy neutrino mass is related to the mass scales in M_H as

$$M_{N_1} \simeq \mathrm{Min}\left\{m_N, m_{R_1}, m_{R_i}\left(\frac{m_N^2}{m_N^2+|\mu_i|^2}\right)\right\}. \tag{5.29}$$

We refer the reader to Appendix 5.B for its analytical derivation in the scenario of three scales (corresponding to degenerate entries of M_R).

Eq. (5.27) allows us to locate the complex phases in the heavy neutrino mass matrix. In the limit $[\mu] \ll [\mathsf{M}_R]$ or m_N, M_H becomes symmetric and real up to $\mathcal{O}([\mu]/\mathrm{Max}\{[\mathsf{M}_\mathsf{R}], m_N\})$. This limit is disfavored by leptogenesis as the complex phases needed to generate the CP asymmetry would be suppressed. For example, if all Yukawa couplings in the new sector were $\mathcal{O}(1)$, heavy masses would be associated with the vevs generating them: $[\mathsf{M}_R] \sim m_N \sim v_S$ and $[\mu] \sim v_{\mathrm{CP}}$. If this is the case, we expect the vevs to be correlated, i.e. $v_{\mathrm{CP}} \sim v_S$.

Fig. 5.2 summarizes different hierarchies that the heavy neutrino mass spectrum can exhibit, depending on the relationships among the scales participating in M_H. Taking into account that scenarios with $[\mu] \gtrsim m_N$, $[\mathsf{M}_R]$ are preferred by leptogenesis,

(i) If $[\mu]$ is the heaviest scale (panel C), $M_{N_1} \simeq \mathrm{Min}\left\{m_{R_1}, m_{R_i} m_N^2/|\mu_i|^2\right\}$. Generically there will be a mass gap between the lightest, N_1, and the second lightest, N_2, heavy neutrinos. Hence, vanilla leptogenesis is expected to occur.

(ii) If $[\mu] \sim m_N \gg [\mathsf{M}_R]$ (panel A), whether vanilla leptogenesis or resonant leptogenesis [185] takes place depends on the hierarchy of the Yukawa couplings in M_R.

(iii) If $[\mu] \sim [\mathsf{M}_R] \gg m_N$ (panel C or panel D), $M_{N_1} \simeq m_{R_i}(m_N/|\mu_i|)^2$, and vanilla leptogenesis is expected.

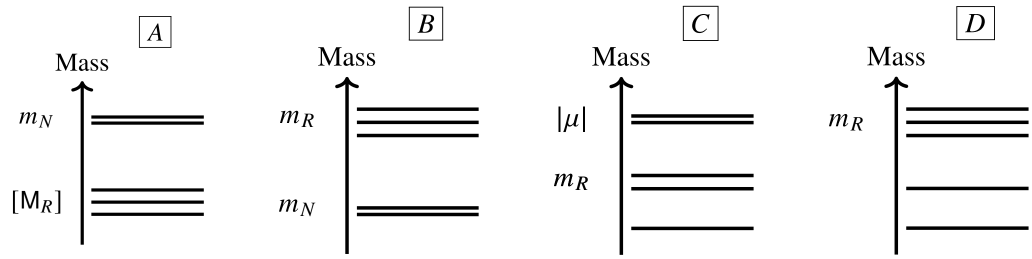


Figure 5.2: Possible heavy neutrino physical mass spectra for different hierarchies. We follow the notation from the main text. Panel A corresponds to the hierarchies $m_N \gg [\mathsf{M}_R] \gg |\mu_i|$ and $m_N \gg |\mu_i| \gg [\mathsf{M}_R]$. In both cases $[\mu]$ does not impact the heavy neutrino masses. Panel B represents the mass spectrum when $[\mathsf{M}_R] \gg m_N \gg |\mu_i|$. Panel C involves the hierarchies $|\mu_i| \gg m_N \gg [\mathsf{M}_R]$, where the lightest heavy neutrino mass is either m_{R_1} or $m_{R_i} m_N^2/|\mu_i|^2$, and $|\mu_i| \gg [\mathsf{M}_R] \gg m_N$, with the lightest heavy neutrino mass always being given by $m_{R_i} m_N^2/|\mu_i|^2$. We note that in the latter case, m_N does not impact the heavy neutrino masses. Panel D corresponds to the hierarchy $[\mathsf{M}_R] \gg |\mu_i| \gg m_N$ (except for very hierarchical M_R, which lead to the spectrum from Panel B instead). Although we assumed three decoupled scales to reproduce the panels above, similar hierarchies can be found relaxing this condition.

(iv) If $m_N \sim [\mathsf{M}_R] \sim [\mu]$, all heavy neutrino masses will be roughly at the same scale, and whether vanilla or resonant leptogenesis occur will depend on the particular values of the Yukawa couplings entering in M_H.

We proceed under the assumption of hierarchical heavy neutrinos, whereby only the lightest heavy neutrino plays a significant role in leptogenesis. This corresponds to scenarios A, C and D in Fig. 5.2. In our numerical analysis in Sec. 5.5, this is enforced by imposing $M_{N_{j\neq 1}} > 5 M_{N_1}$ as a rule of thumb [186].

The CP asymmetry

In the physical basis, the interactions between the heavy neutrinos and the leptonic $\mathrm{SU}(2)_L$ doublets are given by

$$-\mathcal{L} \supset \bar{L}_{Li} \mathsf{h}^{i\ell} \tilde{H} N_\ell + \text{h.c.}, \qquad \text{where} \quad \mathsf{h}^{i\ell} \equiv (\mathsf{Y}_\nu \mathsf{O}_R)^{ik} \mathsf{U}_H^{*k\ell}. \tag{5.30}$$

In our construction of the 5×5 matrix U_H, $k = 2, 3, 4$ correspond to the three flavours of ν_Rs. Meanwhile, the index ℓ runs over the five heavy flavors. In particular, $\ell = 1$ corresponds to the lightest heavy neutrino. The CP asymmetry, generated by the

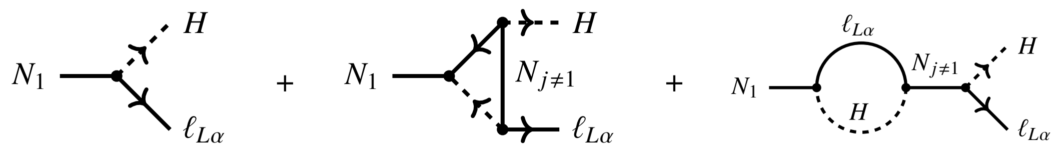


Figure 5.3: Tree-level and one-loop level diagrams for the decays of the lightest heavy neutrino N_1 that can contribute to the CP asymmetry.

interference of the vertex correction and self-energies shown in Fig. 5.3, is defined as (see for example [175, 187])

$$\epsilon_{\rm CP} \equiv \frac{\sum_\alpha \Gamma_{N_1\to H\ell_\alpha} - \Gamma_{N_1\to H^\dagger\bar{\ell}_\alpha}}{\Gamma_D} = \sum_{j\neq 1} \frac{{\rm Im}\{[\mathsf{h}^\dagger \mathsf{h}]_{1j}^2\}}{8\pi[\mathsf{h}^\dagger\mathsf{h}]_{11}} f_{\rm loop}\left[\frac{M_{N_j}^2}{M_{N_1}^2}\right], \tag{5.31}$$

where Γ_D is the total decay rate of N_1, given by[11]

$$\Gamma_D \equiv \sum_\alpha \Gamma_{N_1\to H\ell_\alpha} + \Gamma_{N_1\to H^\dagger\bar{\ell}_\alpha} = \frac{[\mathsf{h}^\dagger\mathsf{h}]_{11}}{8\pi} M_{N_1}. \tag{5.32}$$

For hierarchical heavy neutrinos, $r \equiv M_{N_{j\neq1}}^2/M_{N_1}^2 \gg 1$, the loop function is approximately given by

$$f_{\rm loop}[r] = \sqrt{r}\left[1-(1+r)\log\left(\frac{1+r}{r}\right)+\frac{1}{1-r}\right] \simeq -\frac{3}{2\sqrt{r}}. \tag{5.33}$$

As shown in Ref. [179], for a given M_{N_1} there is a maximum value of $\epsilon_{\rm CP}$ reachable. The proof of this bound, referred to as the Davidson–Ibarra (DI) bound, relies on the assumption that the numbers of light and heavy neutrinos are equal. Although our mass matrix in Eq. (5.27) does not satisfy the latter criterion, a very similar bound can be obtained in our case:

$$|\epsilon_{\rm CP}| \leq \frac{3}{8\pi}\frac{M_{N_1}}{v_H^2} m_{\nu_3}, \tag{5.34}$$

where m_{ν_3} is the heaviest light neutrino. We refer the reader to Appendix 5.C for the proof.

As anticipated earlier, the existence of a minimum required $\epsilon_{\rm CP}$ to generate $\eta_B^{\rm obs}$ combined with the upper bound on the heaviest active neutrino mass, leads Eq. (5.34) to impose a lower bound on the mass of the lightest heavy neutrino.

[11] We are assuming that, in whatever theory that UV-completes our simplified model, the branching ratio ${\rm Br}(N_1 \to H\ell) \sim 1$.

Dynamics

The dynamics of N_1 and the generated $B-L$ asymmetry are dictated by the Boltzmann equations:

$$\frac{dN_{N_1}}{dz} = -(D+S)(N_{N_1} - N_{N_1}^{\rm eq}), \tag{5.35}$$

$$\frac{dN_{\rm B-L}}{dz} = -\epsilon_{\rm CP} D(N_{N_1} - N_{N_1}^{\rm eq}) - W\, N_{\rm B-L}, \tag{5.36}$$

where $z \equiv M_{N_1}/T$, $\epsilon_{\rm CP}$ is defined in Eq. (5.31), N_{N_1} ($N_{\rm B-L}$) is the number density of N_1 (the total $B-L$ asymmetry) normalized by the number density of the photons at $z \longrightarrow 0$. D accounts for the relative strength of the thermal averaged decay and inverse decay rates with respect to the expansion rate of the universe,

$$D(z) = Kz\frac{K_1(z)}{K_2(z)}\,. \tag{5.37}$$

Here, $K_1(z)$ and $K_2(z)$ are the first and second modified Bessel functions, respectively. The *washout parameter* K, defined as

$$K \equiv \frac{\Gamma_D}{H(z=1)} = \frac{\tilde{m}_1}{m_\star}\frac{m_{R_1}}{M_{N_1}}\,, \tag{5.38}$$

quantifies the strength of the interactions responsible for either generating (through inverse decays) or depleting (through decays) the N_1 population. Above, $m_\star \equiv 4\pi v_H^2\sqrt{8\pi^3 g_\star/90}/M_{\rm Pl}$ is a constant (up to variations in the relativistic degrees of freedom $g_\star$ [188]), with $M_{\rm Pl} = 1.2\times 10^{19}$ GeV, and the total decay rate in the vacuum, Γ_D, given in Eq. (5.32). Following the literature, we define the effective neutrino mass $\tilde{m}_1 \equiv [\mathsf{h}^\dagger \mathsf{h}]_{11} v_H^2/(2\,m_{R_1})$ as a proxy for the scale of active neutrino masses. We note that K differs from the usual definition in the literature by the ratio m_{R_1}/M_{N_1} since, in our case, the lightest heavy neutrino mass M_{N_1} does not always correspond to m_{R_1}, as shown in Eq. (5.29).

$\Delta L = 1$ scatterings may affect both the population of heavy neutrinos via S in Eq. (5.35), as well as the washout processes encoded in W in Eq. (5.36). The latter also receives contributions from inverse decays, as well as non-resonant $\Delta L = 2$ scatterings mediated by heavy neutrinos [189].

We will consider the two scenarios where the abundance of N_1 is either dynamically generated (starting from a null thermal abundance at high temperatures) or thermally generated. While the former is more conservative, the latter may occur easily in the context of automatic Nelson-Barr models due to the scattering processes mediated by the new gauge boson, $N_1 N_1 \to (Z')^* \to \bar{f} f$, with f being both new and/or standard model fermions, depending on the force. See appendix 5.A and Refs. [190, 191] for more details.

In our analysis, we make the following assumptions, which simplify considerably the dynamics:

- Flavor effects become relevant at temperatures below 10^{11} GeV [192–194], which covers the regime for automatic Nelson-Barr models, given the upper bound on the SCPV scale coming from demanding quality to the mechanism, see Eq. (5.15). Flavor effects generically help achieve successful leptogenesis, but do not impact the lower bound for M_{N_1} [195–197], which is our main interest. Hence, adopting a conservative approach, we neglect flavor effects and trace over the final lepton flavors, as we did in Eq. (5.36).

- $\Delta L = 1$ scatterings play a relevant role in the weak washout regime ($K < 1$). We find that, in order to be consistent with neutrino data, most of the selected points in our numerical analysis are in the strong washout regime. Therefore, we neglect the effect of $\Delta L = 1$ scatterings in our analysis, although we have checked numerically that they do not impact our conclusions. Including them, the washout term is dominated by the inverse decays, and given by

$$W_{\rm ID} = \frac{1}{2}\frac{N_{N_1}^{\rm eq}(z)}{N_\ell^{\rm eq}} D = \frac{1}{4} K\, z^3 K_1(z)\,, \tag{5.39}$$

  where $N_{N_1}^{\rm eq} = 3z^2 K_2(z)/8$ and $N_\ell^{\rm eq} = 3/4$.

- Non-resonant $\Delta L = 2$ scatterings are neglected, as these are suppressed by two insertions of the Dirac Yukawa couplings.

- Thermal and spectator effects are also neglected. The former can give corrections to the masses and couplings at high temperatures. See Ref. [189] for a detailed study on their impact in leptogenesis. For a thermal population of N_1, the lower bound on its mass stays roughly the same[12]. In the strong washout regime, thermal effects are practically negligible [175]. Spectator processes [198] are interactions in the plasma that can alter the chemical potentials of particles in thermal equilibrium and affect the dynamics of

[12] In the scenario of a dominant population of N_1, the bound could be weakened by an order of magnitude [189].

leptogenesis. Their proper inclusion in the dynamics have at most 20% − 40% effects [199].

Eqs. (5.35) and (5.36) clearly show the three ingredients needed for leptogenesis: (i) a non-zero CP violation, see Eq. (5.31), (ii) lepton number violation induced by the $\Delta L = 1$ decay (and scatterings) of the lightest heavy Majorana neutrino N_1, and (iii) departure of thermal equilibrium, $N_{N_1} \neq N_{N_1}^{\rm eq}$, in order to achieve a non-zero leptonic asymmetry. Sphaleron processes will then transfer the asymmetry generated in the lepton sector to the baryon sector, respecting $B - L$. The latter conversion depends on the $\mathrm{SU}(2)_L$-charged degrees of freedom (the same in our model as in the SM), which gives $a_{\rm sph} = 28/79$ [200]. The baryon asymmetry today is then given by

$$\eta_B = a_{\rm sph} \left(\frac{3}{4} \epsilon_{\rm CP}\, \kappa_f \right) \frac{n_\gamma(T_i)}{n_\gamma(T_{\rm rec})}, \tag{5.40}$$

where $N_{\rm B-L}(T_i)$ is displayed in between brackets, T_i was introduced above as the starting temperature for leptogenesis, and $n_\gamma(T_{\rm rec})$ ($n_\gamma(T_i)$) is the number density of photons evaluated at the recombination temperature $T_{\rm rec}$ (at T_i).

The efficiency factor is defined as

$$\kappa_f = -\frac{4}{3} \int_{z_{\rm RH}}^{\infty} dz' \frac{dN_{N_1}}{dz'} e^{\int_{z'}^{\infty} dz'' W_{ID}(K, z'')}, \tag{5.41}$$

where $z_{\rm RH}$ ideally corresponds to the first temperature of the universe, i.e. $z_{\rm RH} = M_{N_1}/T_{\rm RH}$. For the efficiency factor in the two scenarios studied, we use the analytical approximations from Ref. [176]

$$\kappa_f(K) = \begin{cases} \dfrac{2}{z_B(K) K} \left(1 - e^{-\frac{1}{2} z_B(K) K} \right) & \text{if thermal,} \\ \kappa_f^+(K) + \kappa_f^-(K) & \text{if dynamical,} \end{cases} \tag{5.42}$$

where $\kappa_f^\pm$ are given by

$$\kappa_f^-(K) = 2 e^{-\frac{3}{8}\pi K} \left(1 - e^{\frac{2}{3}\bar{N}(K)} \right), \quad \text{and} \quad \kappa_f^+(K) = \frac{2}{z_B(K) K} \left(1 - e^{-\frac{2}{3} z_B(K) K \bar{N}(K)} \right),$$

with

$$\bar{N}(K) = \frac{9\pi K}{16} \left(1 + \sqrt{\frac{9\pi K/16}{N_{N_1}^{\rm eq}(z=0)}} \right)^{-2}, \tag{5.43}$$

$$z_B(K) \simeq 1 + \frac{1}{2} \ln \left[1 + \frac{\pi K^2}{1024} \left(\ln \left[\frac{5^5 \pi K^2}{1024} \right] \right)^5 \right]. \tag{5.44}$$

When does leptogenesis occur? This is a key question regarding the viability of leptogenesis within Nelson-Barr models. Following Ref. [176], we estimate the lowest possible initial temperature for leptogenesis as

$$T_i \equiv \frac{M_{N_1}}{z_i} \simeq \frac{M_{N_1}}{z_B - 2e^{-3/K}} \,. \tag{5.45}$$

Inflation, or whatever mechanism that solves the horizon problem, must take place after SCPV in order to erase the domain walls. To ensure this, we require the reheating temperature to be lower than $v_{\rm CP}$. We can then interpret T_i as a lower bound on the first temperature of the universe. In the weak washout regime ($K \lesssim 1$), $T_i \sim M_{N_1}$, while in the strong washout regime T_i is up to one order of magnitude smaller than M_{N_1}. The former is understood from the small K expansion of $z_B = 1 + \mathcal{O}(K^2)$. The latter comes from the upper bound on $K \lesssim 10^3$, which suppresses T_i by a factor of ~ 10 with respect to M_{N_1}.

5.5 Methods and Results

In this section, we perform a numerical study of leptogenesis for the simplified model parametrized in Eq. (5.24), which reproduces the features of the Nelson-Barr mechanism in the lepton sector (see discussion in Sec. 5.4).

Sampling

To efficiently explore the parameter space, we first allow the two heavy vevs, $v_{\rm S}$ and $v_{\rm CP}$, to vary logarithmically within the interval $[10^7, 10^{14}]$ GeV. These boundaries are justified as follows: numerical scans show that there are no viable parameter points for the heavy vevs below 10^7 GeV. On the upper end, $v_{\rm CP}$ is constrained by the quality bound, while $v_{\rm S}$ is limited by the canonical seesaw bound. We then pick specific mass matrices by randomly sampling the 13 real Yukawa couplings $(\mathsf{Y}_\nu \mathsf{O}_R)^{ij}$, $(\mathsf{O}_R^T \mathsf{Y}_R \mathsf{O}_R)^{ii}$, and λ_N on a uniform logarithmic scale. We similarly sample the magnitude of the 3 complex parameters $|\mu^i|/v_{\rm CP}$, and we uniformly sample their phases in the interval $[0, 2\pi)$. For each point in a 200×200 grid $(v_{\rm CP}, v_{\rm S})$ such that $v_{\rm CP} < v_{\rm S}$, we randomly sampled the 19 parameters described above 10^3 times.

We fix the overall scale of $(\mathsf{Y}_\nu \mathsf{O}_R)$ in order to obtain a mass for the heaviest active neutrino that saturates the bound $m_{\nu_3} \leq 0.1$ eV. Invoking perturbativity, we bound the magnitude of all Yukawa couplings to be less than one[13], allowing them to vary

[13] We have chosen 1 as an upper bound for the Yukawa couplings (instead of $\sqrt{4\pi}$) to stay safe from RGE effects. Although we do not expect relevant contributions from the new sector to the running as these are neutral fields (under the SM forces), new charged fermions may be present in specific UV completions (see appendix 5.A). We thank Riccardo Rattazzi for pointing this out to us.

along the span of 1, 2, and 5 orders of magnitude. In the plots, we indicate the different spreads in the Yukawa couplings with different colors: blue, orange, and green points are log-uniformly sampled from the intervals $[10^{-5}, 1)$, $[10^{-2}, 1)$, and $[10^{-1}, 1)$ respectively.

Once all the mass matrices are generated, we numerically diagonalize *à la Takagi* to find the masses in the physical basis and we analytically compute - using the formulae listed in Sec. 5.4 - the unflavored CP asymmetry $\epsilon_{\rm CP}$ and the washout parameter K for each one of our specific realizations, inferring the predicted η_B using Eq. (5.40).

Successful Leptogenesis

We consider successful samplings the ones that meet the following three conditions:

- They must generate enough baryon asymmetry to match the observed value $\eta_B^{\rm obs}$ given in Eq. (5.1). Thus, we require $\eta_B^{\rm obs} < |\eta_B| < 1.5 \times \eta_B^{\rm obs}$.
- The initial temperature for leptogenesis T_i, computed according to Eq. (5.45), must be lower than $v_{\rm CP}$ to avoid the restoration of the CP symmetry. Hence, we impose $T_i < v_{\rm CP}$.
- The mass spectrum of the heavy neutrinos must be hierarchical so that only the asymmetry produced by the N_1s population matters. We conventionally keep only the points with $M_{N_{j\neq 1}} > 5 M_{N_1}$ as a rule of thumb [196].

We work within two possible scenarios regarding the initial population of N_1s:

(i) TIA: thermal initial abundance (left panels), where the lightest heavy neutrino N_1 is taken to be in thermal equilibrium prior to leptogenesis, i.e. $N_{N_1}(0) = 3/4$.

(ii) DIA: dynamical initial abundance (right panels), with $N_{N_1}(0) = 0$, i.e. assuming that the N_1s are produced exclusively by inverse decays.

For the two scenarios above, the efficiency factors are computed according to Eq. (5.42).

In Fig. 5.4, we plot all the successful points in our scan in order to show the parameter space consistent with leptogenesis from different perspectives.

In the first row of Fig. 5.4, $|\epsilon_{\rm CP}|$ is plotted against K. The efficiency curve for both scenarios is easily recognized. This is a consequence of the efficiency factor, κ_f,

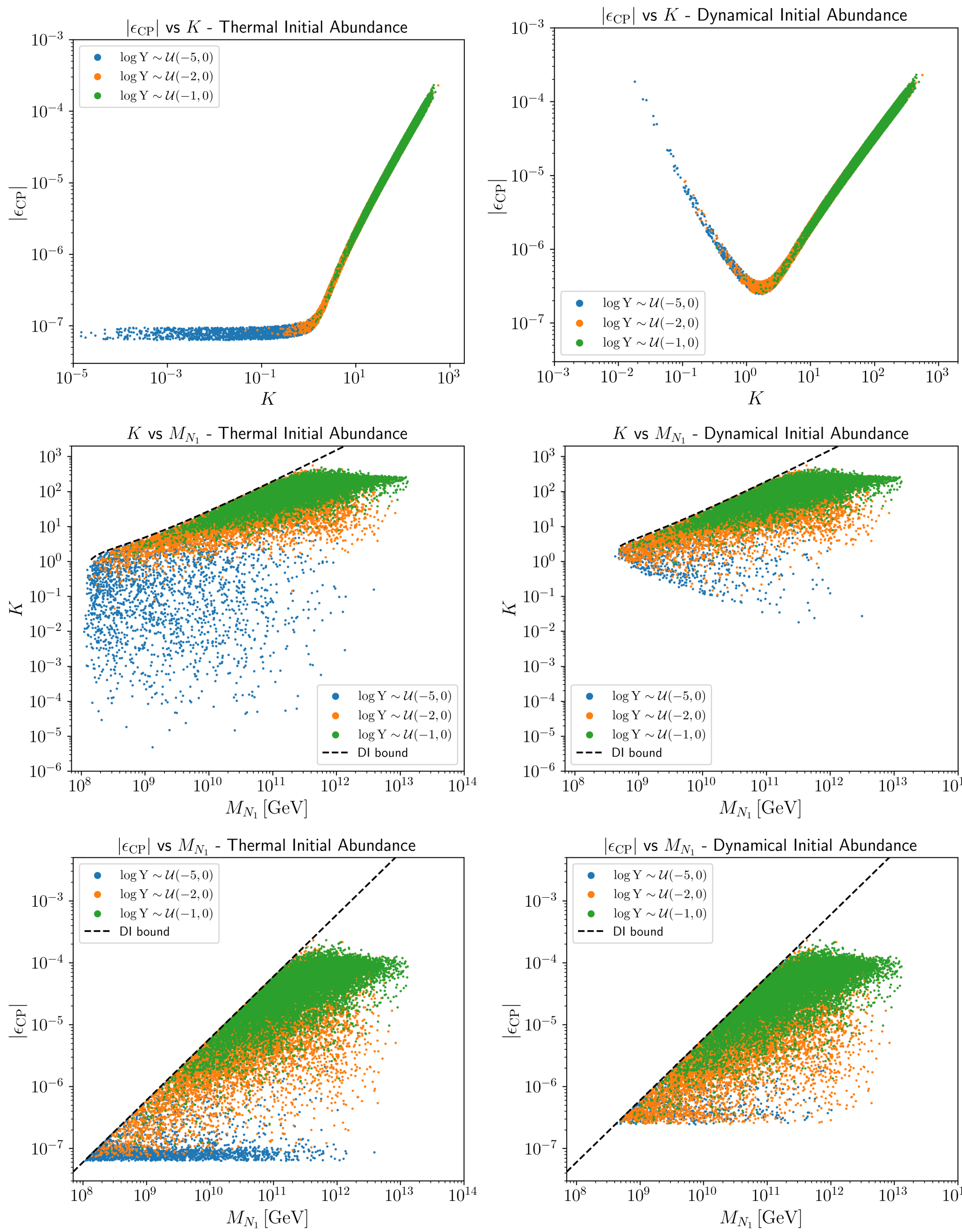


Figure 5.4: Scatter plots showing the parameter space that lead to successful leptogenesis, obtained by randomly sampling the magnitude of the Yukawa couplings, Y, from log-uniform distributions with varying ranges. Specifically, the log values of Y are drawn from a uniform distribution $\log \mathrm{Y} \sim \mathcal{U}(n, 0)$, where $n = -1, -2$, or -5 sets the lower bound of the distribution. This corresponds to logarithmically sampling Y over one (green points), two (orange points), or five (blue points) orders of magnitude below the upper bound, which is fixed at $\mathrm{Y} = 1$ (see discussion in footnote 13). The points satisfy $\eta_B^{\rm obs} < |\eta_B| < 1.5\,\eta_B^{\rm obs}$, $v_{\rm CP} > T_i$, and $M_{N_{j\neq 1}} > 5M_{N_1}$.

scaling inversely proportional to $\epsilon_{\rm CP}$ for a fixed baryon-to-photon ratio, as shown in Eq. (5.40). The thickness of the band is given by our selection rule for the $\eta_{\rm B}$ of the successful points. The reader may immediately identify the lower bound on $|\epsilon_{\rm CP}|$, which corresponds to the maximum efficiency achievable: $K \lesssim 1$ ($K \sim 1$) for the TIA (DIA) case. Note that, as expected [176, 201], the strong washout regime, $K > 1$, is preferred if Yukawa couplings are $\mathcal{O}(1)$ (green points). This is because $\tilde{m}_1 \sim m_{\nu_3}$[14]. Wider ranges of variation (on a logarithmic scale) of the Yukawa couplings allow for a larger dissociation between $\tilde{m}_1$ and m_{ν_3}, which in turn leads to a broader range of K values (see the orange and blue points). Since in the strong washout regime the efficiency curves for both TIA and DIA are the same, the two plots for $K \gtrsim 10$ are perfectly coincident. In the DIA plot, we can identify an upper bound for $\epsilon_{\rm CP}$ for each washout regime. In the weak washout regime, the curve stops because the CP asymmetry is saturated[15]. In the strong washout regime, the common upper bound for both TIA and DIA comes from K reaching its maximum value. As we discuss in Appendix 5.C, the upper bound $K < m_{\nu_3}/m_\star$ [201] derived in the vanilla leptogenesis scenario (3 heavy neutrinos) does not hold in our case – although we do not expect a significant deviation. Indeed, our results include points with $10^2 < K < 10^3$. The plots in the second row show an alternative correlation in the parameter space of the theory, this time K vs M_{N_1}. For a given K, there is a certain $\epsilon_{\rm CP}$ that gives the right asymmetry, which in turn fixes the lightest M_{N_1} allowed by saturating the DI bound (see Eq. (5.34), represented with a dashed line in these figures). We note that, as M_{N_1} grows, the washout parameter tends to $K \sim 10^2$ in both TIA and DIA. This behavior reflects the seesaw mechanism: as M_{N_1} approaches the canonical seesaw limit, the spread in Yukawa coupling narrows, leading to $\tilde{m}_1 \sim m_{\nu_3}$. Conversely, for smaller values of M_{N_1}, the spread widens (for a fixed range, i.e., a given color), resulting in a stronger decoupling between $\tilde{m}_1$ and m_{ν_3}, and thus a broader variation in K.

The third-row plots explicitly depict the DI bound. The upper bound on $\epsilon_{\rm CP}$ for both regimes has already been discussed. The lower bound on $\epsilon_{\rm CP}$ corresponds to the maximum value of the efficiency factor κ_f. The blue points are statistically favored where the efficiency is maximized, that is, $K \lesssim 1$ ($K \sim 1$) for TIA (DIA). This can be clearly seen from the accumulation of points at the minimum $|\epsilon_{\rm CP}|$ allowed.

[14] For $[\lambda_N]$, $[\mathsf{O}_R^T \mathsf{Y}_R \mathsf{O}_R]$, $[\mu] = \mathcal{O}(1)$, $m_{R_1} \sim m_N \sim \mu$ and so, according to Eq. (5.29), $m_{R_1} \sim M_{N_1}$. In this case, the hierarchy between $\tilde{m}_1$ and $m_\star$ determines whether we are in the strong ($\tilde{m}_1 > m_\star$) or weak ($\tilde{m}_1 < m_\star$) wash-out regimes.

[15] We have checked that with more statistics – for $(\mathsf{O}_R^T \mathsf{Y}_R \mathsf{O}_R)^{ii}$, λ_N, $[\mu]/v_{\rm CP} \sim \mathcal{O}(1)$ and the Yukawa couplings relevant for the light neutrinos varying along the span of 5 orders of magnitude – $\epsilon_{\rm CP} > 10^{-4}$ can be achieved for $K \ll 1$, approaching $\epsilon_{\rm CP} \sim 10^{-2}$ (loop effect).

The lowest value of M_{N_1} leading to successful leptogenesis is easily identified: $M_{N_1} \gtrsim 10^8$ GeV ($M_{N_1} \gtrsim 4 \times 10^8$ GeV) for the TIA (DIA) scenario. Such a lower bound sets the lowest possible reheating temperature via Eq. (5.45).

A comment on the mass scale of the new fermions regarding finite naturalness is in order. As we discussed in subsection 5.3, finite naturalness constraints on the radiatively generated quartic coupling require the new quarks to lie near the TeV scale [144]. However, this bound is significantly relaxed in the case of the new neutral leptons. As shown in the one-loop diagram (square brackets) in Fig.5.1, their contribution is suppressed by two insertions of the Dirac Yukawa coupling and by the hierarchy between the mass of the neutral fermions and $v_{\rm CP}$ (width of the viable window for leptogenesis). Using the analog of Eq. (5.19) for the lepton, $\delta\lambda_4|_{1-\rm loop} v_{\rm CP}^2 \lesssim m_H^2$ reads as

$$\frac{1}{16\pi^2}\frac{2m_{\nu_3}}{v_H^2}M_{N_1}^3 \ln\left(\frac{M_{N_1}^2}{m_h^2}\right) \lesssim m_H^2 \;\Rightarrow\; M_{N_1} \lesssim 0.4\times 10^7, \tag{5.46}$$

which agrees with the order of magnitude of the constraint derived in Ref. [32] by demanding finite naturalness in Type-I seesaw models.

Quality bounds

We are now in a position to address whether leptogenesis is viable within automatic Nelson-Barr models. Each point in our sampled parameter space has an associated temperature at which leptogenesis begins, see Eq. (5.45), and a given $\Delta\bar{\theta}_{\rm QCD}$. In this subsection, we apply the quality bounds discussed in Sec. 5.3 to the successful points in order to determine the width of the viable window in energy scale.

In Fig. 5.5, we plot the ratio between the scale of SCPV, $v_{\rm CP}$, and the lowest temperature at which leptogenesis becomes successful, T_i. We take this dimensionless quantity as a proxy for a the width of the viable energy window for leptogenesis.

In the first row, the plots display the successful points from the parameter scan used in our numerical analysis. As indicated by the different colors, a wider spread in Yukawa couplings allows for a broader window. The points forming the sharp boundary on the upper right of the plots maximize the ratio $v_{\rm CP}/T_i$, achieved when $v_{\rm CP} \sim 10^{14}$ GeV and $T_i \sim 0.1\, M_{N_1}$ (strong washout regime). This cutoff is a consequence of our imposed upper limit on $v_{\rm CP}$ in the scan. However, as we will see in the plots below, this cutoff is purely chosen for convenience since the points consistent with the mechanism's quality remain unaffected by it.

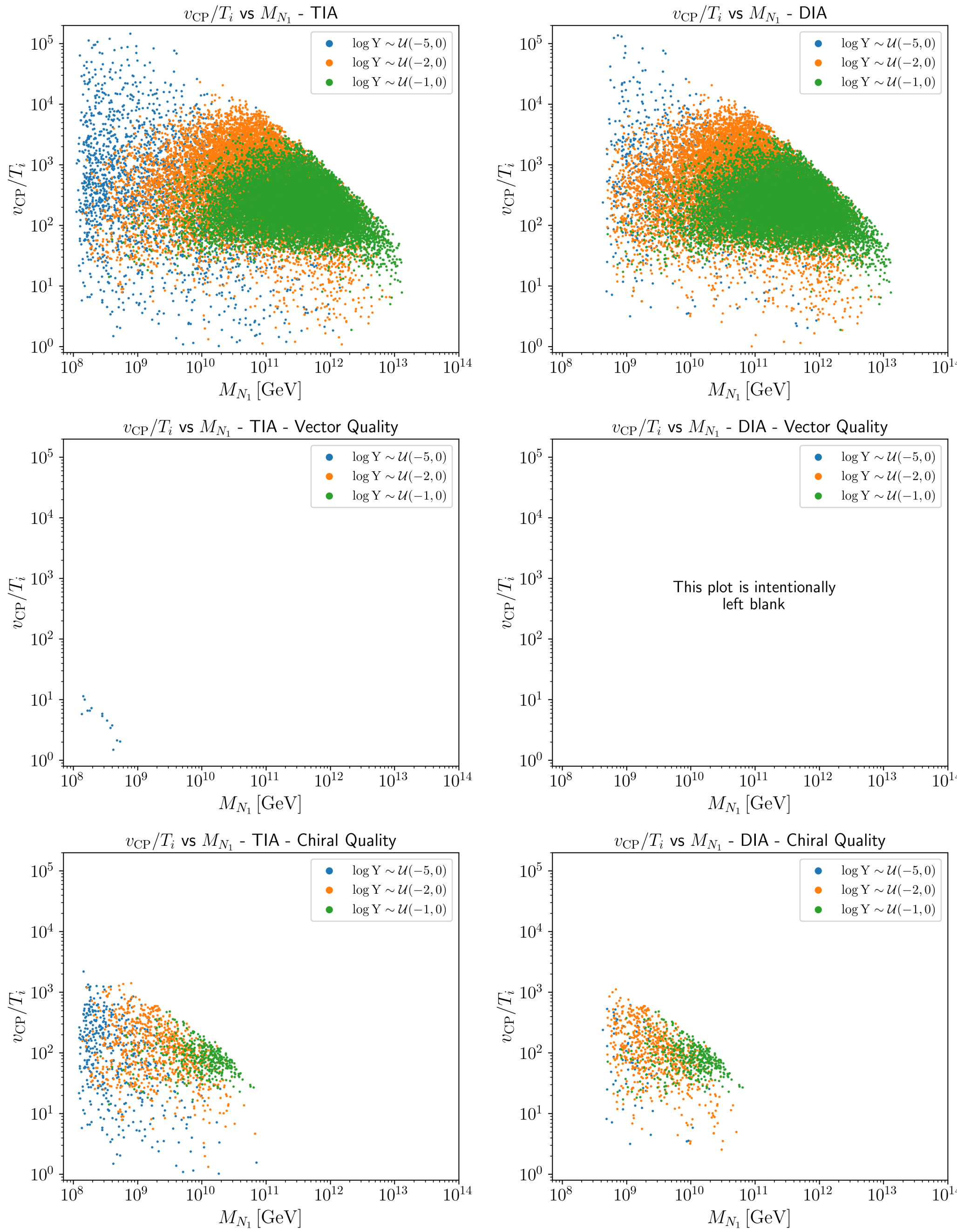


Figure 5.5: Scatter plots of the successful points for leptogenesis as a function of M_{N_1} and the ratio $v_{\rm CP}/T_i$ (required to be greater than one by our selection procedure), for a TIA (left) and DIA (right) of N_1s. In the first row, we show all successful points from our numerical analysis. In the second and third row, we show which of these succesful points are compatible with the quality bounds for automatic Nelson-Barr models with new vector-like and chiral quarks, respectively. Different colors of the points indicated different spreads taken for the Yukawa couplings, according to the legends.

Green points (for which Yukawa couplings are order one) favor the strong washout regime and therefore $T_i \sim 0.1\, M_{N_1}$, and $v_{\rm CP} \lesssim v_S$ (see subsection 5.4). These points cluster around $v_{\rm CP}/T_i \sim 10^2$, bounded by two horizontal cuts. The upper cut is given by the maximum ratio $v_{\rm CP}/T_i \leq v_S/(0.1 M_{N_1})$ allowed. The minimum M_{N_1} accessible within an order of magnitude spread in the Yukawa couplings is $M_{N_1} \sim 10^{-3}\, v_S$ (see panel C in Fig. 5.2). Conversely, the emergent lower cut arises when minimizing the ratio. The largest M_{N_1} possible for our numerics is $v_S/5$, due to the imposed hierarchy $M_{N_1} < M_{N_{j\neq 1}}/5$, which yields a minimum ratio of $v_{\rm CP}/T_i \sim 50$. These horizontal cuts relax for the orange and blue points, with the blue ones spanning all five orders of magnitude, as expected. These are statistically favored in the range of highest efficiency, i.e., $K \lesssim 1$ ($K \sim 1$) for TIA (DIA). For most of these points $T_i \sim M_{N_1}$, and the spread in the plots reflects the spread in the magnitude of the Yukawa couplings.

In the second row, we show the successful points for leptogenesis that are compatible with the quality bound, $\Delta\bar{\theta}_{\rm QCD} < 10^{-10}$, for automatic Nelson-Barr models with new vector-like quarks. See Sec. 5.3, Eq. (5.12). The absence of points shows that this scenario is incompatible with high-scale leptogenesis starting from a DIA; and only a few points survive for the TIA case. We notice that relaxing condition (iv) in subsection 5.4, i.e. considering vector-like neutral leptons, includes a third scale m_N (uncorrelated with $v_{\rm CP}$ and v_S). We explicitly verified by more simulations that the results of the scan do not change the available parameter space shown in Fig. 5.5. The statement that high-scale leptogenesis requires chiral new fermions, already anticipated in Ref. [143], is strongly based on the assumptions we made. We note some of them could be relaxed (e.g. allowing for degenerate neutrinos enabling resonant leptogenesis, or including flavor effects). However, we also note that the bound in Eq. (5.12) assumes a particularly optimistic framework. For instance, when naturalness considerations are included and the new quarks are assumed to reside near the TeV scale, the quality bound becomes more stringent, driving $v_{\rm CP}$ down to the TeV range, and thereby creating clear conflict with high-scale leptogenesis.

Plots from the third row show the successful points for leptogenesis consistent with quality bounds for new chiral fermions (vector-like under the SM). See Sec. 5.3, Eq. (5.15). As illustrated by the points, a broader spread in the Yukawa couplings leads to smaller values of M_{N_1} and a wider viable window.

We conclude that automatic Nelson-Barr models with new chiral fermions (vector-like under the SM group) are compatible with high-scale leptogenesis and with the simplest mechanisms for diluting unobserved topological defects. In the next section, we focus on this type of automatic NB models.

5.6 Potential signatures

High-scale leptogenesis is notoriously inaccessible to experimental verification. However, in the context of automatic Nelson-Barr models, the quark and lepton sectors become correlated due to their shared origin of the CP-violating phase. In this framework, the scale of spontaneous CP violation, $v_{\rm CP}$, provides a connection between leptogenesis and $\bar{\theta}_{\rm QCD}$.

In the previous section we showed that automatic NB models with new chiral fermions are compatible with high scale leptogenesis. As extensively discussed in Sec. 5.3, higher-dimensional operators can induce a shift in $\bar{\theta}_{\rm QCD}$, given by Eq. (5.14). Assuming that higher-dimensional operators are suppressed by negative powers of the Planck scale, and both the CP phase and the magnitude of the Wilson coefficients, $[\xi]$, to be order one, Eq. (5.14) becomes

$$\Delta\bar{\theta}_{\rm QCD} \simeq 3\times 10^{-18}\left(\frac{v_{\rm CP}}{10^8\ {\rm GeV}}\right)^2\left(\frac{3\times 10^{-5}}{[\mathsf{Y}_d]}\right)\left(1+\frac{v_S}{v_{\rm CP}}\right). \tag{5.47}$$

In the above, we picked the new quark to be down-like, $[\mathsf{Y}_d] = 3\times 10^{-5}$, but a similar result is expected for up-type quarks. We also have assumed that $m_D/v_{\rm CP} \gg [\mathsf{Y}_d]$. Relaxing this assumption would shift the points towards more testable shifts in $\bar{\theta}_{\rm QCD}$.

In Fig. 5.6, we show the scatter plots of the points which achieve successful leptogenesis, according to our previous three criteria (see subsection 5.5). On the y-axis we compute the shift $\Delta\bar{\theta}_{\rm QCD}$ according to Eq. (5.47). The red shaded area is excluded by the current leading upper bound on the nEDM, $|d_n| < 2.2\times 10^{-26}\ e\,$cm [124], measured at PSI. Such a bound translates into an upper bound on $\bar{\theta}_{\rm QCD}$ via [207, 208]

$$|d_N| \sim |\bar{\theta}_{\rm QCD}|\frac{m_q^*}{\Lambda_{\rm QCD}}\frac{e}{2m_N} \sim |\bar{\theta}_{\rm QCD}|\times 10^{-16}\ e\,{\rm cm}, \tag{5.48}$$

where $m_q^* = m_u m_d m_s/(m_u m_d + m_s m_u + m_s m_d)$ is the reduced quark mass, $m_N \sim 1$ GeV is the mass of the nucleon, and $\Lambda_{\rm QCD} \sim 200$ MeV. We also show, in gray, potential constraints coming from future nucleon EDM measurements. Experiments of ultra-cold neutrons conducted directly within superfluid helium allow higher densities of neutrons and electric fields, enabling improved sensitivities [203]. For example, the n2EDM experiment, which is currently under construction, is

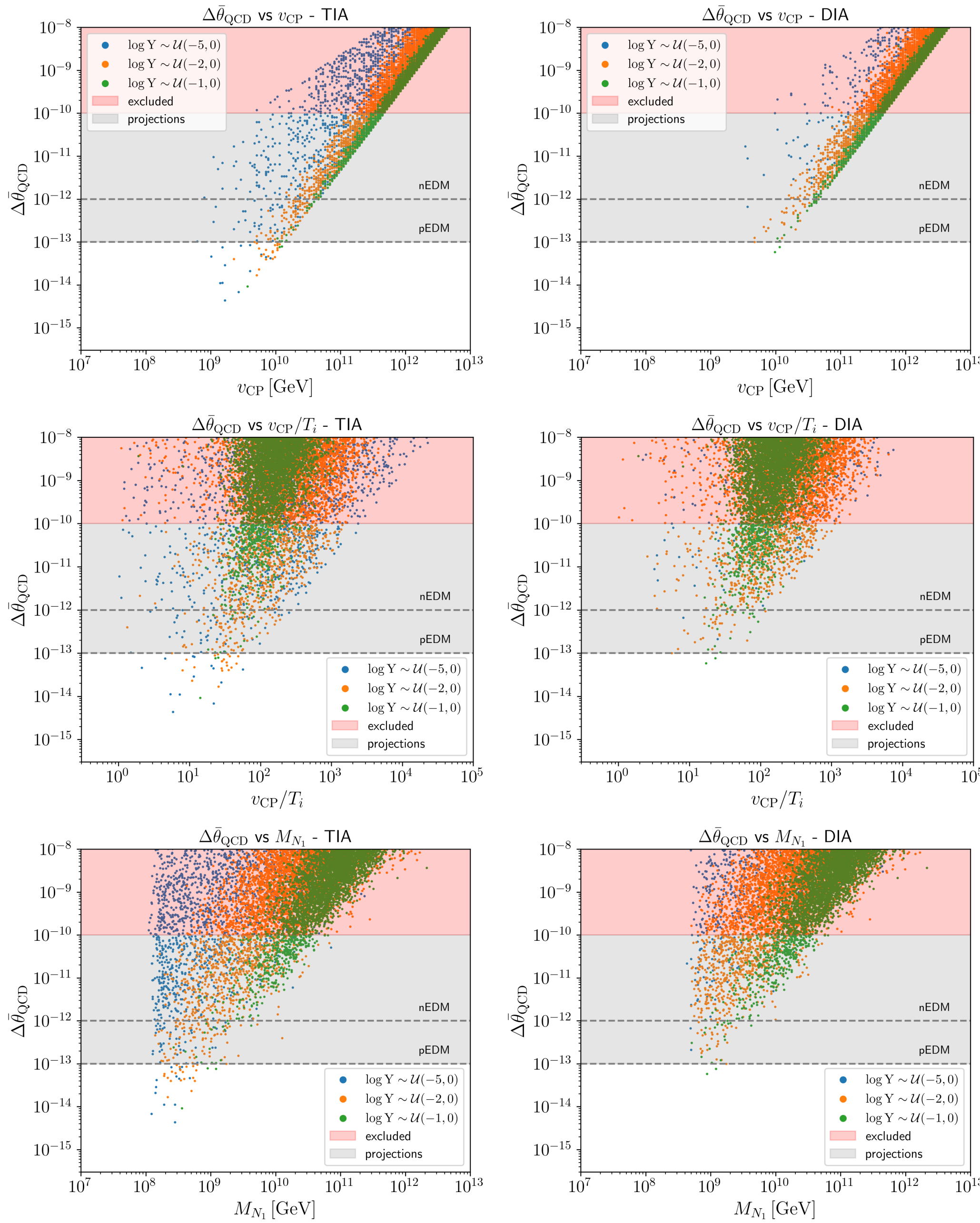


Figure 5.6: Scatter plots of the successful leptogenesis points and their induced shift in $\bar{\theta}_{QCD}$ computed with Eq. (5.47) as a function of $v_{\rm CP}$ (first row), $v_{\rm CP}/T_i$ (second row), and M_{N_1} (third row), for a TIA (left) and DIA (right) of N_1s. In red, the zone currently excluded from electric dipole moment experiments. In gray, delimitated by dashed lines, we show the projective bounds from future experiments on the nEDM [202–205], and the proton electric dipole moment proton electric dipole moment (p-EDM) [206].

designed to measure the nEDM with sensitivity $|d_n| < 1 \times 10^{-27}$ e cm, with further possibility to go below the 10^{-27} e cm regime (e.g. $|d_n| < 5 \times 10^{-28}$ e cm in n2EDM Magic) [202]. Other collaborations [204, 205] aim for similar sensitivities. These experiments will allow to probe $|\bar{\theta}_{\rm QCD}| > 10^{-12}$, shown on the plots with a dashed line labeled by nEDM.

Alternatively, projections on the proton EDM (pEDM) are very promising. The pEDM experiment [209, 210] proposes to search for a pEDM by measuring the vertical rotation of the polarization of a stored proton beam in a storage ring. Phase-1 aims for $|d_p| < 10^{-29}$ e cm, which improves the sensitivity to $\bar{\theta}_{\rm QCD}$ by three orders of magnitude. We show this projection on the plots with a dashed line labeled by pEDM. Phase-2 aims to be sensitive to the SM prediction of 10^{-31} e cm.

In the first row of Fig. 5.6, v_{CP} is shown on the x-axis. As is particularly evident in the TIA scenario, the successful points lie within two diagonal bands with different slopes. The lower band, where most of the green points are located, corresponds to the regime where $v_S \sim v_{\rm CP}$, leading to $\Delta\bar{\theta}_{\rm QCD} \propto v_{\rm CP}^2$. The upper band, reached only by the blue points, arises when a larger spread in the Yukawa couplings allows for $v_{\rm CP} \ll v_S$ while still maintaining a sufficiently large CP-violating phase. In this case, $\Delta\bar{\theta}_{\rm QCD} \propto v_{\rm CP}$.

In the second row, the shift in $\bar{\theta}_{\rm QCD}$ is plotted against the available window for leptogenesis, parametrized by $v_{\rm CP}/T_i$. As already shown in Fig. 5.5, smaller predictions for $\Delta\bar{\theta}_{\rm QCD}$ correspond to narrower windows in which leptogenesis can successfully occur.

In the third row, $\Delta\bar{\theta}_{\rm QCD}$ is plotted against the mass of the lightest heavy neutrino, M_{N_1}. There is an upper bound on M_{N_1} consistent with maintaining the quality of the mechanism, approximately $M_{N_1}^{\rm max} \sim 8 \times 10^{10}$ GeV. Lower neutrino masses lead to a smaller shift in $\bar{\theta}_{\rm QCD}$. This trend reflects a general feature: improved experimental sensitivity to $\Delta\theta_{\rm QCD}$ favors both a lower scale for the new heavy neutral fermions and lower scale for SCPV, as shown in the first-row plots.

In both TIA and DIA scenarios, we can appreciate that the vast majority of the successful points are within reach of future EDM experiments [203].

5.7 Conclusions

In this chapter, we have numerically shown that the observed baryon asymmetry can be reproduced in automatic Nelson-Barr models through leptogenesis while maintaining a high-quality solution to the strong CP problem. In these models, where all sources of CP violation are tied to a single symmetry-breaking scale, v_{CP}, CP violation must account for the nonzero phase in the quark sector (Jarlskog determinant), but also for the observed baryon asymmetry. We consequently studied the resulting correlations among the observed CP-violating phenomena.

In automatic NB models, the required quark mass texture to achieve a vanishing tree-level $\bar{\theta}_{\mathrm{QCD}}$ is a consequence of a gauge symmetry (in our scenario, an abelian force). This framework allows for a higher spontaneous CP violation (SCPV) scale facilitating the implementation of mechanisms to dilute the domain walls associated with its breaking. On the model building side, we mirrored the quark sector by introducing five heavy neutral leptons, following the same structure in the couplings to the SCPV scalar fields. In addition, we let the three right-handed neutrinos to get a Majorana mass term from the vev of the field that breaks the extra gauge symmetry, v_{S}, keeping the number of relevant scales and phases to a minimum. To be maximally conservative, we adopted intentionally pessimistic assumptions to test NB leptogenesis: we modeled leptogenesis dynamics using only decays and inverse decays, leaving out other processes (such as scatterings, flavor, thermal effects,. . .), that usually enhance the generated asymmetry. We further imposed a hierarchical spectrum for the heavy leptons so that the Davidson–Ibarra bound holds, ruling out resonant enhancements.

A numerical scan over the two scales at play, v_{CP} and v_{S}, and a randomized sampling of Yukawa couplings over several orders of magnitude manifest a robust window of two to three orders of magnitude between v_{CP} and the leptogenesis temperature T_i, assuming the new quarks are chiral under the new gauge group. This gap is wide enough to accommodate any mechanism needed to dilute SCPV-induced domain walls and still reheats the universe above T_i. For vector-like quarks, the quality bound tightens and high-scale leptogenesis faces clear tension – though our conservative assumptions suggest that more complex models or a fuller treatment of leptogenesis dynamics could reopen viable parameter space.

High-scale leptogenesis is notoriously difficult to experimentally constrain. However, NB models implementing it quantitatively correlate – due to to the common SCPV scale – the induced shift in the QCD vacuum angle, $\Delta\bar{\theta}_{\mathrm{QCD}}$, coming from higher-dimensional operators, with the baryon asymmetry predicted, η_B. From our numerical analysis, a clear correlation emerges between $\Delta\bar{\theta}_{\mathrm{QCD}}$ and the available leptogenesis window v_{CP}/T_i; improved sensitivity to $\bar{\theta}_{\mathrm{QCD}}$ corresponds to narrower windows, thereby constraining high-scale leptogenesis. Another interesting insight we can deduce from our numerics is that enforcing the NB quality bound caps the lightest heavy–neutrino mass at $M_{N_1} \lesssim 8 \times 10^{10}$ GeV. Remarkably, the vast majority of the parameter space available induces a shift in $\bar{\theta}_{\rm QCD}$ that lies within the reach of current and future nucleon EDM experiments and well above the SM prediction of $\bar{\theta}_{\rm QCD} \sim 10^{-15}$. Although we cannot exclude Nelson-Barr models if such a shift is not observed, further tuning may be needed to naturally accommodate high-scale leptogenesis while still having a high-quality solution to the strong-CP problem.

So far, we have shown that automatic NB models solve a ten order of magnitude discrepancy between the observed upper bound on $\bar{\theta}_{\rm QCD} < 10^{-10}$ and the naive theoretical expectation for it to be $O(1)$ via a high-quality mechanism that is robust under UV corrections and, under certain assumptions, compatible with high-scale leptogenesis. However, the introduction of new heavy fermions generically adds dangerous radiative corrections to the effective scalar potential, and a similar fine tuning to the one that NB models succeed to solve may arise between different contributions to a physical quantity. Demanding finite naturalness, the lowest mass solutions found $M_{N_1} \sim 10^8$ GeV, require three-four orders of magnitude in fine tuning between the squares of the bare mass of the Higgs and its radiative contributions, manifesting the usual tension with high-scale leptogenesis. A possibility to relax the tuning in this case is to consider enhancements of the produced asymmetry in order to evade the Davidson-Ibarra bound (e.g. population dynamics, resonant spectrum, ARS leptogenesis. . .). A second way would be to explore a low scale baryogenesis scenario in which the out-of-equilibrium decays of the complex fields responsible for SCPV produce an asymmetry in the primordial universe. However, quick thermalization with the SM bath due to the hypercharges as well as mixings between all neutral leptons strongly challenge the mechanism. We leave studies in these directions for future work.

In summary, automatic Nelson-Barr models with chiral fermions offer a high-quality solution to the strong CP problem and can naturally accommodate high-scale leptogenesis, successfully accounting for the observed matter-antimatter asymmetry.

Appendix 5.A Automatic Nelson-Barr UV completions

As discussed in the main text, we aim to gauge the $\mathbb{Z}_2$ (or any discrete symmetry imposing the zeroes in the mass matrix) usually introduced in simplified NB models such as the minimal theory proposed by BBP [159]. For simplicity, we assume that it is a consequence of an abelian symmetry, $\mathrm{U}(1)$.

The simplest attempt would be to assume that the new vector-like down (or up) quarks are vector-like under the entire symmetry group $\mathrm{SU}(3) \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y \otimes \mathrm{U}(1)$:

$$D_L \sim (3, 1, -1/3, Q_D), \qquad \text{and} \qquad D_R \sim (3, 1, -1/3, Q_D),$$

and that the complex scalars responsible for SCPV have quantum numbers $X_a \sim (1, 1, 0, Q_D)$. Although the $\mathbb{Z}_2$ symmetry has been gauged, the vector-like nature of the new fermions leads to contributions from dimension five operators to $\Delta\theta_{\mathrm{QCD}}$, with a quality bound that is in tension with thermal leptogenesis (see Eq. 5.12 and the discussion in subsection 5.5, around Fig. 5.5).

In order to make the new quarks chiral under the new symmetry (still vector-like under the SM), we assume they carry different $\mathrm{U}(1)$ charges. In this case, however, the triangle anomalies $\mathcal{A}[\mathrm{SU}(3)^2 \otimes \mathrm{U}(1)]$, $\mathcal{A}[\mathrm{U}(1)_Y \otimes \mathrm{U}(1)^2]$, $\mathcal{A}[\mathrm{U}(1)_Y^2 \otimes \mathrm{U}(1)]$, and $\mathcal{A}[\mathrm{U}(1)^3]$ do not cancel[16]. In particular, $\mathcal{A}[\mathrm{SU}(3)^2 \otimes \mathrm{U}(1)]$ requires at least an extra field charged under QCD and the new force. By limiting ourselves to SM ingredients, one could try charging the quark doublets under $\mathrm{U}(1)$, but that would spoil the anomalies involving $\mathrm{SU}(2)_L$ unless lepton doublets are also $\mathrm{U}(1)$-charged. Alternatively, one could try charging the right-handed quarks under the new symmetry. We encourage the reader to explore all possible charge combinations and see for themselves why this approach fails.

Therefore, we are forced to add either exotic fields (on top of D_L and D_R) or extend the new force to the lepton sector. In the context of Nelson-Barr models, the latter is encouraged by the experimental evidence of flavour number violation and the suggestion of CP violation in the PMNS matrix.

A simple way to generate the textures required by the Nelson-Barr mechanism is to consider a $\mathrm{U}(1)$ symmetry under which the SM is anomaly-free, plus the following new vector-like (under the SM gauge group) fermions: [144]

$$\begin{aligned}
U_R &\sim (3, 1, 2/3, -Q_R), & U_L &\sim (3, 1, 2/3, Q_L) \\
D_R &\sim (3, 1, -1/3, Q_R), & D_L &\sim (3, 1, -1/3, -Q_L), \\
E_R &\sim (1, 1, -1, Q_R), & E_L &\sim (1, 1, -1, -Q_L), \\
N_R &\sim (1, 1, 0, -Q_R), & N_L &\sim (1, 1, 0, Q_L).
\end{aligned}$$

Motivated abelian (anomaly-free) extensions of the standard model are $\mathrm{U}(1)_R$ and $\mathrm{U}(1)_{B-L}$ with three right-handed neutrinos ν_R. The automatization of the NB

[16] For the charges being equal in magnitude but opposite sign, $\mathcal{A}[\mathrm{U}(1)_Y \otimes \mathrm{U}(1)^2] = 0$, but the other two still survive.

mechanism in the quark sector in these models was studied in Ref. [144]. We note that this UV completion satisfies the four requirements listed in subsection 5.4 for the lepton sector.

To implement high-scale leptogenesis in these models, Majorana neutrino masses are needed. In the $\mathrm{U}(1)_R$ scenario, under the charge relation $|Q_L + Q_R| = 2|Q_{\nu_R}|$, the right-handed neutrinos ν_R get a Majorana mass from the spontaneous symmetry breaking of $\mathrm{U}(1)_R$. In the context of $\mathrm{U}(1)_{\mathrm{B-L}}$, the new scalars X_a cannot couple simultaneously to quarks and leptons as they have different $B - L$ charges. In order to implement NB in both quark and lepton sector, and avoid heavy charged relics (assuming that their mass lies below the reheat temperature of the universe), $Q_L = 1/3$ and $X \sim (1, 1, 0, 2/3)$ [144]. With $Q_R = 5/3$, Majorana masses for the right-handed neutrinos would be allowed.

We close this appendix by including another possibility where both lepton and baryon numbers are local, i.e. $\mathrm{U}(1)_B$ and $\mathrm{U}(1)_L$ are gauged. The fermion content of this theory was proposed in Ref. [211]. Gauge anomalies are cancelled by adding a family of vector-like fermions. Their quantum numbers under $\mathrm{SU}(3) \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y \otimes \mathrm{U}(1)_B \otimes \mathrm{U}(1)_L$ are listed in Table 5.1. As long as $B \neq 4/3$, and $L \neq 4$, condition (iii) is satisfied in both the quark and lepton sectors. We assume that the local baryon and lepton numbers are broken spontaneously by the vevs of $S_B \sim (1, 1, 0, 1, 0)$ and $S_L \sim (1, 1, 0, 0, 3)$, respectively, generating heavy masses for the new fermions[17].

To implement the NB mechanism, we need the new scalars X_a responsible for SCPV to bridge both sectors. This fixes the charge of X_a to be $|Q_{X_a}| = |B - 1/3|$. To extend the mechanism to the lepton sector, condition (ii) must be fulfilled, i.e. $|Q_{X_a}| = |L - 1|$. As an automatic consequence of anomaly cancellation, conditions (i) and (iv) are satisfied.

Regarding high-scale leptogenesis in this scenario; for example, by fixing $L = 3$, Majorana masses for ν_R (M_R dynamically generated by $\langle X_a \rangle$) and for N_R (bare mass m_R) are allowed. Instead, fixing $L = -1$, on top of the dynamical Majorana masses for the ν_Rs, the couplings $N_R^T C \nu_R S_L$ and $N_L^T C N_L X^*$ generate dynamically a $m_{RR} \neq 0$ and $m_L \neq 0$, see Eq. (5.24), respectively.

We note that, within this UV completion, baryon number is broken in one unit. As

[17]In Ref. [211], where this model was proposed, $S_L \sim (1, 1, 0, 2)$ generates Majorana masses for the neutrinos via the type-I seesaw, while the new leptons get mass via the electroweak symmetry breaking. The latter is in tension with the branching fraction of the SM Higgs to two photons, as the new Yukawa couplings need to be $O(1)$.

Field	SU(3)	SU(2)$_L$	U(1)$_Y$	U(1)$_B$	U(1)$_L$
Q_L (×3)	3	2	1/6	1/3	0
L_L (×3)	1	2	-1/2	0	1
u_R (×3)	3	1	2/3	1/3	0
d_R (×3)	3	1	-1/3	1/3	0
e_R (×3)	1	1	-1	0	1
ν_R (×3)	1	1	0	0	1
q_L	3	2	1/6	$B-1$	0
ℓ_L	1	2	-1/2	0	$L-3$
U_R	3	1	2/3	$B-1$	0
D_R	3	1	-1/3	$B-1$	0
E_R	1	1	-1	0	$L-3$
N_R	1	1	0	0	$L-3$
q_R	3	2	1/6	B	0
ℓ_R	1	2	-1/2	0	L
U_L	3	1	2/3	B	0
D_L	3	1	-1/3	B	0
E_L	1	1	-1	0	L
N_L	1	1	0	0	L

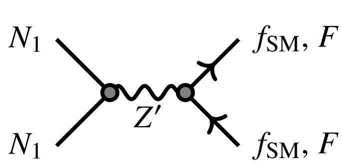
Table 5.1: Quantum numbers of the field content under the SU(3) ⊗ SU(2)$_L$ ⊗ U(1)$_Y$ ⊗ U(1)$_B$ ⊗ U(1)$_L$ gauge group. The first box contains just the SM fields. B and L are unconstrained baryon and lepton charges, respectively.

noted in Ref. [212], dimension-nine operators such as

$$\frac{\xi}{\Lambda^5} Q_L Q_L Q_L L_L S_B^* S_L^* X_a + \cdots \tag{5.49}$$

can mediate proton decay. Here, Λ is the scale that UV-completes the theory presented. In spite of the severe constraints on the lifetime of the proton, $\tau_p \lesssim 10^{34}$ years [213], any of the scales $\langle X \rangle$, $\langle S_L \rangle$ and $\langle S_B \rangle$ are consistent with experiment if the cutoff of the theory is at the Planck scale.

It is worth noting that the new gauge symmetry allowing for the Nelson-Barr mechanism also generates interactions amongst the new and the SM sectors. In the aforementioned examples, the lightest neutral lepton N_1 could easily thermalize at early times via the following interactions:



$$: \quad \sigma_{N_1 N_1 \to (Z')^* \to \bar{f} f} = \frac{(g')^4 \sqrt{s}}{8\pi} \frac{\sqrt{s - 4M_{N_1}^2}}{(s - M_{Z'}^2)^2 + \Gamma_{Z'}^2 M_{Z'}^2}. \tag{5.50}$$

The strength of the gauge coupling (or, alternatively, the mass of the new gauge boson[18]) will determine which scenario of leptogenesis takes place: thermal or dynamical. A study on how these interactions affect the initial thermal population of heavy leptons, as well as the interplay with the out-of-equilibrium condition required at the time of leptogenesis, can be found in Refs. [190, 191].

Appendix 5.B The lightest heavy neutrino mass

In this appendix, we will analytically derive the relation of Eq. (5.29) for a 3×3 matrix with the same features of the heavy mass matrix for the simplified model in Eq. (5.24).

The real positive singular values of the following mass matrix (complex symmetric)

$$\mathcal{M} = \begin{pmatrix} 0 & \mu & m_N \\ \mu & m_R & 0 \\ m_N & 0 & 0 \end{pmatrix}, \tag{5.51}$$

where $m_N, m_R \in \mathbb{R}$ and $\mu \in \mathbb{C}$, can be expressed as the solutions of the following third-order algebraic equations

$$x^3 \pm m_R x^2 - \bar{m}_N^2 x \mp m_R m_N^2 = 0\,, \tag{5.52}$$

where $\bar{m}_N^2 = m_N^2 + |\mu|^2$.

The discriminant is the same for both equations since it is bilinear in the coefficients of the x^2 and x^0 terms, and it reads as

$$\Delta = 4\bar{m}_N^6 - 27 m_N^4 m_R^2 + 18 \bar{m}_N^2 m_N^2 m_R^2 + m_R^2 \bar{m}_N^4 + 4 m_N^2 m_R^4\,. \tag{5.53}$$

Using that $m_R, m_N > 0$, and $\bar{m}_N > m_N$ by definition, we have that $\Delta > 0$. Hence, the two equations above will have three real solutions each. In order to correctly identify the three mass eigenvalues, we want to isolate the three positive solutions.

For clarity, we can rewrite the two equations in (5.52) as

$$m_R = \pm \frac{x(x^2 - \bar{m}_N^2)}{x^2 - m_N^2}\,. \tag{5.54}$$

Using again the property that all the known coefficients are real and positive, the three mass eigenstates can be identified as follows.

[18] The spontaneous breaking scale of the new symmetry U(1), v_S, will determine the new gauge boson mass up to the gauge coupling, $M_{Z'} = \mathcal{O}(g' v_S)$.

- Picking the minus sign, $x > 0$ iff

$$(a): \begin{cases} x > m_N \\ \bar{m}_N > \sqrt{x^2 + x - \frac{m_N^2}{x^2}} \end{cases} \tag{5.55}$$

- Picking the plus sign, $x > 0$ iff

$$(b): \begin{cases} x > m_N \\ \bar{m}_N < \sqrt{x^2 + x - \frac{m_N^2}{x^2}} \end{cases} \qquad \text{or} \qquad (c): x < m_N \tag{5.56}$$

Now, we want to determine which root is the lightest. We can have two mass hierarchies (assuming all non-degenerate mass parameters).

1. $\underline{\mathbf{m_R < m_N}}$: In this case, conditions in case (a) imply that $m_N < x < \bar{m}_N$. Conditions for case (b) give $x > \bar{m}_N$. Hence, case (c) corresponds to the lightest eigenvalues for x.

2. $\underline{\mathbf{m_R > m_N}}$: Case (c) trivially gives the lightest eigenvalue since x is smaller than the lightest scale.

As we proved above, the lightest mass eigenstate is always expressed by the solution x_c to the following equation

$$m_R = \frac{x(\bar{m}_N^2 - x^2)}{m_N^2 - x^2} \quad \text{with} \quad x < m_N\,. \tag{5.57}$$

We can find another upper-bound on the solution, rewriting the above equation as

$$\frac{m_R}{x_c} = \frac{(\bar{m}_N^2 - x_c^2)}{m_N^2 - x_c^2} = 1+\frac{|\mu|^2}{m_N^2 - x_c^2} > 1+\frac{|\mu|^2}{m_N^2} = \left(\frac{\bar{m}_N}{m_N}\right)^2 \Rightarrow x_c < m_R\left(\frac{m_N}{\bar{m}_N}\right)^2\,. \tag{5.58}$$

Hence, we conclude that

$$M_{N_1} \equiv x_c < \mathrm{Min}\left\{m_N, m_R\left(\frac{m_N}{\bar{m}_N}\right)^2\right\}\,. \tag{5.59}$$

Appendix 5.C The Davidson-Ibarra bound for more than three heavy neutrinos

The CP asymmetry summing over the lepton flavors and expanding the loop function around $M_{N_1}^2/M_{N_{j\neq 1}}^2$, see Eq. (5.33), is given by

$$\epsilon_{\mathrm{CP}} = -\frac{3M_{N_1}}{2}\frac{\mathrm{Im}\{\sum_j [\mathsf{h}^\dagger \mathsf{h}]_{1j}^2 M_{N_j}^{-1}\}}{8\pi[\mathsf{h}^\dagger\mathsf{h}]_{11}}, \tag{5.60}$$

where $\mathsf{h}^{i\ell} \equiv (\mathsf{Y}_\nu \mathsf{O}_R)^{ik} \mathsf{U}_H^{*k\ell}$ with $k = 2, 3, 4$. This implies that $\mathsf{h}^\dagger \mathsf{h} = \mathsf{U}_H^T M_D^T M_D \mathsf{U}_H^* (2/v_H^2)$, see Eqs. (5.27). Then, we can rewrite $\mathrm{Im}\{\cdots\}$ in Eq. (5.60) as

$$\begin{aligned}\mathrm{Im}\{[\mathsf{h}^\dagger \mathsf{h}\, \widehat{\mathsf{M}}_H^{-1} \mathsf{h}^T \mathsf{h}^*]_{11}\} &= (2/v_H^2)^2 \mathrm{Im}\{[\mathsf{U}_H^T \mathsf{M}_D^T \mathsf{M}_D \mathsf{U}_H^*\, \widehat{\mathsf{M}}_H^{-1} \mathsf{U}_H^\dagger \mathsf{M}_D^T \mathsf{M}_D \mathsf{U}_H]_{11}\} \\ &= -(2/v_H^2)^2 \mathrm{Im}\{[\mathsf{U}_H^\dagger \mathsf{M}_D^T \mathsf{M}_D \mathsf{U}_H\, \widehat{\mathsf{M}}_H^{-1} \mathsf{U}_H^T \mathsf{M}_D^T \mathsf{M}_D \mathsf{U}_H^*]_{11}\},\end{aligned} \tag{5.61}$$

where in the last step we used that $\mathrm{Im}\{z\} = -\mathrm{Im}\{z^*\}$. By using the identities in Eq. (5.28), i.e. $\mathsf{U}_H\, \widehat{\mathsf{M}}_H^{-1} \mathsf{U}_H^T = \mathsf{M}_H^{-1}$, and $\mathsf{M}_L = -\mathsf{M}_D\, \mathsf{M}_H^{-1} \mathsf{M}_D^T$, we can rewrite the above expression as

$$\mathrm{Im}\{[\mathsf{h}^\dagger \mathsf{h}\, \widehat{\mathsf{M}}_H^{-1} \mathsf{h}^T \mathsf{h}^*]_{11}\} = (2/v_H^2)^2\ \mathrm{Im}\{[\mathsf{U}_H^\dagger \mathsf{M}_D^T \mathsf{M}_L \mathsf{M}_D \mathsf{U}_H^*]_{11}\}. \tag{5.62}$$

In what follows, we will make use of the Casas-Ibarra parametrization [214]. From the seesaw relation in Eq. (5.28) we can write

$$\mathbb{I}_{3\times3} = \left(e^{i\pi/2}\sqrt{\widehat{\mathsf{M}}_L^{-1}}\mathsf{U}_L^T \mathsf{M}_D \mathsf{U}_H\sqrt{\widehat{\mathsf{M}}_H^{-1}}\right)\left(e^{i\pi/2}\sqrt{\widehat{\mathsf{M}}_L^{-1}}\mathsf{U}_L^T \mathsf{M}_D \mathsf{U}_H\sqrt{\widehat{\mathsf{M}}_H^{-1}}\right)^T \equiv \mathsf{R}^T \mathsf{R}. \tag{5.63}$$

Above, $\sqrt{\widehat{\mathsf{M}}} = \mathrm{diag}(\sqrt{\widehat{\mathsf{M}}_{11}}, \cdots, \sqrt{\widehat{\mathsf{M}}_{55}})$. Eq. (5.63) allows us to identify, in between the parenthesis, a complex 5×3 matrix, R, satisfying that $\mathsf{R}^T \mathsf{R} = \mathbb{I}_{3\times3}$. Hence,

$$\mathsf{M}_D = e^{-i\pi/2} \mathsf{U}_L^* \sqrt{\widehat{\mathsf{M}}_L} \mathsf{R}^T \sqrt{\widehat{\mathsf{M}}_H} \mathsf{U}_H^\dagger = e^{i\pi/2} \mathsf{U}_L \sqrt{\widehat{\mathsf{M}}_L} \mathsf{R}^\dagger \sqrt{\widehat{\mathsf{M}}_H} \mathsf{U}_H^T, \tag{5.64}$$

where in the last equality we have made use of the reality of the Dirac mass matrix M_D. This allows us to write the right-hand-side of Eq. (5.62) as a function of R,

$$-(2/v_H^2)^2\ \mathrm{Im}\left\{\left[\sqrt{\widehat{\mathsf{M}}_H} \mathsf{R}^* \widehat{\mathsf{M}}_L^2 \mathsf{R}^\dagger \sqrt{\cancel{\mathsf{M}}_H}\right]_{11}\right\} = (2/v_H^2)^2 M_{N_1} \mathrm{Im}\{(\mathsf{R}\widehat{\mathsf{M}}_L^2 \mathsf{R}^T)_{11}\}. \tag{5.65}$$

In the last step, we have used again that $\mathrm{Im}\{z\} = -\mathrm{Im}\{z^*\}$.

On the other hand, the denominator in Eq. (5.60) can also be written in terms of R,

$$[\mathsf{h}^\dagger \mathsf{h}]_{11} = \frac{2}{v_H^2}\left[\sqrt{\widehat{\mathsf{M}}_H} \mathsf{R} \widehat{\mathsf{M}}_L \mathsf{R}^\dagger \sqrt{\widehat{\mathsf{M}}_H}\right]_{11} = \frac{2}{v_H^2} M_{N_1} [\mathsf{R} \widehat{\mathsf{M}}_L \mathsf{R}^\dagger]_{11}. \tag{5.66}$$

Altogether, combining the numerator and denominator, we get an expression of ϵ_{CP} as a function of R, the light neutrino masses and the lightest heavy neutrino mass:

$$\epsilon_{\mathrm{CP}} = -\frac{3 M_{N_1}}{8\pi v_H^2} \frac{\mathrm{Im}\left\{\sum_{i=1}^3 \mathsf{R}_{1i}^2 m_{\nu_i}^2\right\}}{\sum_{i=1}^3 |\mathsf{R}_{1i}|^2 m_{\nu_i}}. \tag{5.67}$$

Until this step, the result maps to the well-known DI bound [179]. However, since R is not a square matrix, in our case $\mathsf{R}\mathsf{R}^T$ is an idempotent matrix rather than the identity. Let us take a different path to show that $\epsilon_{\rm CP}$ is indeed bounded in this case as well.

Using that $\mathrm{Im}\{z\} < |z|$, and the triangle inequality $|\sum_{i=1}^3 R_{1i}^2 m_{\nu_i}^2| \leq \sum_{i=1}^3 |R_{1i}^2| m_{\nu_i}^2$, we can bound Eq. (5.67) as follows

$$|\epsilon_{\rm CP}| \leq \frac{3M_{N_1}}{8\pi v_H^2} \frac{m_{\nu_3}^2(|\mathsf{R}_{11}^2| + |\mathsf{R}_{12}^2| + |\mathsf{R}_{13}^2|) - (m_{\nu_3}^2 - m_{\nu_1}^2)|\mathsf{R}_{11}^2| - (m_{\nu_3}^2 - m_{\nu_2}^2)|\mathsf{R}_{12}^2|}{m_{\nu_3}(|\mathsf{R}_{13}^2| + |\mathsf{R}_{12}^2| + |\mathsf{R}_{11}^2|) - (m_{\nu_3} - m_{\nu_1})|\mathsf{R}_{11}^2| - (m_{\nu_3} - m_{\nu_2})|\mathsf{R}_{12}^2|}, \tag{5.68}$$

Using the following relations $m_{\nu_3}^2 - m_{\nu_i}^2 = (m_{\nu_3} - m_{\nu_i})(m_{\nu_3} + m_{\nu_i}) \geq (m_{\nu_3} - m_{\nu_i})m_{\nu_3}$ for $i = 1, 2$, the previous expression simplifies to

$$|\epsilon_{\rm CP}| \leq \frac{3M_{N_1}}{8\pi v_H^2} m_{\nu_3}. \tag{5.69}$$

We note that this bound applies generically, regardless of the number of neutral singlets in the theory [215]. As the plots in Sec. 5.5 show, Eq. (5.69) is satisfied by our numerical solutions.

Although having a rectangular matrix, we could still bound $|\epsilon_{\rm CP}|$ thanks to the ratio between weighted sums of R entries. This is not the case, however, for the washout factor K, defined in Eq. (5.38). Writing K *à la* Casas-Ibarra, i.e. in terms of R, according to Eq. (5.66) we have

$$m_\star K = \frac{[\mathsf{h}^\dagger \mathsf{h}]_{11} v_H^2}{2M_{N_1}} = [\mathsf{R}\widehat{\mathsf{M}}_L \mathsf{R}^\dagger]_{11} = \sum_{i=1}^{3} m_{\nu_i} |\mathsf{R}_{1i}^2|. \tag{5.70}$$

Ref. [201] derives an upper bound on the washout parameter, $K \lesssim m_{\nu_3}/m_\star$, by assuming that there are no strong phase cancellations so that $\sum_i |\mathsf{R}_{1i}^2| \simeq |\sum_i \mathsf{R}_{1i}^2|$. In our case, because $\mathsf{R}\mathsf{R}^T \neq \mathbb{I}$, we cannot claim the same. We note that, if R were a real matrix, $\mathsf{R}\mathsf{R}^T$ would be a projection matrix, and hence $(\mathsf{R}\mathsf{R}^T)_{11} = \sum_i \mathsf{R}_{1i}^2 \leq 1$, recovering the bound $K \lesssim m_3/m_\star \sim 10^2$. However, due to the phases of R, we do not expect this bound to be strictly respected.

BIBLIOGRAPHY

[1] S. Patrone, A. Testa, and M. B. Wise. "*Regularization scheme dependence of the counterterms in the galaxy bias expansion*". JCAP 11 (2023). DOI: 10.1088/1475-7516/2023/11/087. arXiv: 2306.08025 [astro-ph.CO] (cit. on p. 3).

[2] P. Fileviez Pérez, C. Murgui, S. Patrone, A. Testa, and M. B. Wise. "*Finite naturalness and quark-lepton unification*". Phys. Rev. D 109.1 (2024). DOI: 10.1103/PhysRevD.109.015011. arXiv: 2308.07367 [hep-ph] (cit. on pp. 3, 91, 92).

[3] S. Patrone, N. Blinov, and R. Plestid. "*Long-lived axionlike particles from electromagnetic cascades*". Phys. Rev. D 113.7 (2026). DOI: 10.1103/38hq-qk7k. arXiv: 2509.14310 [hep-ph] (cit. on p. 3).

[4] C. Murgui and S. Patrone. "*Leptogenesis in automatic Nelson-Barr models*". JHEP 09 (2025). DOI: 10.1007/JHEP09(2025)113. arXiv: 2506.18963 [hep-ph] (cit. on p. 4).

[5] V. Desjacques, D. Jeong, and F. Schmidt. "Large-Scale Galaxy Bias". *Phys. Rept.* 733 (2018), pp. 1–193. DOI: 10.1016/j.physrep.2017.12.002. arXiv: 1611.09787 [astro-ph.CO] (cit. on pp. 6, 16, 20).

[6] M. Ryle and R. W. Clarke. "An examination of the steady-state model in the light of some recent observations of radio sources". *Monthly Notices of the Royal Astronomical Society* 122 (Jan. 1961), p. 349. DOI: 10.1093/mnras/122.4.349 (cit. on p. 6).

[7] N. Kaiser. "On the spatial correlations of Abell clusters." *The Astrophysical Journal Letters* 284 (Sept. 1984), pp. L9–L12. DOI: 10.1086/184341 (cit. on p. 7).

[8] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. "The Statistics of Peaks of Gaussian Random Fields". *Astrophysical Journal* 304 (May 1986), p. 15. DOI: 10.1086/164143 (cit. on p. 7).

[9] S. Cole and N. Kaiser. "Biased clustering in the cold dark matter cosmogony." *Monthly Notices of the Royal Astronomical Society* 237 (Apr. 1989), pp. 1127–1146. DOI: 10.1093/mnras/237.4.1127 (cit. on p. 7).

[10] H. J. Mo and S. D. M. White. "An Analytic model for the spatial clustering of dark matter halos". *Mon. Not. Roy. Astron. Soc.* 282 (1996), p. 347. DOI: 10.1093/mnras/282.2.347. arXiv: astro-ph/9512127 (cit. on p. 7).

[11] R. K. Sheth and G. Tormen. "Large scale bias and the peak background split". *Mon. Not. Roy. Astron. Soc.* 308 (1999), p. 119. DOI: 10.1046/j.1365-8711.1999.02692.x. arXiv: astro-ph/9901122 (cit. on p. 7).

[12] J. N. Fry and E. Gaztanaga. "Biasing and hierarchical statistics in large scale structure". *Astrophys. J.* 413 (1993), pp. 447–452. DOI: 10.1086/173015. arXiv: astro-ph/9302009 (cit. on p. 7).

[13] P. McDonald. "Primordial non-Gaussianity: large-scale structure signature in the perturbative bias model". *Phys. Rev. D* 78 (2008), p. 123519. DOI: 10.1103/PhysRevD.78.123519. arXiv: 0806.1061 [astro-ph] (cit. on pp. 11, 16).

[14] N. Dalal, O. Dore, D. Huterer, and A. Shirokov. "The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects". *Phys. Rev. D* 77 (2008), p. 123514. DOI: 10.1103/PhysRevD.77.123514. arXiv: 0710.4560 [astro-ph] (cit. on p. 12).

[15] P. McDonald. "Clustering of dark matter tracers: Renormalizing the bias parameters". *Phys. Rev. D* 74 (2006). [Erratum: Phys.Rev.D 74, 129901 (2006)], p. 103512. DOI: 10.1103/PhysRevD.74.129901. arXiv: astro-ph/0609413 (cit. on pp. 16, 21).

[16] P. McDonald and A. Roy. "Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS". *JCAP* 08 (2009), p. 020. DOI: 10.1088/1475-7516/2009/08/020. arXiv: 0902.0991 [astro-ph.CO] (cit. on p. 16).

[17] V. Assassi, D. Baumann, D. Green, and M. Zaldarriaga. "Renormalized Halo Bias". *JCAP* 08 (2014), p. 056. DOI: 10.1088/1475-7516/2014/08/056. arXiv: 1402.5916 [astro-ph.CO] (cit. on pp. 16, 19, 24).

[18] V. Assassi, D. Baumann, and F. Schmidt. "Galaxy Bias and Primordial Non-Gaussianity". *JCAP* 12 (2015), p. 043. DOI: 10.1088/1475-7516/2015/12/043. arXiv: 1510.03723 [astro-ph.CO] (cit. on pp. 16, 20–22, 26, 28, 29).

[19] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (cit. on pp. 17, 86, 98, 101).

[20] P. J. Peebles. *The Large-Scale Structure of the Universe*. Princeton University Press, Nov. 1980. ISBN: 978-0-691-20671-4 (cit. on p. 18).

[21] J. N. Fry. "The Galaxy correlation hierarchy in perturbation theory". *Astrophys. J.* 279 (1984), pp. 499–510. DOI: 10.1086/161913 (cit. on p. 18).

[22] M. H. Goroff, B. Grinstein, S. J. Rey, and M. B. Wise. "Coupling of Modes of Cosmological Mass Density Fluctuations". *Astrophys. J.* 311 (1986), pp. 6–14. DOI: 10.1086/164749 (cit. on p. 18).

[23] D. Jeong. “Cosmology with high ($z > 1$) redshift galaxy surveys”. PhD thesis. University of Texas at Austin, 2010. URL: https://repositories.lib.utexas.edu/handle/2152/ETD-UT-2010-08-1781 (cit. on p. 19).

[24] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore. “The Effective Field Theory of Cosmological Large Scale Structures”. *JHEP* 09 (2012), p. 082. DOI: 10.1007/JHEP09(2012)082. arXiv: 1206.2926 [astro-ph.CO] (cit. on p. 19).

[25] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga. “Cosmological Non-Linearities as an Effective Fluid”. *JCAP* 07 (2012), p. 051. DOI: 10.1088/1475-7516/2012/07/051. arXiv: 1004.2488 [astro-ph.CO] (cit. on p. 19).

[26] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga. “LSS constraints with controlled theoretical uncertainties” (Feb. 2016). arXiv: 1602.00674 [astro-ph.CO] (cit. on p. 20).

[27] M. Mirbabayi, F. Schmidt, and M. Zaldarriaga. “Biased Tracers and Time Evolution”. *JCAP* 07 (2015), p. 030. DOI: 10.1088/1475-7516/2015/07/030. arXiv: 1412.5169 [astro-ph.CO] (cit. on pp. 21, 26).

[28] A. Kehagias and A. Riotto. “Conformal Symmetries of FRW Accelerating Cosmologies”. *Nucl. Phys. B* 884 (2014), pp. 547–565. DOI: 10.1016/j.nuclphysb.2014.05.006. arXiv: 1309.3671 [hep-th] (cit. on p. 31).

[29] G. F. Giudice. “Naturalness after LHC8”. *PoS* EPS-HEP2013 (2013), p. 163. DOI: 10.22323/1.180.0163. arXiv: 1307.7879 [hep-ph] (cit. on p. 38).

[30] N. Craig. “Naturalness: A Snowmass White Paper”. *Snowmass 2021*. May 2022. arXiv: 2205.05708 [hep-ph] (cit. on p. 38).

[31] K. G. Wilson. “The renormalization group and critical phenomena”. *Rev. Mod. Phys.* 55 (1983), pp. 583–600. DOI: 10.1103/RevModPhys.55.583 (cit. on p. 38).

[32] M. Farina, D. Pappadopulo, and A. Strumia. “A modified naturalness principle and its experimental tests”. *JHEP* 08 (2013), p. 022. DOI: 10.1007/JHEP08(2013)022. arXiv: 1303.7244 [hep-ph] (cit. on pp. 38, 94, 111).

[33] H. Georgi and S. L. Glashow. “Unity of All Elementary Particle Forces”. *Phys. Rev. Lett.* 32 (1974), pp. 438–441. DOI: 10.1103/PhysRevLett.32.438 (cit. on p. 39).

[34] J. C. Pati and A. Salam. “Lepton Number as the Fourth Color”. *Phys. Rev. D* 10 (1974). [Erratum: Phys.Rev.D 11, 703–703 (1975)], pp. 275–289. DOI: 10.1103/PhysRevD.10.275 (cit. on p. 39).

[35] J. C. Pati and A. Salam. “Is Baryon Number Conserved?” *Phys. Rev. Lett.* 31 (1973), pp. 661–664. DOI: 10.1103/PhysRevLett.31.661 (cit. on p. 39).

[36] G. Valencia and S. Willenbrock. "Quark - lepton unification and rare meson decays". *Phys. Rev. D* 50 (1994), pp. 6843–6848. DOI: 10.1103/PhysRevD.50.6843. arXiv: hep-ph/9409201 (cit. on p. 39).

[37] A. D. Smirnov. "Vector leptoquark mass limits and branching ratios of $K_L^0, B^0, B_s \to l_i^+ l_j^-$ decays with account of fermion mixing in leptoquark currents". *Mod. Phys. Lett. A* 33 (2018), p. 1850019. DOI: 10.1142/S0217732318500190. arXiv: 1801.02895 [hep-ph] (cit. on p. 39).

[38] H. Gedeonová and M. Hudec. "All possible first signals of gauge leptoquark in quark-lepton unification and beyond". *Phys. Rev. D* 107.9 (2023), p. 095029. DOI: 10.1103/PhysRevD.107.095029. arXiv: 2210.00347 [hep-ph] (cit. on p. 39).

[39] A. D. Smirnov. "The Minimal quark - lepton symmetry model and the limit on Z-prime mass". *Phys. Lett. B* 346 (1995), pp. 297–302. DOI: 10.1016/0370-2693(95)00015-D. arXiv: hep-ph/9503239 (cit. on pp. 39, 42).

[40] P. Fileviez Perez and M. B. Wise. "Low Scale Quark-Lepton Unification". *Phys. Rev. D* 88 (2013), p. 057703. DOI: 10.1103/PhysRevD.88.057703. arXiv: 1307.6213 [hep-ph] (cit. on pp. 39, 42, 43, 49).

[41] R. N. Mohapatra. "Mechanism for Understanding Small Neutrino Mass in Superstring Theories". *Phys. Rev. Lett.* 56 (1986), pp. 561–563. DOI: 10.1103/PhysRevLett.56.561 (cit. on p. 43).

[42] R. N. Mohapatra and J. W. F. Valle. "Neutrino Mass and Baryon Number Nonconservation in Superstring Models". *Phys. Rev. D* 34 (1986), p. 1642. DOI: 10.1103/PhysRevD.34.1642 (cit. on p. 43).

[43] T. Faber et al. "A unified leptoquark model confronted with lepton non-universality in *B*-meson decays". *Phys. Lett. B* 787 (2018), pp. 159–166. DOI: 10.1016/j.physletb.2018.10.051. arXiv: 1808.05511 [hep-ph] (cit. on p. 49).

[44] T. Faber et al. "Collider phenomenology of a unified leptoquark model". *Phys. Rev. D* 101.9 (2020), p. 095024. DOI: 10.1103/PhysRevD.101.095024. arXiv: 1812.07592 [hep-ph] (cit. on p. 49).

[45] P. Fileviez Perez, E. Golias, and A. D. Plascencia. "Two-Higgs-doublet model and quark-lepton unification". *JHEP* 08 (2022), p. 293. DOI: 10.1007/JHEP08(2022)293. arXiv: 2205.02235 [hep-ph] (cit. on p. 49).

[46] M. J. Dolan, T. P. Dutka, and R. R. Volkas. "Lowering the scale of Pati-Salam breaking through seesaw mixing". *JHEP* 05 (2021), p. 199. DOI: 10.1007/JHEP05(2021)199. arXiv: 2012.05976 [hep-ph] (cit. on p. 50).

[47] F. Wilczek. "Problem of Strong *P* and *T* Invariance in the Presence of Instantons". *Phys. Rev. Lett.* 40 (1978), pp. 279–282. DOI: 10.1103/PhysRevLett.40.279 (cit. on pp. 52, 87).

[48] S. Weinberg. “A New Light Boson?” *Phys. Rev. Lett.* 40 (1978), pp. 223–226. DOI: 10.1103/PhysRevLett.40.223 (cit. on pp. 52, 87).

[49] J. E. Kim. “Weak Interaction Singlet and Strong CP Invariance”. *Phys. Rev. Lett.* 43 (1979), p. 103. DOI: 10.1103/PhysRevLett.43.103 (cit. on pp. 52, 87).

[50] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov. “Can Confinement Ensure Natural CP Invariance of Strong Interactions?” *Nucl. Phys. B* 166 (1980), pp. 493–506. DOI: 10.1016/0550-3213(80)90209-6 (cit. on pp. 52, 87).

[51] M. Dine, W. Fischler, and M. Srednicki. “A Simple Solution to the Strong CP Problem with a Harmless Axion”. *Phys. Lett. B* 104 (1981), pp. 199–202. DOI: 10.1016/0370-2693(81)90590-6 (cit. on pp. 52, 87).

[52] A. R. Zhitnitsky. “On Possible Suppression of the Axion Hadron Interactions. (In Russian)”. *Sov. J. Nucl. Phys.* 31 (1980), p. 260 (cit. on pp. 52, 87).

[53] J. Beacham et al. “Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report”. *J. Phys. G* 47.1 (2020), p. 010501. DOI: 10.1088/1361-6471/ab4cd2. arXiv: 1901.09966 [hep-ex] (cit. on pp. 52–54, 73).

[54] B. Batell, N. Blinov, C. Hearty, and R. McGehee. “Exploring Dark Sector Portals with High Intensity Experiments”. *Snowmass 2021*. July 2022. arXiv: 2207.06905 [hep-ph] (cit. on p. 52).

[55] P. Ilten et al. “Experiments and Facilities for Accelerator-Based Dark Sector Searches”. *Snowmass 2021*. June 2022. arXiv: 2206.04220 [hep-ex] (cit. on p. 52).

[56] N. Blinov, P. J. Fox, K. J. Kelly, P. A. N. Machado, and R. Plestid. “Dark fluxes from electromagnetic cascades”. *JHEP* 07 (2024), p. 022. DOI: 10.1007/JHEP07(2024)022. arXiv: 2401.06843 [hep-ph] (cit. on pp. 53, 57, 59, 61, 76).

[57] V. Brdar et al. “Axionlike Particles at Future Neutrino Experiments: Closing the Cosmological Triangle”. *Phys. Rev. Lett.* 126.20 (2021), p. 201801. DOI: 10.1103/PhysRevLett.126.201801. arXiv: 2011.07054 [hep-ph] (cit. on pp. 53, 74).

[58] A. A. Aguilar-Arevalo et al. “Prospects for detecting axionlike particles at the Coherent CAPTAIN-Mills experiment”. *Phys. Rev. D* 107.9 (2023), p. 095036. DOI: 10.1103/PhysRevD.107.095036. arXiv: 2112.09979 [hep-ph] (cit. on p. 53).

[59] F. Capozzi et al. “New constraints on ALP couplings to electrons and photons from ArgoNeuT and the MiniBooNE beam dump”. *Phys. Rev. D* 108.7 (2023), p. 075019. DOI: 10.1103/PhysRevD.108.075019. arXiv: 2307.03878 [hep-ph] (cit. on pp. 53, 74).

[60] Y.-S. Tsai. "AXION BREMSSTRAHLUNG BY AN ELECTRON BEAM". *Phys. Rev. D* 34 (1986). Ed. by S. C. Loken, p. 1326. DOI: 10.1103/PhysRevD.34.1326 (cit. on pp. 53, 61).

[61] J. D. Bjorken et al. "Search for Neutral Metastable Penetrating Particles Produced in the SLAC Beam Dump". *Phys. Rev. D* 38 (1988), p. 3375. DOI: 10.1103/PhysRevD.38.3375 (cit. on pp. 53, 74–76).

[62] J. Jerhot, B. Döbrich, F. Ertas, F. Kahlhoefer, and T. Spadaro. "ALPINIST: Axion-Like Particles In Numerous Interactions Simulated and Tabulated". *JHEP* 07 (2022), p. 094. DOI: 10.1007/JHEP07(2022)094. arXiv: 2201.05170 [hep-ph] (cit. on pp. 54, 67, 71, 73–76).

[63] O. Aberle et al. *BDF/SHiP at the ECN3 high-intensity beam facility*. Tech. rep. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2839677 (cit. on pp. 53, 54, 66, 70).

[64] R. Albanese et al. *BDF/SHiP at the ECN3 high-intensity beam facility*. Tech. rep. Geneva: CERN, 2023. URL: https://cds.cern.ch/record/2878604 (cit. on pp. 53, 54, 66).

[65] B. Döbrich, J. Jaeckel, and T. Spadaro. "Light in the beam dump - ALP production from decay photons in proton beam-dumps". *JHEP* 05 (2019). [Erratum: JHEP 10, 046 (2020)], p. 213. DOI: 10.1007/JHEP05(2019)213. arXiv: 1904.02091 [hep-ph] (cit. on pp. 53, 63, 71).

[66] M. Anelli et al. "A facility to Search for Hidden Particles (SHiP) at the CERN SPS" (Apr. 2015). arXiv: 1504.04956 [physics.ins-det] (cit. on p. 53).

[67] M. Battaglieri et al. "Dark Matter Search in a Beam-Dump eXperiment (BDX) at Jefferson Lab" (July 2016). arXiv: 1607.01390 [hep-ex] (cit. on pp. 54, 67, 71).

[68] S. S. Chakrabarty and I. Jaeglé. "Search for dark photon, axion-like particles, dark scalar, or light dark matter in Compton-like processes" (Mar. 2019). arXiv: 1903.06225 [hep-ph] (cit. on p. 55).

[69] M. Bauer, M. Neubert, S. Renner, M. Schnubel, and A. Thamm. "The Low-Energy Effective Theory of Axions and ALPs". *JHEP* 04 (2021), p. 063. DOI: 10.1007/JHEP04(2021)063. arXiv: 2012.12272 [hep-ph] (cit. on pp. 55, 56).

[70] A. Eberhart, M. Fedele, F. Kahlhoefer, E. Ravensburg, and R. Ziegler. "Leptophilic ALPs in Laboratory Experiments" (Apr. 2025). arXiv: 2504.05873 [hep-ph] (cit. on p. 56).

[71] N. Blinov, E. Kowalczyk, and M. Wynne. "Axion-like particle searches at DarkQuest". *JHEP* 02 (2022), p. 036. DOI: 10.1007/JHEP02(2022)036. arXiv: 2112.09814 [hep-ph] (cit. on pp. 56, 74).

[72] M. Bauer, M. Neubert, S. Renner, M. Schnubel, and A. Thamm. "Consistent Treatment of Axions in the Weak Chiral Lagrangian". *Phys. Rev. Lett.* 127.8 (2021), p. 081803. DOI: 10.1103/PhysRevLett.127.081803. arXiv: 2102.13112 [hep-ph] (cit. on p. 56).

[73] G. Alonso-Álvarez, M. B. Gavela, and P. Quilez. "Axion couplings to electroweak gauge bosons". *Eur. Phys. J. C* 79.3 (2019), p. 223. DOI: 10.1140/epjc/s10052-019-6732-5. arXiv: 1811.05466 [hep-ph] (cit. on p. 57).

[74] V. Plakkot and S. Hoof. "Anomaly ratio distributions of hadronic axion models with multiple heavy quarks". *Phys. Rev. D* 104.7 (2021), p. 075017. DOI: 10.1103/PhysRevD.104.075017. arXiv: 2107.12378 [hep-ph] (cit. on p. 57).

[75] R. Daido, F. Takahashi, and N. Yokozaki. "Enhanced axion–photon coupling in GUT with hidden photon". *Phys. Lett. B* 780 (2018), pp. 538–542. DOI: 10.1016/j.physletb.2018.03.039. arXiv: 1801.10344 [hep-ph] (cit. on p. 57).

[76] M. Farina, D. Pappadopulo, F. Rompineve, and A. Tesi. "The photo-philic QCD axion". *JHEP* 01 (2017), p. 095. DOI: 10.1007/JHEP01(2017)095. arXiv: 1611.09855 [hep-ph] (cit. on p. 57).

[77] P. Agrawal, J. Fan, M. Reece, and L.-T. Wang. "Experimental Targets for Photon Couplings of the QCD Axion". *JHEP* 02 (2018), p. 006. DOI: 10.1007/JHEP02(2018)006. arXiv: 1709.06085 [hep-ph] (cit. on p. 57).

[78] L. Di Luzio, M. Giannotti, E. Nardi, and L. Visinelli. "The landscape of QCD axion models". *Phys. Rept.* 870 (2020), pp. 1–117. DOI: 10.1016/j.physrep.2020.06.002. arXiv: 2003.01100 [hep-ph] (cit. on p. 57).

[79] T. Zhou, R. Plestid, K. J. Kelly, N. Blinov, and P. J. Fox. "Long-lived vectors from electromagnetic cascades at SHiP". *JHEP* 02 (2025), p. 107. DOI: 10.1007/JHEP02(2025)107. arXiv: 2412.01880 [hep-ph] (cit. on pp. 57, 62, 63, 65, 75, 76, 80).

[80] S. Patrone. *ALPETITE*. 2025. URL: https://github.com/spatrone/ALPETITE (cit. on p. 57).

[81] J. D. Bjorken, R. Essig, P. Schuster, and N. Toro. "New Fixed-Target Experiments to Search for Dark Gauge Forces". *Phys. Rev. D* 80 (2009), p. 075018. DOI: 10.1103/PhysRevD.80.075018. arXiv: 0906.0580 [hep-ph] (cit. on pp. 59, 61).

[82] A. Celentano, L. Darmé, L. Marsicano, and E. Nardi. "New production channels for light dark matter in hadronic showers". *Phys. Rev. D* 102.7 (2020), p. 075026. DOI: 10.1103/PhysRevD.102.075026. arXiv: 2006.09419 [hep-ph] (cit. on pp. 59, 62, 76).

[83] C. F. von Weizsacker. “Radiation emitted in collisions of very fast electrons”. *Z. Phys.* 88 (1934), pp. 612–625. DOI: 10.1007/BF01333110 (cit. on p. 60).

[84] E. J. Williams. “Nature of the high-energy particles of penetrating radiation and status of ionization and radiation formulae”. *Phys. Rev.* 45 (1934), pp. 729–730. DOI: 10.1103/PhysRev.45.729 (cit. on p. 60).

[85] V. M. Budnev, I. F. Ginzburg, G. V. Meledin, and V. G. Serbo. “The Two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation”. *Phys. Rept.* 15 (1975), pp. 181–281. DOI: 10.1016/0370-1573(75)90009-5 (cit. on p. 60).

[86] G. P. Lepage. “Adaptive multidimensional integration: VEGAS enhanced”. *J. Comput. Phys.* 439 (2021), p. 110386. DOI: 10.1016/j.jcp.2021.110386. arXiv: 2009.05112 [physics.comp-ph] (cit. on p. 61).

[87] R. Plestid and M. B. Wise. “Atomic binding corrections for high-energy fixed target experiments”. *Phys. Rev. D* 110.5 (2024), p. 056032. DOI: 10.1103/PhysRevD.110.056032. arXiv: 2403.12184 [hep-ph] (cit. on p. 62).

[88] L. Marsicano et al. “Novel Way to Search for Light Dark Matter in Lepton Beam-Dump Experiments”. *Phys. Rev. Lett.* 121.4 (2018), p. 041802. DOI: 10.1103/PhysRevLett.121.041802. arXiv: 1807.05884 [hep-ex] (cit. on pp. 62, 76).

[89] E. Nardi, C. D. R. Carvajal, A. Ghoshal, D. Meloni, and M. Raggi. “Resonant production of dark photons in positron beam dump experiments”. *Phys. Rev. D* 97.9 (2018), p. 095004. DOI: 10.1103/PhysRevD.97.095004. arXiv: 1802.04756 [hep-ph] (cit. on pp. 62, 76).

[90] L. Marsicano et al. “Dark photon production through positron annihilation in beam-dump experiments”. *Phys. Rev. D* 98.1 (2018), p. 015031. DOI: 10.1103/PhysRevD.98.015031. arXiv: 1802.03794 [hep-ex] (cit. on pp. 62, 76).

[91] F. Arias-Aragón, G. G. di Cortona, and E. Nardi. “Unveiling atomic electron motion effects in e^+e^- high energy collisions at NA64” (July 2025). arXiv: 2507.08941 [hep-ph] (cit. on p. 62).

[92] N. Serra. Private communication. 2024 (cit. on p. 65).

[93] R. Essig et al. “Probing millicharged particles at an electron beam dump with ultralow-threshold sensors”. *JHEP* 04 (2025), p. 057. DOI: 10.1007/JHEP04(2025)057. arXiv: 2412.09652 [hep-ph] (cit. on p. 65).

[94] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on p. 67).

[95] G. Magill, R. Plestid, M. Pospelov, and Y.-D. Tsai. "Millicharged particles in neutrino experiments". *Phys. Rev. Lett.* 122.7 (2019), p. 071801. DOI: 10.1103/PhysRevLett.122.071801. arXiv: 1806.03310 [hep-ph] (cit. on p. 67).

[96] G. Machado Santos Soares. *Optimization of the Selection of Hidden Particles in the SHiP Experiment*. May 2021 (cit. on p. 70).

[97] M. Battaglieri et al. "The BDX-MINI detector for Light Dark Matter search at JLab". *Eur. Phys. J. C* 81.2 (2021), p. 164. DOI: 10.1140/epjc/s10052-021-08957-5. arXiv: 2011.10532 [physics.ins-det] (cit. on p. 71).

[98] D. Aloni, C. Fanelli, Y. Soreq, and M. Williams. "Photoproduction of Axionlike Particles". *Phys. Rev. Lett.* 123.7 (2019), p. 071801. DOI: 10.1103/PhysRevLett.123.071801. arXiv: 1903.03586 [hep-ph] (cit. on p. 74).

[99] M. J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer, and K. Schmidt-Hoberg. "Revised constraints and Belle II sensitivity for visible and invisible axion-like particles". *JHEP* 12 (2017). [Erratum: JHEP 03, 190 (2021)], p. 094. DOI: 10.1007/JHEP12(2017)094. arXiv: 1709.00009 [hep-ph] (cit. on pp. 74–76).

[100] D. Banerjee et al. "Search for Axionlike and Scalar Particles with the NA64 Experiment". *Phys. Rev. Lett.* 125.8 (2020), p. 081801. DOI: 10.1103/PhysRevLett.125.081801. arXiv: 2005.02710 [hep-ex] (cit. on p. 74).

[101] R. Mammen Abraham et al. "Shining light on the dark sector: search for axion-like particles and other new physics in photonic final states with FASER". *JHEP* 01 (2025), p. 199. DOI: 10.1007/JHEP01(2025)199. arXiv: 2410.10363 [hep-ex] (cit. on p. 74).

[102] R. Balkin et al. "Probing axion-like particles at the Electron-Ion Collider". *JHEP* 02 (2024), p. 123. DOI: 10.1007/JHEP02(2024)123. arXiv: 2310.08827 [hep-ph] (cit. on p. 74).

[103] F. Acanfora, R. Franceschini, A. Mastroddi, and D. Redigolo. "Fusing photons into diphoton resonances at Belle II and beyond". *JHEP* 12 (2024), p. 099. DOI: 10.1007/JHEP12(2024)099. arXiv: 2406.14614 [hep-ph] (cit. on p. 74).

[104] J. Blumlein et al. "Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment". *Z. Phys. C* 51 (1991), pp. 341–350. DOI: 10.1007/BF01548556 (cit. on pp. 74–76).

[105] J. Blumlein et al. "Limits on the mass of light (pseudo)scalar particles from Bethe-Heitler e+ e- and mu+ mu- pair production in a proton - iron beam dump experiment". *Int. J. Mod. Phys. A* 7 (1992), pp. 3835–3850. DOI: 10.1142/S0217751X9200171X (cit. on p. 74).

[106] A. A. Aguilar-Arevalo et al. "Dark Matter Search in Nucleon, Pion, and Electron Channels from a Proton Beam Dump with MiniBooNE". *Phys. Rev. D* 98.11 (2018), p. 112004. DOI: 10.1103/PhysRevD.98.112004. arXiv: 1807.06137 [hep-ex] (cit. on p. 74).

[107] D. F. G. Fiorillo, T. Pitik, and E. Vitagliano. "Energy Transfer by Feebly Interacting Particles in Supernovae: The Trapping Regime". *Phys. Rev. Lett.* 135.7 (2025), p. 071005. DOI: 10.1103/cz94-dqxt. arXiv: 2503.13653 [hep-ph] (cit. on p. 74).

[108] P. F. Depta, M. Hufnagel, and K. Schmidt-Hoberg. "Robust cosmological constraints on axion-like particles". *JCAP* 05 (2020), p. 009. DOI: 10.1088/1475-7516/2020/05/009. arXiv: 2002.08370 [hep-ph] (cit. on p. 74).

[109] M. Diamond, D. F. G. Fiorillo, G. Marques-Tavares, I. Tamborra, and E. Vitagliano. "Multimessenger Constraints on Radiatively Decaying Axions from GW170817". *Phys. Rev. Lett.* 132.10 (2024), p. 101004. DOI: 10.1103/PhysRevLett.132.101004. arXiv: 2305.10327 [hep-ph] (cit. on p. 74).

[110] Z. Bai et al. "New physics searches with an optical dump at LUXE". *Phys. Rev. D* 106.11 (2022), p. 115034. DOI: 10.1103/PhysRevD.106.115034. arXiv: 2107.13554 [hep-ph] (cit. on p. 74).

[111] F. Bergsma et al. "Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions". *Phys. Lett. B* 157 (1985), pp. 458–462. DOI: 10.1016/0370-2693(85)90400-9 (cit. on pp. 75, 76).

[112] B. Döbrich, J. Jaeckel, F. Kahlhoefer, A. Ringwald, and K. Schmidt-Hoberg. "ALPtraum: ALP production in proton beam dump experiments". *JHEP* 02 (2016), p. 018. DOI: 10.1007/JHEP02(2016)018. arXiv: 1512.03069 [hep-ph] (cit. on p. 76).

[113] S. N. Gninenko, D. V. Kirpichnikov, M. M. Kirsanov, and N. V. Krasnikov. "The exact tree-level calculation of the dark photon production in high-energy electron scattering at the CERN SPS". *Phys. Lett. B* 782 (2018), pp. 406–411. DOI: 10.1016/j.physletb.2018.05.010. arXiv: 1712.05706 [hep-ph] (cit. on p. 79).

[114] Y.-S. Liu, D. McKeen, and G. A. Miller. "Validity of the Weizsäcker-Williams approximation and the analysis of beam dump experiments: Production of a new scalar boson". *Phys. Rev. D* 95.3 (2017), p. 036010. DOI: 10.1103/PhysRevD.95.036010. arXiv: 1609.06781 [hep-ph] (cit. on p. 79).

[115] Y.-S. Liu and G. A. Miller. "Validity of the Weizsäcker-Williams approximation and the analysis of beam dump experiments: Production of an axion, a dark photon, or a new axial-vector boson". *Phys. Rev. D* 96.1 (2017), p. 016004. DOI: 10.1103/PhysRevD.96.016004. arXiv: 1705.01633 [hep-ph] (cit. on p. 79).

[116] S. Navas et al. "Review of particle physics". *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001 (cit. on p. 80).

[117] R. H. Cyburt, B. D. Fields, K. A. Olive, and T.-H. Yeh. "Big Bang Nucleosynthesis: 2015". *Rev. Mod. Phys.* 88 (2016), p. 015004. DOI: 10.1103/RevModPhys.88.015004. arXiv: 1505.01076 [astro-ph.CO] (cit. on p. 86).

[118] R. L. Workman et al. "Review of Particle Physics". *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097 (cit. on p. 86).

[119] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497 (cit. on pp. 86, 88, 98).

[120] M. C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz. "NuFIT: Three-Flavour Global Analyses of Neutrino Oscillation Experiments". *Universe* 7.12 (2021), p. 459. DOI: 10.3390/universe7120459. arXiv: 2111.03086 [hep-ph] (cit. on p. 86).

[121] Q. R. Ahmad et al. "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory". *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. arXiv: nucl-ex/0204008 (cit. on p. 86).

[122] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003 (cit. on p. 86).

[123] K. Eguchi et al. "First results from KamLAND: Evidence for reactor antineutrino disappearance". *Phys. Rev. Lett.* 90 (2003), p. 021802. DOI: 10.1103/PhysRevLett.90.021802. arXiv: hep-ex/0212021 (cit. on p. 86).

[124] C. Abel et al. "Measurement of the Permanent Electric Dipole Moment of the Neutron". *Phys. Rev. Lett.* 124.8 (2020), p. 081803. DOI: 10.1103/PhysRevLett.124.081803. arXiv: 2001.11966 [hep-ex] (cit. on pp. 87, 114).

[125] M. Pospelov and A. Ritz. "Theta vacua, QCD sum rules, and the neutron electric dipole moment". *Nucl. Phys. B* 573 (2000), pp. 177–200. DOI: 10.1016/S0550-3213(99)00817-2. arXiv: hep-ph/9908508 (cit. on p. 87).

[126] J. Liang et al. "Nucleon electric dipole moment from the θ term with lattice chiral fermions". *Phys. Rev. D* 108.9 (2023), p. 094512. DOI: 10.1103/PhysRevD.108.094512. arXiv: 2301.04331 [hep-lat] (cit. on p. 87).

[127] A. V. Manohar. "Introduction to Effective Field Theories" (Apr. 2018). Ed. by S. Davidson, P. Gambino, M. Laine, M. Neubert, and C. Salomon. DOI: 10.1093/oso/9780198855743.003.0002. arXiv: 1804.05863 [hep-ph] (cit. on p. 87).

[128] R. D. Peccei and H. R. Quinn. "CP Conservation in the Presence of Instantons". *Phys. Rev. Lett.* 38 (1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440 (cit. on p. 87).

[129] R. D. Peccei and H. R. Quinn. "Constraints Imposed by CP Conservation in the Presence of Instantons". *Phys. Rev. D* 16 (1977), pp. 1791–1797. DOI: 10.1103/PhysRevD.16.1791 (cit. on p. 87).

[130] A. E. Nelson. "Naturally Weak CP Violation". *Phys. Lett. B* 136 (1984), pp. 387–391. DOI: 10.1016/0370-2693(84)92025-2 (cit. on pp. 87, 89).

[131] S. M. Barr. "Solving the Strong CP Problem Without the Peccei-Quinn Symmetry". *Phys. Rev. Lett.* 53 (1984), p. 329. DOI: 10.1103/PhysRevLett.53.329 (cit. on pp. 87, 89).

[132] A. E. Nelson. "Calculation of θ Barr". *Phys. Lett. B* 143 (1984), pp. 165–170. DOI: 10.1016/0370-2693(84)90827-X (cit. on p. 87).

[133] H. M. Georgi, L. J. Hall, and M. B. Wise. "Grand Unified Models With an Automatic Peccei-Quinn Symmetry". *Nucl. Phys. B* 192 (1981), pp. 409–416. DOI: 10.1016/0550-3213(81)90433-8 (cit. on p. 87).

[134] M. Kamionkowski and J. March-Russell. "Planck scale physics and the Peccei-Quinn mechanism". *Phys. Lett. B* 282 (1992), pp. 137–141. DOI: 10.1016/0370-2693(92)90492-M. arXiv: hep-th/9202003 (cit. on p. 87).

[135] R. Holman et al. "Solutions to the strong CP problem in a world with gravity". *Phys. Lett. B* 282 (1992), pp. 132–136. DOI: 10.1016/0370-2693(92)90491-L. arXiv: hep-ph/9203206 (cit. on p. 87).

[136] Q. Bonnefoy. "Heavy fields and the axion quality problem". *Phys. Rev. D* 108.3 (2023), p. 035023. DOI: 10.1103/PhysRevD.108.035023. arXiv: 2212.00102 [hep-ph] (cit. on p. 87).

[137] E. Witten. "Some Properties of O(32) Superstrings". *Phys. Lett. B* 149 (1984), pp. 351–356. DOI: 10.1016/0370-2693(84)90422-2 (cit. on p. 87).

[138] X.-G. Wen and E. Witten. "Electric and Magnetic Charges in Superstring Models". *Nucl. Phys. B* 261 (1985), pp. 651–677. DOI: 10.1016/0550-3213(85)90592-9 (cit. on p. 87).

[139] P. Svrcek and E. Witten. "Axions In String Theory". *JHEP* 06 (2006), p. 051. DOI: 10.1088/1126-6708/2006/06/051. arXiv: hep-th/0605206 (cit. on p. 87).

[140] M. Reece. "Extra-Dimensional Axion Expectations" (June 2024). arXiv: 2406.08543 [hep-ph] (cit. on p. 87).

[141] N. Craig and M. Kongsore. "High-quality axions from higher-form symmetries in extra dimensions". *Phys. Rev. D* 111.1 (2025), p. 015047. DOI: 10.1103/PhysRevD.111.015047. arXiv: 2408.10295 [hep-ph] (cit. on p. 87).

[142] P. Agrawal, M. Nee, and M. Reig. "Axion couplings in heterotic string theory". *JHEP* 02 (2025), p. 188. DOI: 10.1007/JHEP02(2025)188. arXiv: 2410.03820 [hep-ph] (cit. on p. 87).

[143] P. Asadi, S. Homiller, Q. Lu, and M. Reece. "Chiral Nelson-Barr models: Quality and cosmology". *Phys. Rev. D* 107.11 (2023), p. 115012. DOI: 10.1103/PhysRevD.107.115012. arXiv: 2212.03882 [hep-ph] (cit. on pp. 87, 88, 92, 113).

[144] P. F. Perez, C. Murgui, and M. B. Wise. "Automatic Nelson-Barr solutions to the strong CP puzzle". *Phys. Rev. D* 108.1 (2023), p. 015010. DOI: 10.1103/PhysRevD.108.015010. arXiv: 2302.06620 [hep-ph] (cit. on pp. 87, 93–95, 98, 111, 119, 120).

[145] J. Preskill, M. B. Wise, and F. Wilczek. "Cosmology of the Invisible Axion". *Phys. Lett. B* 120 (1983). Ed. by M. A. Srednicki, pp. 127–132. DOI: 10.1016/0370-2693(83)90637-8 (cit. on p. 87).

[146] L. F. Abbott and P. Sikivie. "A Cosmological Bound on the Invisible Axion". *Phys. Lett. B* 120 (1983). Ed. by M. A. Srednicki, pp. 133–136. DOI: 10.1016/0370-2693(83)90638-X (cit. on p. 87).

[147] M. Dine and W. Fischler. "The Not So Harmless Axion". *Phys. Lett. B* 120 (1983). Ed. by M. A. Srednicki, pp. 137–141. DOI: 10.1016/0370-2693(83)90639-1 (cit. on p. 87).

[148] M. Dine, G. Perez, W. Ratzinger, and I. Savoray. "Nelson-Barr ultralight dark matter" (May 2024). arXiv: 2405.06744 [hep-ph] (cit. on p. 88).

[149] V. A. Kuzmin, M. E. Shaposhnikov, and I. I. Tkachev. "Strong CP violation, electroweak baryogenesis, and axionic dark matter". *Phys. Rev. D* 45 (1992), pp. 466–475. DOI: 10.1103/PhysRevD.45.466 (cit. on p. 88).

[150] G. Servant. "Baryogenesis from Strong CP Violation and the QCD Axion". *Phys. Rev. Lett.* 113.17 (2014), p. 171803. DOI: 10.1103/PhysRevLett.113.171803. arXiv: 1407.0030 [hep-ph] (cit. on p. 88).

[151] R. T. Co and K. Harigaya. "Axiogenesis". *Phys. Rev. Lett.* 124.11 (2020), p. 111602. DOI: 10.1103/PhysRevLett.124.111602. arXiv: 1910.02080 [hep-ph] (cit. on p. 88).

[152] T. D. Lee. "A Theory of Spontaneous T Violation". *Phys. Rev. D* 8 (1973). Ed. by G. Feinberg, pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226 (cit. on p. 88).

[153] J. McNamara and M. Reece. "Reflections on Parity Breaking" (Nov. 2022). arXiv: 2212.00039 [hep-th] (cit. on pp. 88, 92, 99).

[154] P. A. Zyla et al. "Review of Particle Physics". *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104 (cit. on pp. 88, 91).

[155] K. Abe et al. "Measurements of neutrino oscillation parameters from the T2K experiment using 3.6×10^{21} protons on target" (Mar. 2023). arXiv: 2303.03222 [hep-ex] (cit. on pp. 88, 96).

[156] M. A. Acero et al. "Improved measurement of neutrino oscillation parameters by the NOvA experiment". *Phys. Rev. D* 106.3 (2022), p. 032004. DOI: 10.1103/PhysRevD.106.032004. arXiv: 2108.08219 [hep-ex] (cit. on pp. 88, 96).

[157] M. A. Acero et al. "Expanding neutrino oscillation parameter measurements in NOvA using a Bayesian approach". *Phys. Rev. D* 110.1 (2024), p. 012005. DOI: 10.1103/PhysRevD.110.012005. arXiv: 2311.07835 [hep-ex] (cit. on pp. 88, 96).

[158] V. Mikola. "Results from the T2K+NOvA Joint Analysis". *EPJ Web Conf.* 312 (2024), p. 02002. DOI: 10.1051/epjconf/202431202002 (cit. on p. 88).

[159] L. Bento, G. C. Branco, and P. A. Parada. "A Minimal model with natural suppression of strong CP violation". *Phys. Lett. B* 267 (1991), pp. 95–99. DOI: 10.1016/0370-2693(91)90530-4 (cit. on pp. 88, 89, 92, 118).

[160] G. C. Branco, P. A. Parada, and M. N. Rebelo. "A Common origin for all CP violations" (July 2003). arXiv: hep-ph/0307119 (cit. on p. 88).

[161] K. Murai and K. Nakayama. "Revisiting the minimal Nelson-Barr model". *JHEP* 11 (2024), p. 098. DOI: 10.1007/JHEP11(2024)098. arXiv: 2407.16202 [hep-ph] (cit. on p. 88).

[162] D. Suematsu. "CP issues in the SM from a viewpoint of spontaneous CP violation". *Phys. Rev. D* 108.9 (2023), p. 095046. DOI: 10.1103/PhysRevD.108.095046. arXiv: 2309.04783 [hep-ph] (cit. on p. 88).

[163] J. M. Alves et al. "Vector-like singlet quarks: A roadmap". *Phys. Rept.* 1057 (2024), pp. 1–69. DOI: 10.1016/j.physrep.2023.12.004. arXiv: 2304.10561 [hep-ph] (cit. on pp. 90, 94).

[164] A. Valenti and L. Vecchi. "The CKM phase and $\overline{\theta}$ in Nelson-Barr models". *JHEP* 07.203 (2021), p. 203. DOI: 10.1007/JHEP07(2021)203. arXiv: 2105.09122 [hep-ph] (cit. on pp. 91, 92).

[165] M. Dine and P. Draper. "Challenges for the Nelson-Barr Mechanism". *JHEP* 08 (2015), p. 132. DOI: 10.1007/JHEP08(2015)132. arXiv: 1506.05433 [hep-ph] (cit. on p. 91).

[166] R. S. Chivukula and H. Georgi. "Composite Technicolor Standard Model". *Phys. Lett. B* 188 (1987), pp. 99–104. DOI: 10.1016/0370-2693(87)90713-1 (cit. on p. 92).

[167] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia. "Minimal flavor violation: An Effective field theory approach". *Nucl. Phys. B* 645 (2002), pp. 155–187. DOI: 10.1016/S0550-3213(02)00836-2. arXiv: hep-ph/0207036 (cit. on p. 92).

[168] G. R. Dvali and G. Senjanovic. "Is there a domain wall problem?" *Phys. Rev. Lett.* 74 (1995), pp. 5178–5181. DOI: 10.1103/PhysRevLett.74.5178. arXiv: hep-ph/9501387 (cit. on p. 93).

[169] G. R. Dvali, A. Melfo, and G. Senjanovic. "Nonrestoration of spontaneously broken P and CP at high temperature". *Phys. Rev. D* 54 (1996), pp. 7857–7866. DOI: 10.1103/PhysRevD.54.7857. arXiv: hep-ph/9601376 (cit. on p. 93).

[170] H. E. Haber and Z. Surujon. "A Group-theoretic Condition for Spontaneous CP Violation". *Phys. Rev. D* 86 (2012), p. 075007. DOI: 10.1103/PhysRevD.86.075007. arXiv: 1201.1730 [hep-ph] (cit. on p. 93).

[171] P. Minkowski. "$\mu \to e\gamma$ at a Rate of One Out of 10^9 Muon Decays?" *Phys. Lett. B* 67 (1977), pp. 421–428. DOI: 10.1016/0370-2693(77)90435-X (cit. on p. 98).

[172] T. Yanagida. "Horizontal gauge symmetry and masses of neutrinos". *Conf. Proc. C* 7902131 (1979). Ed. by O. Sawada and A. Sugamoto, pp. 95–99 (cit. on p. 98).

[173] M. Gell-Mann, P. Ramond, and R. Slansky. "Complex Spinors and Unified Theories". *Conf. Proc. C* 790927 (1979), pp. 315–321. arXiv: 1306.4669 [hep-th] (cit. on p. 98).

[174] R. N. Mohapatra and G. Senjanovic. "Neutrino Mass and Spontaneous Parity Nonconservation". *Phys. Rev. Lett.* 44 (1980), p. 912. DOI: 10.1103/PhysRevLett.44.912 (cit. on p. 98).

[175] S. Davidson, E. Nardi, and Y. Nir. "Leptogenesis". *Phys. Rept.* 466 (2008), pp. 105–177. DOI: 10.1016/j.physrep.2008.06.002. arXiv: 0802.2962 [hep-ph] (cit. on pp. 98, 103, 105).

[176] W. Buchmuller, P. Di Bari, and M. Plumacher. "Leptogenesis for pedestrians". *Annals Phys.* 315 (2005), pp. 305–351. DOI: 10.1016/j.aop.2004.02.003. arXiv: hep-ph/0401240 (cit. on pp. 98, 106, 107, 110).

[177] M. Fukugita and T. Yanagida. "Baryogenesis Without Grand Unification". *Phys. Lett. B* 174 (1986), pp. 45–47. DOI: 10.1016/0370-2693(86)91126-3 (cit. on p. 98).

[178] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov. "On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe". *Phys. Lett. B* 155 (1985), p. 36. DOI: 10.1016/0370-2693(85)91028-7 (cit. on p. 99).

[179] S. Davidson and A. Ibarra. "A Lower bound on the right-handed neutrino mass from leptogenesis". *Phys. Lett. B* 535 (2002), pp. 25–32. DOI: 10.1016/S0370-2693(02)01735-5. arXiv: hep-ph/0202239 (cit. on pp. 99, 103, 125).

[180] W. Grimus and L. Lavoura. "The Seesaw mechanism at arbitrary order: Disentangling the small scale from the large scale". *JHEP* 11 (2000), p. 042. DOI: 10.1088/1126-6708/2000/11/042. arXiv: hep-ph/0008179 (cit. on p. 100).

[181] L. Autonne. *Sur les matrices hypohermitiennes et sur les matrices unitaires*. Annales de l'Université de Lyon, Série I: Sciences, Médecine, Fasc. 38. Lyon and Paris: A. Rey and Librairie Gauthier-Villars, 1915, p. 77 (cit. on p. 101).

[182] T. TAKAGI. "On an Algebraic Problem Related to an Analytic Theorem of Carathéodory and Fejér and on an Allied Theorem of Landau". *Japanese journal of mathematics :transactions and abstracts* 1 (1924), pp. 83–93. DOI: 10.4099/jjm1924.1.0_83 (cit. on p. 101).

[183] M. Aker et al. "Direct neutrino-mass measurement with sub-electronvolt sensitivity". *Nature Phys.* 18.2 (2022), pp. 160–166. DOI: 10.1038/s41567-021-01463-1. arXiv: 2105.08533 [hep-ex] (cit. on p. 101).

[184] A. G. Adame et al. "DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations" (Apr. 2024). arXiv: 2404.03002 [astro-ph.CO] (cit. on p. 101).

[185] A. Pilaftsis and T. E. J. Underwood. "Resonant leptogenesis". *Nucl. Phys. B* 692 (2004), pp. 303–345. DOI: 10.1016/j.nuclphysb.2004.05.029. arXiv: hep-ph/0309342 (cit. on p. 101).

[186] S. Blanchet and P. Di Bari. "Leptogenesis beyond the limit of hierarchical heavy neutrino masses". *JCAP* 06 (2006), p. 023. DOI: 10.1088/1475-7516/2006/06/023. arXiv: hep-ph/0603107 (cit. on p. 102).

[187] S. Blanchet and P. Di Bari. "The minimal scenario of leptogenesis". *New J. Phys.* 14 (2012), p. 125012. DOI: 10.1088/1367-2630/14/12/125012. arXiv: 1211.0512 [hep-ph] (cit. on p. 103).

[188] S. Borsanyi et al. "Calculation of the axion mass based on high-temperature lattice quantum chromodynamics". *Nature* 539.7627 (2016), pp. 69–71. DOI: 10.1038/nature20115. arXiv: 1606.07494 [hep-lat] (cit. on p. 104).

[189] G. F. Giudice, A. Notari, M. Raidal, A. Riotto, and A. Strumia. "Towards a complete theory of thermal leptogenesis in the SM and MSSM". *Nucl. Phys. B* 685 (2004), pp. 89–149. DOI: 10.1016/j.nuclphysb.2004.02.019. arXiv: hep-ph/0310123 (cit. on pp. 104, 105).

[190] M. Plumacher. "Baryogenesis and lepton number violation". *Z. Phys. C* 74 (1997), pp. 549–559. DOI: 10.1007/s002880050418. arXiv: hep-ph/9604229 (cit. on pp. 105, 122).

[191] P. Fileviez Perez, C. Murgui, and A. D. Plascencia. “Baryogenesis via leptogenesis: Spontaneous B and L violation”. *Phys. Rev. D* 104.5 (2021), p. 055007. DOI: 10.1103/PhysRevD.104.055007. arXiv: 2103.13397 [hep-ph] (cit. on pp. 105, 122).

[192] R. Barbieri, P. Creminelli, A. Strumia, and N. Tetradis. “Baryogenesis through leptogenesis”. *Nucl. Phys. B* 575 (2000), pp. 61–77. DOI: 10.1016/S0550-3213(00)00011-0. arXiv: hep-ph/9911315 (cit. on p. 105).

[193] A. Abada et al. “Flavour Matters in Leptogenesis”. *JHEP* 09 (2006), p. 010. DOI: 10.1088/1126-6708/2006/09/010. arXiv: hep-ph/0605281 (cit. on p. 105).

[194] E. Nardi, Y. Nir, E. Roulet, and J. Racker. “The Importance of flavor in leptogenesis”. *JHEP* 01 (2006), p. 164. DOI: 10.1088/1126-6708/2006/01/164. arXiv: hep-ph/0601084 (cit. on p. 105).

[195] S. Davidson. “Flavoured Leptogenesis”. *12th International Workshop on Neutrinos Telescopes: Twenty Years after the Supernova 1987A Neutrino Bursts Discovery*. May 2007, pp. 531–545. arXiv: 0705.1590 [hep-ph] (cit. on p. 105).

[196] S. Blanchet and P. Di Bari. “Flavor effects on leptogenesis predictions”. *JCAP* 03 (2007), p. 018. DOI: 10.1088/1475-7516/2007/03/018. arXiv: hep-ph/0607330 (cit. on pp. 105, 108).

[197] F. X. Josse-Michaux and A. Abada. “Study of flavour dependencies in leptogenesis”. *JCAP* 10 (2007), p. 009. DOI: 10.1088/1475-7516/2007/10/009. arXiv: hep-ph/0703084 (cit. on p. 105).

[198] W. Buchmuller and M. Plumacher. “Spectator processes and baryogenesis”. *Phys. Lett. B* 511 (2001), pp. 74–76. DOI: 10.1016/S0370-2693(01)00614-1. arXiv: hep-ph/0104189 (cit. on p. 105).

[199] E. Nardi, Y. Nir, J. Racker, and E. Roulet. “On Higgs and sphaleron effects during the leptogenesis era”. *JHEP* 01 (2006), p. 068. DOI: 10.1088/1126-6708/2006/01/068. arXiv: hep-ph/0512052 (cit. on p. 106).

[200] J. A. Harvey and M. S. Turner. “Cosmological baryon and lepton number in the presence of electroweak fermion number violation”. *Phys. Rev. D* 42 (1990), pp. 3344–3349. DOI: 10.1103/PhysRevD.42.3344 (cit. on p. 106).

[201] W. Buchmuller, P. Di Bari, and M. Plumacher. “The Neutrino mass window for baryogenesis”. *Nucl. Phys. B* 665 (2003), pp. 445–468. DOI: 10.1016/S0550-3213(03)00449-8. arXiv: hep-ph/0302092 (cit. on pp. 110, 125).

[202] N. J. Ayres et al. “The design of the n2EDM experiment: nEDM Collaboration”. *Eur. Phys. J. C* 81.6 (2021), p. 512. DOI: 10.1140/epjc/s10052-021-09298-z. arXiv: 2101.08730 [physics.ins-det] (cit. on pp. 115, 116).

[203] M. Athanasakis-Kaklamanakis et al. "Community input to the European Strategy on particle physics: Searches for Permanent Electric Dipole Moments" (May 2025). arXiv: 2505.22281 [hep-ex] (cit. on pp. 114–116).

[204] M. W. Ahmed et al. "A New Cryogenic Apparatus to Search for the Neutron Electric Dipole Moment". *JINST* 14.11 (2019), P11017. DOI: 10.1088/1748-0221/14/11/P11017. arXiv: 1908.09937 [physics.ins-det] (cit. on pp. 115, 116).

[205] R. Matsumiya et al. "The Precision nEDM Measurement with UltraCold Neutrons at TRIUMF". *JPS Conf. Proc.* 37 (2022), p. 020701. DOI: 10.7566/JPSCP.37.020701. arXiv: 2207.09880 [physics.ins-det] (cit. on pp. 115, 116).

[206] J. Alexander et al. "The storage ring proton EDM experiment" (Apr. 2022). arXiv: 2205.00830 [hep-ph] (cit. on p. 115).

[207] V. Baluni. "CP-nonconserving effects in quantum chromodynamics". *Phys. Rev. D* 19 (7 1979), pp. 2227–2230. DOI: 10.1103/PhysRevD.19.2227. URL: https://link.aps.org/doi/10.1103/PhysRevD.19.2227 (cit. on p. 114).

[208] R. J. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten. "Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics". *Phys. Lett. B* 88 (1979). [Erratum: Phys.Lett.B 91, 487 (1980)], p. 123. DOI: 10.1016/0370-2693(79)90128-X (cit. on p. 114).

[209] J. Alexander et al. "The storage ring proton EDM experiment" (Apr. 2022). arXiv: 2205.00830 [hep-ph] (cit. on p. 116).

[210] F. Abusaif et al. *Storage ring to search for electricdipole moments of charged particles: Feasibility study*. Geneva: CERN, June 2021. DOI: 10.23731/CYRM-2021-003. arXiv: 1912.07881 [hep-ex] (cit. on p. 116).

[211] P. Fileviez Perez and M. B. Wise. "Breaking Local Baryon and Lepton Number at the TeV Scale". *JHEP* 08 (2011), p. 068. DOI: 10.1007/JHEP08(2011)068. arXiv: 1106.0343 [hep-ph] (cit. on p. 120).

[212] M. Duerr, P. Fileviez Perez, and M. B. Wise. "Gauge Theory for Baryon and Lepton Numbers with Leptoquarks". *Phys. Rev. Lett.* 110 (2013), p. 231801. DOI: 10.1103/PhysRevLett.110.231801. arXiv: 1304.0576 [hep-ph] (cit. on p. 121).

[213] A. Takenaka et al. "Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ with an enlarged fiducial volume in Super-Kamiokande I-IV". *Phys. Rev. D* 102.11 (2020), p. 112011. DOI: 10.1103/PhysRevD.102.112011. arXiv: 2010.16098 [hep-ex] (cit. on p. 121).

[214] J. A. Casas and A. Ibarra. "Oscillating neutrinos and $\mu \to e, \gamma$". *Nucl. Phys. B* 618 (2001), pp. 171–204. DOI: 10.1016/S0550-3213(01)00475-8. arXiv: hep-ph/0103065 (cit. on p. 124).

[215] M.-T. Eisele. “Leptogenesis with many neutrinos”. *Phys. Rev. D* 77 (2008), p. 043510. DOI: 10.1103/PhysRevD.77.043510. arXiv: 0706.0200 [hep-ph] (cit. on p. 125).