Probing Undiscovered Particles with Theory and Data-Driven Tools

HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Physics

have examined a dissertation entitled

Probing Undiscovered Particles with Theory and Data-Driven Tools

presented by Katherine Fraser

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Matthew Reece

Signature _____

Typed name: Prof. Matthew Schwartz

Signature _____

Typed name: Prof. Cora Dvorkin

Date: April 23, 2024

Probing Undiscovered Particles with Theory and Data-Driven Tools

a dissertation presented
by
Katherine Fraser
to
The Department of Physics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Physics

Harvard University
Cambridge, Massachusetts
April 23 2024

Probing Undiscovered Particles with Theory and Data-Driven Tools

Abstract

The Standard Model has been precisely tested by a plethora of experiments and has proved extremely successful at describing the fundamental interactions of subatomic particles. Despite this, there are numerous exciting questions motivating the existence of additional physics beyond the Standard Model. In this dissertation, we study a variety of theoretical and data-driven tools for discovering this physics. For theoretical tools, we explore various effective field theories, including those describing: axions mixing with other axions or non-compact scalars, axion interactions with magnetic monopoles, axion strings interacting with massive fermions, CP violating Higgs portal dark matter, scalar triplet and singlet-doublet models as solutions to the CDF-II W mass anomaly, and modifications to Froggatt-Nielsen models for explaining the hierarchy between different flavors of Standard Model Yukawa couplings. Additionally, we investigate machine learning based data-driven tools for precision Top-quark mass measurement and for anomaly detection.

Contents

Previously Published Components

Most of this dissertation comprises slightly modified versions of work that was originally published as journal articles. This includes:

Chapter 2, which was published as [1]:

K. Fraser and M. Reece,
Axion Periodicity and Coupling Quantization in the Presence of Mixing,
JHEP 05 (2020) 066, arXiv: 1910.11349,

Chapter 3, which was published as [2]:

J. Fan, K. Fraser, M. Reece and J. Stout,
Axion Mass from Magnetic Monopole Loops,
Phys. Rev. Lett. 127 (2021) 131602, arXiv: 2105.09950,

Chapter 4 and Appendix A, which were published as [3]:

H. Bagherian, K. Fraser, S. Homiller and J. Stout,
Zero Modes of Massive Fermions Delocalize from Axion Strings,
Submitted to JHEP, arXiv: 1910.11349,

Chapter 5 and Appendix B, which were published as [4]:

Katherine Fraser, Aditya Parikh and Weishuang Linda Xu,
A Closer Look at CP-Violating Higgs Portal Dark Matter as a Candidate for the GCE,
JHEP 03, 123 (2021), arXiv: 2010.15129,

Chapter 6 and Appendices C and D, which were published as [5]:

P. Asadi, C. Cesarotti, K. Fraser, S. Homiller and A. Parikh,
Oblique lessons from the W-mass measurement at CDF II,
Phys. Rev. D 108 (2023) 055026, arXiv: 2204.05283,

Chapter 7 and Appendix E, which were published as [6]:

P. Asadi, A. Bhattacharya, K. Fraser, S. Homiller and A. Parikh,
Wrinkles in the Froggatt-Nielsen mechanism and flavorful new physics,
JHEP 10 (2023) 069, arXiv: 2308.01340,

Chapter 8 and part of Appendix F, which were published as [7]:

F. Flesher, K. Fraser, C. Hutchison, B. Ostdiek and M. D. Schwartz,
Parameter inference from event ensembles and the top-quark mass,
JHEP 09 (2021) 058, arXiv: 2011.04666,

Chapter 9 and part of Appendix F, which were published as [8]:

K. Fraser, S. Homiller, R.K. Mishra, B. Ostdiek and M.D. Schwartz,
Challenges for unsupervised anomaly detection in particle physics,
JHEP 03 (2022) 066, arXiv: 2110.06948.

To my parents, who have always been my greatest supporters.

Acknowledgments

There are so many people who have been invaluable to my success during my PhD. First and foremost, I need to thank my advisor Matt Reece. It's exceedingly rare to find someone who is both so brilliant and down to earth, and I feel enormously lucky to have had the opportunity to be his student. Despite his incredible intuition and extensive knowledge of the literature, Matt always treats even his youngest students as collaborators and answers the smartest version of any question. I have learned so much from him: how to think about research and ask questions, how to communicate clearly, how to effectively prioritize and balance projects, and gained an understanding of countless physics topics. I am also grateful for his focus on setting me up for long term success. From giving encouraging and pragmatic advice, to organizing reading groups and journal clubs, to encouraging me to take on research projects with other faculty and younger members of the group, Matt has always prioritized preparing me to be an independent researcher.

I also need to thank Matt Schwartz, my longtime collaborator, undergraduate research advisor, and committee member. I've learned a great deal from working with him: how to brainstorm effectively, how to break down projects into manageable pieces, how to revise a paper, and the patience to tackle difficult problems. I aspire to develop the type of intuition he has for knowing so early on when a project is worthwhile. My first exposure to QFT was his class and it has had a pivotal influence on how I think about the subject, and I would not be where I am today without my first project with him. Matt has been a wonderful mentor, and I am grateful for how he has helped me build connections within the community and passed along opportunities. I appreciate how he is always willing to make time to chat, whether it be about a research project, for advice, or just to discuss interesting physics.

I also need to thank the many other collaborators I've had the pleasure of working with: Pouya Asadi, Hengameh Bagherian, Arindam Bhattacharya, Cari Cesarotti, JiJi Fan, Sam Homiller, Qianshu Lu, Rashmish Mishra, Bryan Ostdiek, Aditya Parikh, John Stout, and Linda Xu. Collaboration with many of them is a large part of what has made my PhD both enjoyable and productive. I am especially grateful for Aditya, Linda, and Hengameh; without the ability to discuss any and all confusions (no matter how long ago we should have learned the answer) with them, graduate school would have been a very different experience. Also, I have grown and learned so much from the opportunity to work with Sam, John, Rashmish, and Pouya, and I appreciate their patience with my endless questions. I am also thankful for the many other

current and former members of Matt's group: Prateek Agrawal, Daniel Aloni, Junyi Cheng, Francis-Yan Cyr-Racine, Rabia Husain, Qianshu Lu, Noah McNeal, David Pinner, and Motoo Suzuki. I never expected to have so much fun (or be part of a slack filled with so many memes) during my PhD, and I am so appreciative of all the group members who made that possible.

Next, I am grateful to my graduate school cohort. From first year game nights to brunches to our puppet show, I've really enjoyed getting to know my classmates, especially: Alex Atanasov, Alek Bedroya, Taylor Contreras, Dominic Chang, Nicholas Cuccia, Anne Fortman, Soumya Ghosh, Kara Hartig, Ella King, Madelyn Leembruggen, Caroline Martin, Arvind Srinivasan, Yanting Teng, and Arthur Tsang. I'm especially thankful for the opportunity to work with Caroline, Ella, Justina Yang, Madelyn, Taylor, and Yan on various different outreach projects, including Women in Physics and WOWSTEM. Finally, this list would not be complete without acknowledging my friendship with Anne, whose desk was next to mine G1 year.

Additionally, I am thankful for all of the other members of the department that I have had the pleasure of interacting with and learning from. In the high energy theory group, this includes: Nick DePorzio, Aurelien Dersy, Cora Dvorkin, Naomi Gendler, Howard Georgi, Hofie Hannesdottir, Hayden Lee, Jackie Lodman, Jake McNamara, Siddharth Mishra-Sharma, Julian Munoz, Michael Nee, Lisa Randall, Matthew Strassler, Cumrun Vafa, Georgios Valogiannis, Xiaoyuan Zhang. In IAIFI and at MIT, this includes: Rakib Gambir, Sarah Geller, Gaia Grosso, Jim Halverson, Phil Harris, Marisa LaFleur, Sang Eon Park, Wenzer Qin, Tracy Slatyer, Zhiquan Sun, Yitian Sun, and Jesse Thaler. I am also grateful to Jesse for being one of my letter writers, and to Cora for being on my committee and helping me learn about neutrino cosmology. Additionally, I should thank some others who played a pivotal role in setting me on my current path during my time as an undergraduate: Masahiro Morii, for giving me my first research project and always finding the time to chat with me about it, Jenny Hoffman, for being an incredible role model and being so supportive as I applied to graduate school, all those who were part of the Harvard ATLAS group from 2016-2017, for creating a great environment to do my first research project in, my QFT problem set group (Adam Ehrenberg, Adam Frim, Ben Sorscher, and Sebastian Wagner-Carena), for their positivity and clever insights during our long nights working together, and Girma Hailu, for being the first one to encourage me to try a research project, back when didn't know if I was cut out for this.

I would also like to thank the many administrators who helped with WiP, and also make the graduate program and the department run from day-to-day: Jacob Barandes, Hannah Belcher, Gerry Byrne, Lisa Cacciabaudo, Jade Chin, Erica Colwell, Carol Davis, Silke Exner, Rae Gilchrist, Morgan Holly, Elise Krims, Mary McCarthy, and Jennifer Pollock. Carol and Lisa in particular have always gone so far above and beyond what they had to do and are so appreciated.

I am also grateful for the many others in the broader high energy physics community who I have met

at conferences, workshops and schools. I am especially grateful for those who attended TASI with me: Jared Barron, Arushi Bodas, Zach Bogorad, Anne-Katherine Burns, Bryce Cyr, Tianji Cai, Max Fieg, Margarita Gavrilova, Cara Giovanetti, Anna Hallin, Ameen Ismail, Melissa Joseph, Rachel Nguyen, Yujin Park, Bethany Suter, and Kevin Zhou.

Finally, I am grateful to my family and friends for their support during these last six years. My parents in particular have always gone and above and beyond to support me. I also appreciate the support of my brothers James and Joseph, my future sister-in-law Bella, and my whole extended family. Additionally, I want to thank a few friends who have always been around to cheer me up through the tougher parts of my PhD: Pavitra Chari, Katie Kixmoeller, Brittany Logan, Emily Tiberi, Melissa Wong, and Amy Zheng. I deeply appreciate their support.

1

Introduction and Literature Review

1.1 Introduction

The Standard Model is a remarkably successful description of the fundamental interactions of subatomic particles. Since it was first theorized around 50 years ago, it has been precisely tested at a variety of experiments. The largest and most general purpose experiments are colliders, including the Large Hadron Collider (LHC), the Large Electron–Positron Collider (LEP), and the Tevatron. However, a variety of other smaller scale experiments, including fixed-target experiments, beam dump experiments, cherenkov detectors, time projection chambers, haloscopes, helioscopes, and many others, have also been used to precisely measure Standard Model parameters, and to set limits on new physics beyond the Standard Model (BSM).

The Standard Model contains three generations each of leptons and quarks, as well as gauge bosons (the photon, W and Z bosons, and gluons) and the Higgs boson, all transforming in representations of the Standard Model gauge group[1] $G = SU(3)_C \times SU(2)_W \times U(1)_Y$. Out of these, the final piece of the Standard Model to be discovered was the Higgs boson, which was discovered in 2012 at the LHC. While this discovery was exciting, it was not surprising – a discovery was anticipated due to a "no-lose theorem," where longitudinal gauge boson scattering starts to violate perturbative unitary bounds at scales accessible to the

[1] Technically this is an oversimplification; there are actually four possible options for the Standard Model gauge group. The other three are a quotient of G by $\mathbb{Z}_2$, $\mathbb{Z}_3$, or $\mathbb{Z}_6$, respectively [9].

LHC [10]. Since this discovery, however, the roadmap to find additional particles is less clear. Prior to the first operation of the LHC, there was a sociological predisposition in much of the particle physics community to expect that the LHC would find BSM physics in the form of low-scale supersymmetry (SUSY) shortly after it found the Higgs. However, these hopes have gone unfulfilled by the LHC.

Despite this, there is reason to be optimistic that physics beyond the Standard Model (BSM) is still waiting to be discovered. Many exciting questions remain unanswered. Some, such as the fundamental nature of dark matter and dark energy, the source of the baryon anti-baryon asymmetry in the early universe, and the origin of neutrino masses, are tied to experimentally measured quantities that are inexplicable within the Standard Model and are guaranteed to have a solution beyond it. Others, including several problems with fine-tuning and explaining large hierarchies (for example, the hierarchy problem, the Strong CP problem, the cosmological constant problem, and the hierarchy between the size of the masses of different fermion generations), are more theoretical. While it is in principle possible that there is no fundamental answer to these more theoretical questions and nature is just tuned by chance, this explanation is unsatisfying and does not align with our current understanding of small numbers and symmetries. Still other questions are more contested experimental discrepancies, such as the CDF-II W mass measurement [11], the value of R_D [12–18], and Hubble tension [19–23], which disagree with the Standard Model prediction but could be due to either new BSM physics or mismodelled systematic errors. In addition to these questions which can motivate new low scale physics, there are also questions which must be answered by physics at much higher scales, such as the correct underlying characterization of quantum gravity or the detailed particle content of the underlying mechanism (inflation [24–28] being the most compelling currently known possibility) which makes the Cosmic Microwave Background (CMB) isotropic. Extended solitonic objects like topological defects also have the potential to generate interesting effects and motivate further study of BSM physics. More details regarding some of these questions and potential answers will be discussed in §1.2.

Without a clear theoretical prior about which of these questions will first yield a discovery, the most sensible course of action is a simultaneous exploration of a wide variety of options. However, testing such a wide variety of theories will be challenging. Making progress will require innovative approaches. In addition to experimental progress, the development of a variety of theoretical and data-based tools will be necessary. This includes effective field theory (EFT) based theoretical tools and machine learning (ML) based data-driven ones.

ML tools in particular are necessary to sift through the enormous quantities of data produced by experiments, to process data quickly, to find subtle patterns in sparse data, and to fully utilize and synthesize the low level information that detectors collect. In the last decade in particular, the use of these tools in particle physics has advanced tremendously from the previously state-of-the-art Boosted Decision Trees [29]. This

is partly driven by deploying similar advances from industry, and partly driven by customized architectures adapted for particle physics tasks. Many different supervised and unsupervised architectures[2] have been utilized, including autoencoders, convolutional networks, deep sets, diffusion models, fully-connected dense networks, generative adversarial networks, graph networks, normalizing flows, recurrent networks, recursive networks, and transformers. It is outside of the scope of this dissertation to discuss the details of each of these architectures, but references describing them can be found in [30].

These architectures have been applied to a variety of applications. Many of these tools were studied and developed in simplified proof of concept scenarios. One example is the many different types of classifiers which have been explored, including both a wide variety of jet taggers [31–97] and BSM signal versus background classifiers [38, 48, 98–109]. Another is anomaly detection algorithms [8, 108–207], which attempt to design searches which are less model dependent, and will be discussed more in §9. Further examples include tools for reconstruction and calibration [208–237], decorrelation [36,173,238–250], more efficient simulations [251–313], parameter estimation [7, 314–338] (an example of which is discussed in §8), pile-up mitigation [339–345], uncertainty quantification and robustness [32, 213, 249, 346–364], and unfolding [365–383].

On the experimental side, progress has been sufficient that essentially every state of the art experimental analysis uses ML tools at some point in the data collection or analysis pipeline. In addition to analyses which use ML for various smaller pieces (for example, in triggering, calibration, tagging, reconstruction, and simulation), other analyses are starting to be completed that deploy ML as an indispensable part of the underlying design of the analysis, such as unsupervised anomaly detection searches [384–387]. More discussion of these tools can be found in the reviews [253, 388–400]. Despite these substantial advances, there is still room for the development of additional ML tools and also deployment of additional existing tools to experiment.

Even if we do not discover new particles in the near future, innovations in experimental, theoretical, and data-based tools are also vital to precisely measuring Standard Model parameters. These precise measurements are both interesting questions in their own right, and also have the potential to indirectly characterize sources of BSM physics.

The rest of this dissertation is organized as follows. First, in the remainder of this introduction, we provide more detailed motivation for the existence of BSM physics. Then in §2, §3, and §4, we discuss some results about and following from the periodicity of axion fields, including consequences of axion interactions with topological defects. Next, we discuss several BSM models which are motivated by recent experimental results. In §5, we consider models of CP violating Higgs portal majorana dark matter and discuss the consistency of these models with various experimental constraints, including the Galactic Center Excess

[2]Supervised approaches refer to cases where labels are available; in the unsupervised case labels are not used. Weakly- and Semi-supervised approaches which use incomplete or imperfect labels also exist.

(GCE) of gamma rays. In §6, we study the implications of the recent CDF-II measurement of the W-mass by performing electroweak fits and exploring the viability of a scalar triplet model or the singlet-doublet model for explaining the measurement. In §7, we show how BSM physics can be embeddded in nontrivial ways in Froggatt Nielsen (FN) models (models which are a possible explanation of the flavor hierarchy between SM couplings), and show how the recent Belle II BR($B \to K\nu\nu$) measurement can be interpreted in the context of these models. Finally, in §8 and §9, we investigate ML tools which can be used to precisely measure the top mass and to perform anomaly detection searches for new physics, respectively. We conclude in §10. More details about some calculations are also provided in appendices A through F.

1.2 Hints of BSM Physics

Many different questions motivate the search for BSM physics. These questions range from experimental problems (e.g. dark matter, the origin of neutrino masses) where measured quantities cannot be explained by the Standard Model, to theoretical ones (i.e. fine tuning problems, the nature of quantum gravity).

In this section, we review several different motivations for the existence of BSM physics in more detail. Since this is an extremely broad subject, we cannot hope to do justice to all of these topics with thorough explanations here. Instead, we will provide brief introductions to a subset of topics, primarily those which will be relevant for later chapters in this dissertation. These topics include the hierarchy problem, weakly interacting massive particle (WIMP) dark matter, axions and the strong CP problem, aesthetic concerns within the Standard Model (including the flavor hierarchy), and current experimental anomalies. Some other motivations for BSM physics, including models of inflation, the origin of neutrino masses, the cosmological constant problem and the nature of dark energy, and the origin of baryon anti-baryon asymmetry can be reviewed in [401–410].

1.2.1 The Hierarchy Problem

One of the most well studied motivations for BSM physics is the hierarchy problem. Here, we provide a very brief introduction to the hierarchy problem and its solutions. A more thorough review can be found in [409, 411] and references therein.

At the heart of the hierarchy problem is the question "Why is the Standard Model Higgs so light?" Two aspects of this question are worth emphasizing. First, the hierarchy problem applies to the Higgs in particular because it is a scalar, and scalars get large loop corrections to their mass from heavy scales. This is true regardless of regulator. For example, using a hard cutoff we will see the mass gets corrections that depend quadratically on the cutoff, whereas using $\overline{\text{MS}}$ the scalar mass corrections depend quadratically on the heavy matching scale. Even though we have not observed new BSM particles heavier than the Standard

Model, we know that eventually the Standard Model has to break down as an EFT (either at the Planck scale due to quantum gravity or below it if additional new physics is discovered), so corrections from these high scales matter. In order to get a small Higgs mass despite these large corrections, these corrections must be tuned to cancel very precisely. Second, since the Higgs mass is a dimensionful quantity, we need a scale to compare it to: the important question is really why the Higgs mass is so much lighter than the Planck mass.

The hierarchy problem can be made more precise through the idea of naturalness. In its simplest form, naturalness is the idea dimensionless numbers should be approximately $\mathcal{O}(1)$, and we should not expect very small or very large numbers without justification; accepted justifications are typically related to the existence of symmetries. More rigorously, naturalness can be characterized in a couple of different ways. 't Hooft naturalness is the statement that it is natural for a parameter to be small if a symmetry would be restored when it is set to zero. Technical naturalness requires that values we set to be small do not receive large loop corrections, but does not explain the existence of small quantities in the first place. Another valuable question about naturalness is how to quantify it precisely. While it is impossible to know exactly how strictly satisfied these criteria should be by nature, several interesting measures have been proposed for benchmarking different models [412–415].

The skeptical reader might ask why we expect a natural theory in the first place. The strongest reason to expect naturalness in the case of the Higgs mass is that other small parameters in the Standard Model are natural. For example, the proton mass scale arises naturally out of confinement and dimensional transmutation. Other examples include: pion masses, which are small because pions are pseudo-Nambu-Goldstone bosons for chiral symmetry breaking; fermion masses, which are protected by chiral symmetry; and the Standard Model flavor hierarchy, which is 't Hooft natural since the flavor symmetries are only violated by the yukawas. Finally, mass differences such as those between the charged and neutral pion and between the K_L^0 and K_S^0 are described by the expected quadratic sensitivity to heavy mass scales (the ρ meson and charm quark, respectively).

Numerous solutions to the hierarchy problem have been proposed, many of which rely on the existence of new particles at or near the weak scale. Historically one of the most popular solutions is supersymmetry (SUSY) [416–419]. SUSY is an extension of the Lorentz algebra by fermionic generators Q_α and $Q_{\dot{\alpha}}^\dagger$ called supercharges. Because they are fermionic and therefore obey anticommutation relations, these supercharges transform fermions into bosons and vice versa. Because of this, bosons and fermions always come in pairs with the same quantum numbers in SUSY[3], and they are organized together into supermultiplets.

[3]The exception to this is R-charge, under which fields in the same supermultiplet can have different charges. This is allowed because the R-symmetry generator does not commute with the supercharges, but still leaves the SUSY algebra invariant.

Theories can also have more than one supersymmetry. In four dimensions, there can be up to $\mathcal{N} = 8$ supercharges if gravity is included and up to $\mathcal{N} = 4$ if it is not. However, supersymmetric extensions of the Standard Model that are actively searched for at experiments only have $\mathcal{N} = 1$ supersymmetry because only SUSY with a single supercharge allows chiral matter like the Standard Model fermions. The simplest example SUSY model that has been searched for is the Minimal Supersymmetric Standard Model (MSSM), but there are many other possibilities.

The reason SUSY would solve the hierarchy problem is simple: in a theory with perfect SUSY, all large bosonic corrections to the Higgs mass would come with fermionic corrections with equal magnitude but opposite sign. However, we know that in the real world superpartners don't exist at exactly the same scale as their Standard Model counterparts, or we would have seen them. This means that SUSY must be softly broken. We will not discuss the details of the physics responsible for soft SUSY breaking here, but there are several different mechanisms that can be responsible. See [416, 420, 421] for reviews. Regardless of the mechanism responsible for it, soft SUSY breaking prevents a perfect cancellation between bosonic and fermionic mass corrections. Nonetheless, approximate cancellation with substantially less tuning than the Standard Model is still retained.

SUSY is also useful for far more than solving the hierarchy problem. In fact, SUSY was not even originally proposed in the context of the hierarchy problem; instead, early studies of SUSY were motivated by understanding the maximum amount of symmetry a theory could have.[4] If some form of SUSY were discovered at accessible energies, there would also be a natural dark matter candidate in the form of the lightest supersymmetric particle (also see §1.2.2). Additionally, unlike the Standard Model itself, unification at GUT scales is still a possibility in supersymmetric theories. Even if the universe is not described by low-scale SUSY, SUSY more generally is still a very useful theoretical tool: many quantities can be calculated in SUSY theories that are too difficult to calculate otherwise, and the consistent quantum gravity theories that we know of are supersymmetric string theories.

Another class of solutions to the hierarchy problem are composite Higgs models. These rely on using strong dynamics to spontaneously break a symmetry, and are motivated in spirit by the light pions from chiral symmetry breaking. These are similar to earlier Technicolor models [424, 425], which introduced a new strongly coupled gauge sector which was broken by a condensate of fermions. In the case of Technicolor, the same dynamics that breaks the strongly coupled sector also breaks $\mathrm{SU(2)_W} \times \mathrm{U(1)_Y}$ and gives mass to the W and Z bosons. However, the original form of technicolor theories were long ago ruled out – they generically generate flavor changing neutral currents, do not obey electroweak precision constraints, do not

[4] Supersymmetry was proposed as an exception to the Coleman-Mandula theorem [422], which states that the maximal amount of allowed symmetry in a QFT (with a valid S-matrix) is the direct product of Poincare symmetry and an internal symmetry. It was Haag, Lopuszanski, and Sohnius who showed that the Poincare algebra could be extended to the supersymmetry algebra by allowing fermionic generators [423].

have a Higgs boson, and have many (unobserved) light resonances around the confinement scale. After technicolor, composite Higgs models were developed [426] (also see [427, 428]). In these models, the Higgs is still generated from strong dynamics breaking a symmetry, but now that symmetry is global, so the Higgs is a pseudo-Nambu-Goldstone (pNGB) boson. Specifically, these models separate the scale of confinement and electroweak symmetry breaking, so that the pNGB is uneaten between the two. The original composite Higgs proposal is now strongly constrained by Higgs coupling measurements to require substantial tuning, but other variations allow more flexibility [429, 430].

Models with extra dimensions can also lead to interesting solutions to the hierarchy problem. In these models, the separation between the Planck scale and the electroweak scale is due to either a volume factor or a warping factor, or a combination of both. In large extra dimension models [431, 432], the higher dimensional Planck scale is not too far above the TeV scale, and the four dimensional Planck scale is large because it is the higher dimensional scale times the volume of compactified extra dimensions. In Randall-Sundrum models [433, 434], the Standard Model lives on a brane in five dimensional space. The five dimensional space is warped, so that the four dimensional metric is multiplied by an exponentially suppressed function of the fifth dimension, called the 'warp factor'. In RS models, it is this exponential suppression from the warp factor that generates the hierarchy between the electroweak and Planck scales. More recently, a third class of extra dimensional models utilizing a linear dilaton geometry has also been proposed [435–437]. In these models, the hierarchy is generated by the combination of an exponentially large separation between branes and a linear dependence of the metric on the extra dimension. See [409, 411, 438] for additional review of extra dimensional models.

A qualitatively different solution to the hierarchy problem comes in the form of anthropic arguments. The central idea of anthropic arguments is that there exists a multiplicity of different vacua, but only some of them are conducive to life forming. The fact that we are here to make predictions, therefore, selects those vacua. If some parameters are required to be small in the vacua that are compatible with life, then this argument requires those parameters should be small in our universe. Anthropic arguments were first made precise in the context of the cosmological constant problem [439], but have since been applied to the hierarchy problem through the Atomic principle [440]. The atomic principle states that in order for life to form, neutrons cannot decay in bound systems of nuclei, which places a bound on the Standard Model Higgs vacuum expectation value. However, this argument is not airtight; loopholes include the case where other Standard Model parameters besides the Higgs vev are allowed to vary [441].

Aside from anthropic arguments, these solutions all predict many particles starting to appear around the Higgs mass. The lack of observation of these extra particles at the LHC has started to call into question many of these solutions. Because of this, there has been a renewed focus on developing qualitatively different

solutions to the hierarchy problem. Often, these solutions can explain the hierarchy between the Higgs mass and some intermediate scale Λ that is above LHC accessibility but far below the Planck scale, but require one of the older solutions to take effect above this intermediate scale to explain the Hierarchy between Λ and M_{Pl}. This is sometimes referred to as the "Little Hierarchy Problem."

One type of solution to the "Little Hierarchy Problem" are neutral naturalness [442] approaches which rely on discrete symmetries. In these models, there are additional mirror copies of the Standard Model related to the Standard Model by a discrete symmetry. The additional Higgses have a potential with an accidental symmetry that mixes them, and the Standard Model Higgs is a pseudo-NGB of that accidental broken symmetry. By choosing the potential so that one loop corrections to the Higgs mass respect the accidental symmetry, these models can postpone the scale at which the hierarchy problem needs to be explained by one loop factor to $\sim \mathcal{O}(10\text{TeV})$. The simplest example of this type of models is the Twin Higgs model [443]. Another approach is the relaxion [444]. These models involve adding an axion like particle which couples to gluons through the usual $\phi G\tilde{G}$ coupling as well as to the Standard Model Higgs through $g\phi|H|^2$, where g is a dimensionful shift symmetry breaking coupling. In this model, as the universe evolves, the Higgs rolls down its potential. Once QCD confines, the potential gets instanton contributions whose amplitude is proportional to the Higgs vev, which stops the Higgs from rolling further and allows for a light Higgs. Typically relaxion-inflaton couplings are engineered to take advantage of slow roll inflation, and to get the right value of the CP violating phase $\overline{\theta}$. Since the relaxion solution was proposed, there have also been a variety of other cosmological approaches to solving the hierarchy problem which we will not discuss here [445–448]. A third solution is N-naturalness. Like anthropic solutions, N-naturalness [449] is another probabilistic approach. The difference is that in this case, there are N copies of the Standard Model in the same universe, and there is a dynamic mechanism (typically during reheating) that selects the sector that looks like ours to have virtually all of the energy density of the universe. The potentially unappealing part of this model is that it requires an exponentially large number of copies of the Standard Model since the cutoff scales as $\sqrt{N}$; even extending up to 10 TeV requires 10^4 copies, and solving the full hierarchy problem requires many more (and also lowers the scale where gravity becomes strongly coupled).

Finally, in addition to the recent solutions to the "Little Hierarchy Problem," recent developments also instill hope for qualitatively different hierarchy problem solutions which do not lead to new physics at the weak scale. For example, there has been recent progress deriving constraints from quantum gravity by utilizing Swampland conjectures like the Weak Gravity Conjecture [450–455].[5] Another area where there is

[5]The Swampland program uses constraints from quantum gravity to impose constrains on low energy EFTs that aren't apparent from the EFT point of view. The weak gravity conjecture is an example of a Swampland Conjecture. There are several different versions of it; the original (electric) weak gravity conjecture [450] says that in an abelian gauge theory coupled to gravity, there must exist a state of charge q and mass m such that $qg > m/M_{Pl}$. See [456,457] for reviews.

hope for additional progress is finding generalized symmetries (symmetries where the charged operators are not just particles, but can have other dimensions) [458] which might make the Higgs mass 't Hooft natural.

1.2.2 Dark Matter

Astrophysical and cosmological evidence from a variety of scales necessitates the existence of dark matter. The earliest hints of missing mass came almost a century ago from discrepancies between the potential calculated from velocities due to Doppler shift and the potential calculated from the amount of luminous matter directly observed in galaxies and clusters of galaxies [459, 460]. Since then, additional evidence at galactic, cluster, and cosmological scales has been uncovered. In addition to the famous galactic rotation curves first measured by Vera Rubin and collaborators [461], evidence at galactic scales includes the stability of dwarf spheroidal galaxies against tidal forces [462], strong gravitational lensing (and its weak modulation) by elliptical galaxies [463–465], weak gravitational lensing of galaxies by foreground structure[6] [467], and many other examples. At cluster scales, some of the strongest evidence comes from collisions of clusters such as the Bullet cluster and MACS J0025.4-1222, where x-ray and gravitational lensing data were both available to simultaneously and independently study luminous and total matter content, respectively [468–470]. There is also more general strong lensing evidence from other clusters [471, 472]. Many of the most precise measurements of dark matter now come from cosmology, such as the most recent fit of the CMB power spectrum to Planck data (including temperature, polarization, and lensing), which requires DM to constitute 27% of the energy density of the universe [20], as well as the matter power spectrum [473]. See [474–476] for more detailed discussions.

The fundamental nature of the dark matter generating these gravitational effects is unknown. Many dark matter candidates have been proposed with a diverse set of different masses. These range from ultralight dark matter models with masses of order 10^{-20} eV [477–480] to primordial black holes with masses of order $10M_{\odot}$ [481–485]. Despite its unknown fundamental nature, dark matter is required to satisfy some general constraints: it is required to be neutral with respect to the electromagnetic and strong forces,[7] to be cold, to have a lifetime greater than the age of the universe, and to interact at most very weakly with Standard Model particles. There is somewhat more flexibility for dark matter interactions with itself or other particles in secluded dark sectors, though even self-interactions can still be constrained. There are many excellent

[6] Gravitational lensing is the distortion of light by its gravitational interactions with matter. Strong lensing occurs when the distortion is substantial enough to form multiple images or arcs. With weak lensing, the image is only distorted (rather than forming multiple images or arcs) and distributions of lensed objects often must be studied to draw conclusions. See [466] for a review.

[7] One proposed exception to this is millicharged dark matter, where dark matter has a very tiny electric charge. While millicharged dark matter can arise in sensible models (for example, dark photon models) [486–489], its freeze-out production is ruled out by structure formation and the CMB [490–494] and its freeze-in production is now also strongly constrained [495].

reviews on dark matter candidates and constraints, see, e.g. [474–476, 488, 496–502].

Historically, the most popular dark matter candidates were thermal relic WIMPs. These particles were proposed to be at approximately the same scale as the W and Z bosons ($\mathcal{O}$(100 GeV)), and to interact with standard model particles either through the weak force or a new mediator at a similar scale, depending on the exact model. As thermal relics, WIMPs followed the standard cosmological freeze-out story. Because they were in equilibrium with the Standard Model plasma in the early universe, their population started to decrease exponentially due to annihilation after becoming non-relativistic. Before the dark matter could annihilate completely, the universe cooled enough such that WIMPs were no longer in equilibrium with the Standard Model plasma, leaving a relic density of dark matter. Specifically, thermal WIMPs in the mass range were motivated by the so-called "WIMP miracle," which is the observation that the interaction strength required to correctly set the dark matter thermal relic abundance with thermal freeze-out is the same size as that of the Standard Model weak interactions. Another reason for the popularity of thermal relic WIMPs was their prevalence in many BSM theories; notable examples include the lightest supersymmetry particle (LSP) in R-parity preserving SUSY and the lightest Kaluza-Klein particle (LKP) in extra dimensional Kaluza-Klein theories. Minimal WIMPs in singlet-doublet and doublet-triplet models will be discussed more in §5.

More recently, the parameter space for WIMP models has been extensively probed by a variety of experiments which have substantially limited (but not eliminated) their feasibility. While there are still some viable regions of parameter space in certain thermal WIMP models, increasingly stringent constraints have motivated interest in a wide collection of other potential dark matter candidates. These other models relax several of the implicit assumptions made for the simplest thermal WIMP models, including thermal production through $2 \to 2$ annihilation, a weak scale mass, and the types and strengths of interactions with Standard Model and other dark sector particles. These dark matter candidates are characterized by several features, including mass, spin, type of portal to the Standard Model, production mechanism, and interactions with other dark particles. While the dark matter candidates are too numerous to comprehensively list, examples include:

- Ultralight dark matter candidates with masses below the keV scale, which behave as classical fields. The lightest of these candidates are referred to as fuzzy dark matter and have masses around $10^{-21} - 10^{-22}$ eV [477, 478], but they are now strongly constrained [479, 480]. Axions, which will be discussed more in §1.2.3, also fall into this range.

- Light dark matter models, where the dark matter mass is between keV and GeV scale. Some of the simplest examples include warm dark matter models like sterile neutrinos [503], but these are substantially constrained by structure formation. More complicated examples of light dark matter can be either thermal or non-thermal depending on the details of a given model and can sometimes

evade these constraints (see [488] and references therein for examples).

- The heaviest possible dark matter candidates: black holes which formed in the early universe, called primordial black holes (PBHs) [481, 482, 504, 505]. In order to be all of dark matter, masses of roughly $10^{-11}M_\odot$ to $10^{-16}M_\odot$ are required, but PBHs can be 5-10% of dark matter up to masses $\mathcal{O}(10)M_\odot$ [483–485].

- Dark matter candidates with masses between WIMPs and PBHs. At the lower end of this is particles like Wimpzillas with masses between $\mathcal{O}(100)$ TeV and M_{Pl}, which must be produced non-thermally [506–508]. Above M_{Pl}, dark matter candidates must be composite objects, such as the PBHs previously discussed, quark nuggets [509], or Q-balls [510].

- Weakly interacting candidates with a different cosmological history, including: asymmetric DM, where the abundance depends on the asymmetry between the dark matter and its antiparticle [511–515]; forbidden dark matter, where kinematically forbidden dark matter annihilations occur for dark matter particles with momenta in the upper tail of the thermal momenta distribution [516, 517]; coannihilations, where the abundance of dark matter changes because it annihilates against heavier states [517, 518]; semi-annihilation, where two dark matter particles annihilate to one dark matter particle and one unstable particle [519]; and co-scattering, where the dark matter relic abundance is determined by inelastic scattering of the dark matter off another particle [520].

- Light and weak scale models which change the interaction strength. Examples are feebly interacting dark matter (FIMPs) [521, 522], where interactions are too weak to ever reach equilibrium and the dark matter abundance freezes-in because the interaction strength increases with decreasing temperature, and strongly interacting massive particles (SIMPs) [523], where dark matter is at or below the GeV scale and the relic abundance is set by $3 \to 2$ annihilation from dark matter strong self-interactions. Other similar models that rely on $2 \leftrightarrow n$ interactions in the dark sector include cannibal DM [524, 525] and elastically decoupling relics (ELDERs) [526, 527].

- Many models with more complicated self-interactions and secluded dark sectors, some of which are reviewed in [528]. Sometimes these dark sectors are realized in explicit models, but it is also common to parameterize these in terms of their portal to the Standard Model. Example portals include kinetic mixing of dark photons with ordinary photons, Higgs boson interactions with the dark sector, axion-like particles (ALPs), and neutrinos.

See [488, 529] for more details on many of these dark matter models.

This extensive collection of dark matter models has motivated a similarly extensive experimental program. In addition to providing evidence for the existence of dark matter, many cosmological datasets can

also constrain its properties. One example is the number of effective light degrees of freedom N_{eff} from the CMB and BBN, which can constrain warm dark matter models and other light dark sector particles. Another is CMB data more generally: both anisotropies and spectral distortions can be used to constrain dark matter interactions. A third example is the 21cm signal, which is sensitive to interactions which change the temperature of baryons. Additionally, other constraints come from the matter power spectrum and from gravitational lensing at an assortment of different scales. This is only a small fraction of the many cosmological constraints on dark matter models; a more thorough list can be found in [476].

In addition to cosmological constraints, efforts to confirm the nature of dark matter include searches at colliders, indirect detection experiments, and direct detection experiments. Collider searches try to observe dark matter production, but are typically very model dependent and only sensitive to a small subset of dark matter models. Indirect detection looks for annihilation or decay of dark matter from cosmic sources, usually to photons, neutrinos, or charged particles. The brightest potential indirect detection signals often come from the Milky Way, but searches using other dark matter sources can also be valuable because they tend to be less contaminated. For example, whether the source of the galactic center excess (GCE) that has been observed is from dark matter annihilation is an area of active debate (see §5 for more discussion of the GCE), but the dark matter interpretation of this excess can be constrained by other indirect detection measurements such as those from ultra-faint dwarfs. In addition, there are indirect detection signals from purely gravitational interactions in some models, such as gravitational waves. See [494, 496] for more discussion of indirect detection. Finally, direct detection [530] searches look for dark matter scattering off of Standard Model particles. Traditionally, this has involved looking for nuclear recoils from scattering at experiments like XENON1T [531–533], XENONnT [534], PICO [535, 536], and LZ [537]. More recently there has been substantial progress in methods of looking for lighter WIMPs. Instead of nuclear recoils, these proposals look for phonons, electrons and various different collective excitations [538–549]. See [488, 497, 500] for recent reviews of direct detection, including of light dark matter.

1.2.3 Axions and the Strong CP Problem

The strong CP problem is another fine tuning problem. At its heart is the puzzle of why the neutron electric dipole moment (EDM) is so small. Since CP is not a good symmetry of the Standard Model (it is violated in the CKM matrix), there is nothing to forbid the Standard Model from possessing a $\theta G\tilde{G}$ term. This term generates a neutron EDM of size $10^{-16}\ \overline{\theta}$ e cm.[8] Since the neutron EDM is actually measured to be $|d_n| \leq 1.8 \times 10^{-26}$ e cm [552], the very small value $|\theta| \lesssim 10^{-10}$ is required, which violates our naturalness

[8] See [550] for a review of this calculation; $\overline{\theta} = \theta - \text{argdet}(Y_u Y_d)$ is the invariant combination of θ and the phases of the quark Yukawa matrices. Also see [409, 496, 551] for reviews of the Strong CP problem more generally.

expectations of $\mathcal{O}(1)$ numbers. It turns out that $\overline{\theta}$ is technically natural in the Standard Model because the running of $\overline{\theta}$ first comes in at seven loops [553], but is not 't Hooft natural because there is no symmetry that is restored if we set $\overline{\theta} = 0$.

There are several classes of solutions to the strong CP problem. The simplest of these are massless up quark solutions. The simplified form of this argument says that if we take one of the quarks to have zero mass (the up-quark is usually chosen since it is known to have the smallest mass), then the phase of that quark can be rotated freely, making $\overline{\theta}$ unphysical. More precisely, only the CP odd part of the mass parameter, which is the imaginary part of $M = e^{i\theta} \det(Y_u Y_d)$, needs to vanish for this solution to work. Since instantons can generate purely real M, it is enough to have a quark whose mass is purely generated by instantons [554–557] (also see [558] for a longer discussion). However, data, lattice, and large N calculations have ruled this solution out by showing that QCD instantons are insufficient to achieve the correct quark mass [559–562]. There have been some recent efforts to revive this type of solution with additional instanton effects that come from embedding within a larger gauge group (see [558, 563]), but as of now these solutions are either not fully consistent with various constraints or not yet worked out for the full case of the Standard Model. There have also been some efforts to choose a hidden sector quark be massless instead of the up-quark, see e.g. [564].

Another type of solution makes generalized parity or CP a good symmetry in the UV, and then spontaneously breaks it. CP solutions are called Nelson-Barr [565, 566] models. For these solutions, spontaneous breaking must be engineered so that $\overline{\theta}$ is small but the CKM phase is large. The simplest example of these models [567] has extra SU(2) singlet quarks with hypercharge $\pm 1/3$ which communicate the CP breaking from additional scalars to the Standard Model quarks. In minimal parity solutions [568–570], the Standard Model is extended by $SU(2)_R$ and an additional set of mirror particles charged under this group is introduced. The parity operation is then supplemented with $SU(2)_L \leftrightarrow SU(2)_R$. Although this generalized parity symmetry doesn't explicitly forbid a CKM phase, the symmetry still needs to be broken for the model to be phenomenologically viable. There are some challenges that must be overcome for the minimal realizations of both of these types of models (for example, stabilizing a low P or CP breaking scale; see [571–573] for more details), but model-building can overcome many of them [574–582].

Axions are the third (and most popular) type of solution to the Strong CP problem [583–586]. In what follows, I will briefly describe the QCD axion solution to the Strong CP problem, as well as summarize a few other applications of axions more generally and outline some experimental techniques to search for them. More about axions will be discussed in §2 - §4, and can also be found in many reviews including [496, 550, 551, 587–591].

The QCD axion is a periodic scalar a(x) whose purpose is to generate a $\overline{\theta}$ dependent potential which is

minimized when $d_n = 0$, so that the strong CP problem is solved dynamically when $\overline{\theta}$ relaxes to its minimum. The Lagrangian for the axion includes a kinetic term as well as a coupling to gluons

$$\mathcal{L} \supset \frac{N}{32\pi^2 f_a} \int a(x) \mathrm{Tr}[G\tilde{G}]. \tag{1.1}$$

where f_a is the axion decay constant. $N \in \mathbb{Z}$ is an integer because $a(x)$ is periodic; coupling quantization is required in order for the path integral to be well defined [551]. In addition, derivative fermion couplings and couplings to other gauge fields of the same form are also allowed;[9] other couplings which do not respect a continuous axion shift symmetry should not be present. The axion potential is generated nonperturbatively from the gluon coupling. It can be computed by expanding the Chiral Lagrangian mass term,[10] after using an anomalous Chiral rotation to move the gluon coupling to the phase of that mass term. The result is

$$V(a) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4 m_u m_d}{(m_u + m_d)^2} \sin^2\left(\frac{Na}{2f_a} + \frac{\overline{\theta}}{2}\right)}. \tag{1.2}$$

This potential gives the axion an expectation value $\langle a \rangle = -\overline{\theta} f_a$. Since $d_n \sim \overline{\theta} + a/f_a$, this expectation value sets $d_n = 0$ at the minimum of the axion potential and therefore solves the Strong CP problem.

Since axions couple to gauge bosons through dimension five operators, axion models are only EFTs that must be UV completed. There are several different types of possible UV completions, which can be divided into four dimensional field theory UV completions and higher dimensional ones. In four dimensional UV completions, the axion is typically a pseudo-NGB of a broken U(1) Pecci-Quinn symmetry. There are two classes of four dimensional UV completions: KSVZ [592, 593] and DFSZ models [594, 595]. In KSVZ models, new heavy fermions which have gauge charge are added, along with a scalar φ whose phase is the axion. The scalar and some of the fermions are charged under the PQ symmetry so that the Yukawa couplings between the new fermions and φ are respected by the symmetry, and the scalar has a PQ breaking potential. After symmetry breaking and integrating out the fermions, the yukawa couplings generate $aG\tilde{G}$. An example KSVZ like model will be discussed in §2. In DFSZ models, a second Higgs doublet and a heavy scalar singlet are added, and all Higgs fields carry PQ charge. Again, the Higgs potential breaks $U(1)_{PQ}$, leading to an axion that is a combination of the three scalars. In DFSZ models, since the Higgses are all charged under $U(1)_{PQ}$, Standard Model fermions must also have $U(1)_{PQ}$ charges. This leads to tree-level Standard Model fermion interactions with the axion which can be constrained. In extra dimensional UV completions,

[9] In addition to quantized photon couplings of the the same form as (1.1), the QCD axion also has an axion coupling to photons from the mixing with the pion. As discussed in §2, the coefficient of this coupling violates quantization by a factor proportional to the axion mass squared.

[10] This is the zero temperature potential. At finite temperature, there is also a contribution from thermal instantons which is important for understanding axion cosmology.

there is no PQ symmetry that is spontaneously broken. Instead, the axion is a zero mode of a higher dimensional gauge field, and the higher dimensional gauge symmetry protects the approximately conserved axion current. The axion kinetic term is inherited from the higher dimensional kinetic term (with decay constant depending on the volume of the compactification), and the $\theta G\tilde{G}$ coupling is inherited from the higher dimensional Chern-Simons term. Extra dimensional models of axions were first motivated by their appearance in string theory constructions [596–598], but there are also field theory examples. More details and references on extra dimensional axions can be found in the review [551].

An important challenge for the axion solution to the strong CP problem is the axion quality problem [599–601]. In the four dimensional field theory UV completions described above, there is nothing to forbid operators which explicitly break PQ from spoiling the solution. In fact, we expect these operators to appear at the quantum gravity scale, since quantum gravity is expected to break all global symmetries [602–604]. In order for the axion Strong CP solution to be preserved, Planck suppressed operators with dimension up to 14 must be forbidden. While it is possible to do this using discrete symmetries or compositeness [605,606], these solutions are complicated and spoil much of the simplicity which makes axions especially appealing. However, the quality problem is circumvented in higher dimensional UV completions because other contributions to the axion potential are also exponentially suppressed, and therefore it is reasonable to expect them to be subdominant to the potential from QCD. To see this, note that in extra dimensional theories contributions to the axion potential only come from couplings that depend non-derivatively on the higher dimensional gauge field. This means there are only two additional kinds of contributions to the axion potential. One is from Chern-Simons couplings to additional gauge fields, which are non-perturbative for the same reason as the gluon potential. The second is from the covariant derivative coupling of the gauge field to charged particles. In this case, the contribution to the potential is also an instanton effect coming from the charged field worldlines winding around the compactified dimensions, resulting in exponential suppression by the volume of the extra dimensions. See [551] for a more extensive discussion.

In addition to the QCD axion, there are also a variety of more general axion models which do not solve the strong CP problem. These non-QCD axions are also sometimes referred to as axion like particles (ALPs). Although ALPs do not solve the Strong CP problem, they are motivated by factors including the multiple cosmological roles axions can play, their prevalence in string theory, and by wanting to understand the parameter space of models that are now being probed by ongoing experiments. In these models, the axion is still a periodic scalar with a continuous shift symmetry violated primarily by gauge boson couplings of the form $aF\tilde{F}$. However, unlike in the QCD axion case where this coupling is to gluons, in the more general case F can be the field strength of any gauge field. Additionally, many of these models are more complicated and have more than one axion. While a general discussion of axion models is beyond the scope

of this dissertation, some simple examples are discussed in more detail in §2.

Next, we turn to a discussion of the many different uses of axions. As mentioned in §1.2.2, one of the most important roles for axions is to be dark matter [607–609]. Axions serve this purpose through the misalignment mechanism: they behave as coherent classical fields, which redshift like matter. More explicitly, the equation of motion of the axion field looks like that of a harmonic oscillator. At early times, the axion behaves like an overdamped oscillator due to a large damping term called Hubble friction which is proportional to the Hubble constant. This fixes the value of the axion field to a constant θ_0.[11] As the universe cools, Hubble friction stops dominating, and the axion starts oscillating in its potential. Solving for the axion field in terms of temperature and utilizing it to compute the energy density then shows that the axion redshifts like matter once the axion mass stops changing with temperature. In order to get the correct dark matter relic abundance for the QCD axion, this typically requires an axion decay constant of $f_a = \mathcal{O}(10^{12})$ GeV and a mass $\mathcal{O}(\mu\text{eV})$.[12] In the ALP case, a larger section of parameter space is permitted. Additional modifications [610–621] such as nonstandard cosmologies, kinetic misalignment, trapped misalignment, or axiogenesis can further enlarge the allowed parameter space; see [591] for more discussion.

The details of axion dark matter formation depend strongly on whether there was a PQ symmetry which was broken after inflation. If there was, then in addition to the the misalignment mechanism, there is another contribution to axion dark matter which comes from axion strings. Axion strings (which are discussed more in §4) are solitonic objects which form through the Kibble-Zurek mechanism [622, 623]. In the Kibble-Zurek mechanism, different regions of space take different degenerate values of the vaccuum expectation value after symmetry breaking, which causes solitonic objects to form between these vacua. In the axion string case, different patches of the universe will have different values of θ_0 after post-inflationary PQ breaking. There will be regions around which θ_0 varies from 0 to 2π; at the core of these regions axion strings will form. The network of strings that is formed in this way will persist and evolve as the universe expands (see [496, 591] for details). As the string network evolves, it radiates axions which can be an important contribution to axion dark matter. While this contribution can be estimated analytically, state-of-the-art estimations rely on detailed numerical simulations like [624]. Once the continuous axion shift symmetry is broken by QCD, domain walls can also form. If the domain wall number $N = 1$, then these domain walls will self-annihilate. Otherwise, they will become stable, which is a problem cosmologically because we do not see them.

In addition to being dark matter, axions can also perform several other important functions in the early universe. One example is that the axion can be the inflaton which drives the exponential expansion of the

[11] Depending on whether there is a PQ symmetry which broke after inflation, there might be multiple patches with different fixed θ_0, but locally θ_0 remains fixed for large H.

[12] This assumes the initial θ_0 is $\mathcal{O}(1)$. If PQ breaks after inflation, this is a necessary assumption, since different patches are expected to have different values of θ_0 and we should average them. Otherwise, a smaller dark matter abundance can be obtained by tuning the initial θ_0 to be less than $\mathcal{O}(1)$.

universe during inflation. The axion is a natural candidate for the inflaton [625] because the shift symmetry can protect the flatness of the potential which is necessary for the slow-roll conditions to be satisfied. However, there are challenges associated with engineering a large enough field range for these models to work; see [590] or §2 for more discussion. Also note that an axion that plays the role of the inflaton cannot also be a dark matter candidate. Other potential applications of axions include producing large-scale magnetic fields [626, 627], contributing to the matter/antimatter asymmetry in the Universe [621, 628–630], and helping to generate neutrino masses [631, 632]. Finally, as previously mentioned, axions show up very naturally in string theory as zero modes of higher dimensional gauge fields [596–598, 633–637].

The axion's abundant promise for solving multiple physics puzzles has motivated many different experimental probes. Some of these probes rely on astrophysical and cosmological data, while others are dedicated laboratory experiments. Many probe the axion-photon coupling, but others are sensitive to the axion nucleon coupling or other axion fermion couplings. The validity of the experimental constraints also depends on the axion mass, and on other assumptions such as the axion's cosmological history and whether the axion is dark matter. Since this is such an active area of research, we will only highlight a few examples; a more complete set of constraints can be found at [638].

As mentioned above in the context of dark matter formation, the axion's cosmology is strongly dependent on whether there was a PQ symmetry which broke after inflation. In the case where such a symmetry existed, constraints come from the previously discussed string-domain wall network. In the case where PQ broke before inflation ended (or there was no PQ symmetry to begin with), the strongest constraints come from limits on isocurvature perturbations, which exist for preinflationary axions because the quantum fluctuations of axion fields would also be inflated and turn into isocurvature perturbations once the axion attains mass. Additionally, whether PQ broke before or after inflation, cosmology also strongly constrains thermally produced axions.

We conclude this subsection by listing a few example astrophysical and laboratory constraints on axions. The most well-known laboratory constraints come from haloscopes like ADMX [639–644] which have probed the QCD axion line. They use a resonant cavity with a magnetic field to try to convert dark matter axions to an electromagnetic signal. Other terrestrial experiments searching for dark matter axions include broadband experiments like ABRACADABRA [645–648] and DMRadio [649], which utilize the effective current that is generated by an oscillating axion field in the presence of a magnetic field. Other laboratory experiments which don't assume the axion is dark matter include light shining through walls experiments like ALPs [650, 651] and OSQAR [652], which work by applying a magnetic field to convert an intense laser into axions which could pass through a material, and then convert the axions back to photons on the other side. There are also an increasing number of experiments which can directly probe the axion nucleon coupling,

such as CASPER [653–655] and NASDUCK [656, 657]. In addition to terrestrial experiments, other limits on dark matter axions (especially those with relatively high masses) come from indirect detection of axions annihilating into photons at various wavelengths [658–667]. There are also other astrophysical bounds which do not rely on axions being dark matter, such as those from stellar cooling [668], from superradiance [669–672], from the x-rays emitted by the decay of the irreducible cosmological axion background [673], and from solar or globular cluster axion emission [674–677]. Finally, we reiterate that this is only a subset of the many experimental probes of axion parameter space, and that a more complete list of constraints can be found at [638].

1.2.4 Aesthetic Concerns

In the sections above, we have described in detail two important fine tuning problems with the Standard Model. In addition to these precisely defined naturalness questions, there are also a host of other aesthetic questions we can ask about the Standard Model.

One example of these types of questions is why the CKM and PMNS matrices have the structure that they do. The CKM matrix in particular leads to a hierarchy of six orders of magnitude between the masses of different flavors of charged fermions. Although (as previously mentioned) this hierarchy is technically natural, setting fermion masses so far apart by hand is hardly appealing. Technical naturalness allows us to generate this hierarchy at high scales, but it would still be preferable to have a dynamical explanation at those high scales. We discuss models for the flavor hierarchy in more detail in §7.

Other aesthetic questions have to do with why the Standard Model isn't simpler. For example, why are there three generations of fermions? Or, why does the gauge group take the structure that it does? Why do we live in 3+1 dimensions? There are models which are (partially) motivated by these questions, such as Grand Unified Theories (GUTs), which embed the Standard model gauge group into a simpler unified gauge group at high energies [407, 678–682]. Another example is [558], which proposes a solution (in a toy example) to the Strong CP problem utilizing $N_{colors} = N_{generations}$.

1.2.5 Experimental Anomalies

The potential sources of BSM discussed in the previous several sections are due to fairly well established experimental and theoretical concerns. However, there are many other experimental hints of physics beyond the Standard Model. These are measurements that do not currently agree with Standard Model predictions, but whose disagreement could be transient due to theoretical or experimental mismodeling, including not yet understood systematics. In this section, we briefly discuss some of these current experimental hints (as of Spring 2024) of new physics.

There are several anomalies in flavor physics data. One example which has gotten substantial attention is the muon magnetic dipole moment. When the first results from the muon g - 2 experiment from Fermilab were released in 2021 [683] and confirmed the results of the Brookhaven E821 experiment [684], the experimental result disagreed with the concensus of the Standard Model theoretical value [685] by more than 5σ. Since then, additional results from the Fermilab experiment have been released [686] which are consistent with previous results, but the theoretical value from [685] has been questioned. The challenge arises because the largest contribution to the uncertainty (the hadronic vacuum polarization) cannot be computed from first principles. Instead, it must either be derived using a data-driven dispersion relation approach (from the optical theorem) or using the lattice. More recently, updated lattice results [687] as well as new data from the CMD3 experiment [688] seem to point to a much smaller discrepancy between theory and experiment. Another flavor physics anomaly is the Cabibo angle anomaly. The expectation from a combination of CKM unitarity and the comparatively small size of V_{ub} is that $|V_{ud}|$ and $|V_{us}|$ should be determined from the Cabibbo angle by $|V_{ud}| \simeq \cos\theta_{12}$ and $|V_{us}| \simeq \sin\theta_{12}$. However, values of $|V_{ud}|$, $|V_{us}/V_{ud}|$, and $|V_{us}|$ obtained from beta decay [689, 690], leptonic Kaon decays [691], and semi-leptonic Kaon decays [692] disagree with this expectation at $\sim 3\sigma$. There is a similar $\sim 4\sigma$ discrepancy between the two values of $|V_{ud}|$ which can be extracted from the neutron lifetime [693–695] when it is measured two different ways: one which counts the number of remaining neutrons after waiting for decays, and the other which measures the number of protons which are decay products. An additional anomaly is $R_{D^{(*)}} \equiv BR\left(B \to D^{(*)}\tau\nu\right)/BR\left(B \to D^{(*)}\ell\nu\right)$, which disagrees with the Standard Model at $\sim 3\sigma$ [696]. This is in contrast to $R(K^{(*)}) \equiv Br(B \to K^{(*)}\mu^+\mu^-)/Br(B \to K^{(*)}e^+e^-)$, which was thought to be anomalous but now agrees well with the Standard Model [697]. Other flavor physics anomalies include the recent (not yet statistically significant) Belle-II $BR(B \to K\nu\nu)$ [698] measurement (discussed more in §7) and the amount of CP violation in $B \to K\pi$ decays, which is referred to as the $K\pi$ puzzle [699, 700].

There are also a few other claims of anomalies from collider data. The most significant of these is the CDF-II measurement of the W boson mass [11], which disagrees with the Standard Model at 7σ. Since this anomaly is discussed extensively in §6, we defer additional discussion of it until then. There have also been claims of hints of resonances with low global significance $\approx 1-3\sigma$ at several energies [701, 702]. More detailed discussion of these resonances and other collider anomalies can be found in [703, 704].

Neutrino physics is another source of anomalies in data. These anomalies can be broadly divided into two categories: electron (anti-)neutrino appearance anomalies in short baseline oscillation experiments, and electron (anti-)neutrino disappearance anomalies in reactors and radioactive source experiments. The first category comprises the LSND [705, 706] and MiniBooNE excesses [707]. Both the LSND and MiniBooNE experiments used pion decay to produce muon (anti-)neutrinos, and searched for their oscillation into electron

(anti-)neutrinos. LSND utilized pion decays at rest and could only run in anti-neutrino mode; MiniBooNE used pions which decayed in-flight and could run in either neutrino or anti-neutrino mode. Both experiments found excesses in electron (anti-)neutrino appearance. However, a simultaneous explanation of both excesses is in tension with other constraints. Constraints also make explaining each of these excesses individually a challenge. The simplest models which have been proposed are models with an extra sterile neutrino, however, these are strongly constrained by both cosmology and recent experiments such as MicroBooNE [708]. The second category includes electron neutrino disappearance experiments like SAGE [709], GALLEX [710], and BEST [711, 712], which use Gallium detectors to probe electron anti-neutrinos emitted by radioactive sources. In these experiments, oscillations are too rapid to resolve but aggregating them gives the appearance of a deficit. However, it is also difficult to come up with models which explain this anomaly and are consistent with other constraints, particularly from solar and reactor data [713]. In addition, this second category contains reactor experiments, which were thought to have similar hints of electron anti-neutrino disappearance, but dependence on nuclear physics makes the modeling of the anti-neutrino spectrum in these experiments complicated; recently it has been argued these reactor anomalies can actually be explained by source mismodeling. More recent reactor experiments have also ruled out much of the parameter space favored by previous reactor space in sterile neutrino models. See [714,715] for more details, as well as a more thorough discussion of the various neutrino anomalies, constraints, and models to explain them.

Finally, we conclude this subsection by describing several anomalies in astrophysical and cosmological data which could have explanations stemming from new fundamental particles. An important example which has driven substantial model building in the last several years is the Hubble tension, which is the discrepancy between early universe measurements of the Hubble expansion rate from the CMB data and late universe measurements of the same rate [19–23]. Many different solutions [19,716] which introduce BSM particles have been proposed to reduce this tension, however they run into obstacles completely eliminating the tension. A similar cosmological tension is the S_8 tension. S_8 characterizes the clustering of matter at scales of 8 h^{-1} Mpc; the S_8 tension arises because the measurements of S_8 from late universe data do not agree with those inferred from the CMB. The S_8 tension is reviewed in [716]. Additionally, there have been several claims of excesses in astrophysical charged cosmic ray data, including in positrons [717, 718], anti-protons [719, 720], and anti-helium nuclei [721–723]. While these have been suggested to be evidence of various WIMP dark matter models, more recently it has been argued that the positron and anti-proton excesses can be explained due to more conventional means. Specifically, positrons could be created and accelerated by pulsars [724], and the anti-positron excess disappears when correlated systematic errors are taken into account [725, 726]. In contrast, astrophysical explanations for the anti-helium excess are more difficult [727]. See [724] for more details on cosmic ray excesses. Another anomaly is the very recent measurement of the dark energy equation

of state using baryon acoustic oscillation data from DESI [728], which shows a preference for time-dependent dark energy over a cosmological constant. However, it has been argued that we should be cautious in taking this result at face value, both because the phenomenological model that was fit to data is hard to describe with sensible particle physics models and because the results depend on the choice of prior for the parameters in the equation of state [729]. The final anomaly we will discuss is the statistically significant excess of gamma rays found at the galactic center called the galactic center excess (GCE). Various analyses have showed that the GCE is well fit by either dark matter or millisecond pulsars, and that we are not currently able to conclusively determine its source. The dark matter explanation of this galactic center excess (GCE) will be discussed in more detail in §5.

2

Axion Periodicity and Coupling Quantization in the Presence of Mixing

2.1 Introduction

Axion fields are ubiquitous in theories of physics beyond the Standard Model. For this chapter, the defining feature of an axion (as compared with a generic scalar field) is that it is a compact boson, whose target space is a circle. That is, an axion field by definition is identified under a discrete shift symmetry:

$$a(x) \cong a(x) + 2\pi n F_a, \quad n \in \mathbb{Z} \tag{2.1}$$

where $2\pi F_a$ is the fundamental period of the field.

The periodicity (2.1) of an axion has significant consequences for the structure of axion effective field theories. An immediate consequence is that axion potentials must be periodic functions. Less obviously, for reasons that we will review below, couplings of the form $aF_{\mu\nu}\widetilde{F}^{\mu\nu}$ of axions to gauge fields must have *quantized* coefficients, which are integer multiples of $e^2/(16\pi^2 F_a)$. This poses a significant challenge for many phenomenological models that rely on axions. For example, in a cosmological model one might be interested in an axion potential with a very large field range, but at the same time may want a large coupling of

the axion to gauge fields (e.g., for reheating [730], magnetogenesis [626, 627], or for the structure of the inflationary model itself [731, 732]). Because the axion potential and the $a F\widetilde{F}$ couplings both depend on F_a, our options for building such models are very limited, unless the constraints imposed by periodicity (2.1) can be relaxed.

In this chapter, our chief interest is in the robustness of the constraints associated with axion periodicity. Can an effective field theory containing periodic axions flow in the infrared to a new effective field theory in which some of the axion fields have become effectively non-compact, and hence have fewer constraints on their couplings? By considering various examples, in which axions mix with other axions or with non-compact scalar fields, we will argue that the options are very limited. In particular, we claim that whenever some of the axions remain massless in the IR, they will continue to *exactly* respect periodicity constraints. Deviations from these constraints are always proportional to powers of the axion mass. This is reminiscent of the fact that quantum field theories with compact gauge groups in the UV flow to quantum field theories with (possibly different) compact gauge groups in the IR. As we will discuss below, this is more than a superficial similarity.

Before summarizing our results in more detail, let us briefly review the properties of axion effective field theories enforced by the shift symmetry (2.1).

2.1.1 Review: quantized couplings in axion EFT

Readers who are thoroughly familiar with the reason why $a F\widetilde{F}$ couplings are quantized, and how to precisely formulate this condition in theories with fermions, can safely skip this subsection, though it may be useful for establishing our conventions.

Because we will be studying scenarios in which axions may not have canonical kinetic terms, it is often useful to consider *dimensionless* axion fields θ which are normalized to have period 2π,

$$\theta(x) \cong \theta(x) + 2\pi n, \quad n \in \mathbb{Z}. \tag{2.2}$$

These identifications on field space may be thought of as discrete gauge symmetries. In certain theories, such gauge symmetries may be *spontaneously broken*, in which case an axion may appear to acquire a non-periodic potential or other interactions that violate the symmetry. In such cases, there is a monodromy when the axion traverses its fundamental circle, so that the full set of states of the theory actually respects the underlying symmetry. We will refer to such fields as "monodromy axions." Monodromy axions have played a major role in inflationary model-building [733, 734].

The periodicity (2.2) imposes important, well-known constraints on the effective field theory of an axion. An obvious one is that (in the absence of monodromy) the potential is periodic, $V(\theta) = V(\theta + 2\pi)$. In many

theories of axions, there are important couplings between axions and gauge fields of the form

$$\mathcal{L}_{\theta F\widetilde{F}} = k\frac{\theta}{16\pi^2}F_{\mu\nu}\widetilde{F}^{\mu\nu}, \quad k \in \mathbb{Z}, \tag{2.3}$$

where the dual gauge field is defined as $\widetilde{F}^{\mu\nu} := \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. The requirement that k is quantized follows from the axion periodicity (2.2). Here we have assumed that the normalization of F is such that the kinetic term is $-\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}$ and that particles of gauge charge $q \in \mathbb{Z}$ couple to the gauge field through the action $S = q\int_\gamma A$, where γ is the charged-particle worldline and $A = A_\mu\, dx^\mu$ is the 1-form gauge field. The reason that the coupling k in (2.3) is quantized is that the interaction Lagrangian is not gauge invariant: its coefficient changes value under the shift $\theta \mapsto \theta + 2\pi n$. However, the path-integral measure is well-defined whenever $k \in \mathbb{Z}$, because $\exp\bigl(i\int d^4x\,\mathcal{L}_{\theta F\widetilde{F}}\bigr)$ is well-defined.

Our statement of the quantization of the coupling (2.3) applies when we consider this coupling in isolation. In theories with fermions that couple to θ and transform under the gauge field, the correct statement of coupling quantization refers to an invariant combination of couplings. For example, if we consider a Lagrangian containing the terms

$$i\overline{\Psi}\not{D}\Psi + c_\partial(\partial_\mu\theta)\overline{\Psi}\gamma^\mu\gamma^5\Psi - \bigl[me^{ic_m\theta}\overline{\Psi}_L\Psi_R + \text{h.c.}\bigr] + c_F\frac{\theta}{16\pi^2}F_{\mu\nu}\widetilde{F}^{\mu\nu}, \tag{2.4}$$

where the Ψ transform in the fundamental representation of the gauge group, then the field redefinition

$$\Psi_L \mapsto e^{ia\theta}\Psi_L, \quad \Psi_R \mapsto e^{-ia\theta}\Psi_R \tag{2.5}$$

produces a different Lagrangian with replacements

$$c_\partial \mapsto c_\partial - a, \quad c_m \mapsto c_m - 2a, \quad c_F \mapsto c_F + 2a, \tag{2.6}$$

with the shift in c_F arising due to the chiral anomaly (e.g., from the anomalous transformation of the fermion measure in the path integral). As a result, it is clearly not correct to demand that $c_F \in \mathbb{Z}$ in general. However, if we first decouple the axion from the fermions by performing a field redefinition to set $c_m = 0$, so that the axion couples *only* through interactions like c_∂ that preserve a continuous shift symmetry and through $\theta F\widetilde{F}$ type terms, *then* the latter terms are quantized. In other words, the correct quantization condition in the case of the Lagrangian (2.4) is

$$c_m + c_F \in \mathbb{Z}. \tag{2.7}$$

This suffices to ensure that the path integral is well-defined under the identification (2.2). Invariant combi-

nations of couplings including the derivative term, such as $c_\partial - \frac{1}{2}c_m$, can take any real value.

The quantization rules (2.3) or (2.7) apply for axion couplings to U(1) gauge fields or to nonabelian gauge fields, up to a change in the linear combination of coefficients appearing in (2.7) that depends on the Dynkin index of the gauge representation of the fermions. In most of the equations in this chapter, a factor of $F_{\mu\nu}\widetilde{F}^{\mu\nu}$ may be replaced by $\frac{1}{2}F^a_{\mu\nu}\widetilde{F}^{a\mu\nu}$ for a nonabelian group without changing the validity of our statements. The only necessarily abelian gauge fields that we discuss will be those in §2.3 that eat axions to acquire a mass, and the higher-dimensional gauge field in §2.4 that is used to engineer a simple scenario with monodromy. (In both cases one could consider nonabelian extensions, but this would complicate the physics without obvious dividends.)

2.1.2 Summary: motivation and results

Axions, like more general scalar fields, can mix with other fields in a variety of ways. They may have mass or kinetic mixing with other axions (e.g., [735–737]). Some linear combinations of the axions may be eaten by massive spin-1 fields (via the Higgs mechanism or Stückelberg couplings, e.g., [738–742]). Axions may even mix with other fields that are not periodic, whether these are ordinary scalar fields or monodromy axions (e.g., [743]). When some of these fields acquire mass, we can integrate them out to obtain an effective field theory involving only the light fields.

The central question of this chapter is: does the EFT of the light fields always inherit a periodicity condition like (2.2) and the associated constraints? For example, can one begin with a theory of two axions, one linear combination of which acquires a mass (either through a potential or through being eaten by a massive spin-1 field) so that the remaining, light combination is no longer an axion (i.e., has no well-defined period)? The answer to an analogous question in gauge theory is familiar: if we consider a theory with a compact gauge group, which is reduced to a smaller gauge group in the infrared through Higgsing, then the infrared gauge group will still be compact. For example, in the Standard Model, the photon couples to an electromagnetic charge whose quantization is inherited from the quantization of $SU(2)_L$ and $U(1)_Y$ charges. This follows from the fact that the Higgs field itself carries quantized charges. Similarly, even in theories with kinetic mixing, there is a discrete charge lattice for the *massless* U(1) bosons, whether or not they mix with massive spin-1 bosons [744, 745]. Despite the existence of such analogous results, we emphasize that our results for spin-0 bosons do *not* all precisely map to familiar results for spin-1 bosons. For example, we will discuss cases in which spin-0 bosons are eaten by spin-1 bosons, quantization of $a\widetilde{FF}$ couplings and the role of massless chiral fermions in determining the invariant quantized couplings, and axion monodromy. These additional ingredients require different arguments from those of [744, 745].

Apart from its intrinsic interest as a question about the structure of quantum field theory, our motiva-

tion for studying this question is that the constraints imposed by the periodicity (2.2) can provide serious obstructions to building interesting phenomenological models. The literature on applications of axions in phenomenology is vast, so we cannot provide a complete bibliography, but some of the main themes and specific examples to which our work is relevant include:

- **Hierarchies between couplings.** One interesting goal is to have axion couplings to $F\widetilde{F}$ terms with very different sizes. In a theory where these couplings are quantized, this can only be achieved by invoking a large integer, which one could then attempt to explain from within a UV completion (e.g., [736, 746–748]). An obvious application is to the QCD axion, where one might like to separate the coupling to gluons (which determines the axion mass) from the coupling to photons (which is often invoked to provide experimental tests of the theory). Various models can alter the ratio of these couplings [749, 750].
- **Achieving large field ranges.** Especially in cosmological applications, it is often of great interest to have a field that can evolve over a long distance in field space. For example, this is necessary to produce large primordial gravitational wave signals from standard inflation models [751], or to allow novel mechanisms like dynamical relaxation of the weak scale to operate [444]. In the context of string theory, it is known to be difficult to find axions with fundamental period larger than the Planck scale (e.g., [752–755]), which has motivated many efforts to build models where small field ranges in the UV become large field ranges in the IR (which are too numerous to review here).
- **Reconciling a large field range with a large coupling.** In some cases, the challenge is a blend of the two previous ideas. One might want a large axion field range f appearing in terms like cos(a/f), but also a large coupling $\frac{\alpha}{8\pi}\frac{a}{f'}F\widetilde{F}$, and hence a small scale f'. Because f and f' are both related to the axion period, again, it can be difficult to achieve a large separation of these scales. This issue arises in chromonatural inflation [732], which in any single-axion model requires an enormous integer to appear in the effective action [756, 757]. Similar issues arise when using axion couplings to gauge fields for preheating [730], to suppress the axion dark matter abundance [758, 759], or to produce dark photon dark matter [760–762].

Separate from these specific phenomenological or model-building goals, if an axion field is discovered experimentally in the future, precisely measuring its couplings and understanding whether they are quantized could play a critical role in interpreting the signal. Clearly, it is important to understand our theoretical expectations before any such discoveries are made.

We will see that in studying simple theories in which multiple axions mix, interesting subtleties arise in examining the periodicity and couplings of a light axion. If one simply examines formulas that are present in the literature, one might suspect that the IR theory in general does not inherit any periodicity constraint

from the UV theory. We will encounter a case in which the light axion field appears to be non-compact, and yet inherits periodic couplings just as a compact field would. We will also encounter a case in which the light axion field at first sight appears to be compact, but periodicity-violating couplings appear in the EFT. These results provide tantalizing hints for the construction of phenomenological models that can evade various constraints, and in some cases claims of large effective field ranges in such models have been made in the literature [740, 741, 763]. However, in every case that we study, a more careful examination reveals that the periodicity of the axion field and quantization of (properly defined, invariant) couplings are properties of the infrared theory whenever the light axions remain massless. In the particular cases referenced above in [740,741,763], the authors overlook subtleties related to the absence of anomalies, which relates the various parameters in the Lagrangian and enforces quantization. In particular, these relations prevent some of the scenarios discussed in [740, 741, 763] from being able to generate large effective field ranges. Once a mass is generated, the constraints are loosened. However, all such effects are proportional to powers of the light axion mass.

In order to achieve hierarchies between an axion's coupling to different gauge fields, or between an axion field range and the scale suppressing its coupling to a gauge field, we find the following options:

- The axion couplings remain quantized due to periodicity, and the hierarchy arises from a large integer, as in the clockwork scenario [736, 746–748, 764, 765].
- The axion is massive, and its couplings deviate significantly from their expected quantization due to mixing with other axions with masses generated at the same scale. This possibility is familiar from the QCD axion's coupling to the photon, which obtains a non-quantized contribution from mixing with the π^0 [766–768].
- Mixing between monodromy axions and ordinary axions can "realign" monodromy to a light axion with a larger field range than the original monodromy axion, as in the "Dante's Inferno" model [743]. This effectively extends the axion field range by allowing it to "unwind."

Some aspects of our claims have been noted in other recent work, including [750] by one of us and [742]. We extend earlier work by surveying a wider range of models, but also by situating the question in the broader theoretical context of compactness of gauge groups. Some of our arguments in §2.2 resemble those made in the past about mixing of spin-1 gauge fields [744, 745], though various details (e.g., our use of the Smith normal form in §2.2.2, or the effects of turning on a mass for the light axion) are not directly analogous to results in those references.

The outline of this chapter is as follows: in §2.2, we discuss scenarios in which some linear combinations of axions obtain periodic potentials. We show that the remaining, light scalar fields are always periodic (their field space is a torus) and their couplings are quantized as expected. In §2.3, we consider the possibility that

a linear combination of axions decouples because it is eaten by a massive spin-1 field (either via the Higgs or Stückelberg mechanisms). Again, we show that the uneaten combination is a periodic field with quantized couplings. The results of this section were also obtained independently in [742], which appeared while this work was being completed. In §2.4, we discuss the possible mixing of axions with other, non-compact scalars. We show that a theory in which a monodromy axion mixes with a heavier ordinary axion can lead to a "realignment" of monodromy to a linear combination of the original axions, so that the axion decay constant is larger in the low-energy effective field theory. In §2.5, we discuss the relationship between our studies of compactness in axion field spaces and the question of compactness of gauge groups. In particular, we point out that in some cases these questions are related by Hodge (electric/magnetic) duality. We suggest that our results fit into a larger picture in which theories with compact gauge groups in the UV always flow to theories with compact gauge groups in the IR. Finally, we very briefly conclude this chapter in §2.6.

2.2 Mixing with a Heavier Axion with a Periodic Potential

2.2.1 Light axion remaining massless

The first scenario we will consider is when two axions mix and a periodic potential gives a mass to one linear combination of them, leaving one massless axion in the IR. We will argue that there is a consistent EFT description in which the light axion is periodic and has quantized couplings to gauge fields. Elements of our discussion, involving the diagonalization of kinetic mixing in the case of a massive axion, have previously appeared in [735, 769], and some of the conclusions about quantized couplings were previously emphasized in [750]. Nonetheless, it is useful to highlight a confusing aspect of the calculation that has not previously been emphasized, and then explain how this confusion is resolved. We will encounter a similarly confusing intermediate result in §2.3, which our experience in this section will help to resolve correctly.

Setting up the problem in a convenient lattice basis

We will denote our two axion fields θ_1 and θ_2 and assume that they both have period 2π. A different way to say this is that our field space is a torus, obtained by taking the quotient of the plane (θ_1, θ_2) by the lattice $(2\pi n_1, 2\pi n_2)$, $n_i \in \mathbb{Z}$. A linear transformation

$$\begin{pmatrix} \theta_1' \\ \theta_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \tag{2.8}$$

preserves this structure provided that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2,\mathbb{Z}). \tag{2.9}$$

We will call any such basis for our field space a "lattice basis." Other bases are, of course, possible, but they require us to reparametrize the lattice of identifications of the plane.

We will consider an effective Lagrangian of the form

$$\mathcal{L} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + K_{ij}\partial_\mu\theta_i\partial^\mu\theta_j - V(j_1\theta_1 + j_2\theta_2) + \frac{k_1\theta_1 + k_2\theta_2}{16\pi^2}F_{\mu\nu}\widetilde{F}^{\mu\nu}, \tag{2.10}$$

where $j_i, k_i \in \mathbb{Z}$ but K_{ij} is an arbitrary symmetric real matrix of rank 2. For concreteness, one could imagine the potential to take the form

$$V(j_1\theta_1 + j_2\theta_2) = \Lambda^4\left[1 - \cos(j_1\theta_1 + j_2\theta_2)\right], \tag{2.11}$$

which might be the leading approximation to the potential generated by a confining, pure glue sector via the coupling

$$\frac{j_1\theta_1 + j_2\theta_2}{32\pi^2}G^a_{\mu\nu}\widetilde{G}^{a\mu\nu}. \tag{2.12}$$

However, the only important assumption we will make is that $V(x)$ has a minimum at $x = 0$, and a Taylor expansion $V(x) \approx V_0 + \frac{1}{2}\mu^4x^2 + \mathcal{O}(x^3)$. Without loss of generality, we will assume that $\gcd(j_1, j_2) = 1$, by absorbing any common factor into the normalization of the function V.

In the subsequent discussion, we will often drop the $-\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}$ term when writing our Lagrangian. It is understood to be present whenever a coupling to $F\widetilde{F}$ appears.

It is always possible to perform a $\mathrm{GL}(2,\mathbb{Z})$ transformation so that the massive axion is a basis vector, $\theta'_1 = j_1\theta_1 + j_2\theta_2$. To see this, observe that there must exist integers ℓ_1, ℓ_2 such that $j_1\ell_2 - j_2\ell_1 = 1$, as a consequence of our assumption that $\gcd(j_1, j_2) = 1$. Thus, we can define a new lattice basis as $\theta'_1 = j_1\theta_1 + j_2\theta_2$ and $\theta'_2 = \ell_1\theta_1 + \ell_2\theta_2$. The Lagrangian (2.10), written in the new basis, has the same form, with

$$\begin{aligned} k'_1 &= \ell_2 k_1 - \ell_1 k_2, \\ k'_2 &= -j_2 k_1 + j_1 k_2, \\ K' &= (M^{-1})^{\mathrm{T}} K M^{-1} \text{ where } M = \begin{pmatrix} j_1 & j_2 \\ \ell_1 & \ell_2 \end{pmatrix}. \end{aligned} \tag{2.13}$$

Here K denotes the kinetic matrix whose entries K_{ij} appeared in (2.10). Notice that the $\mathrm{GL}(2,\mathbb{Z})$ transformation maintains the quantization of couplings, $k'_i \in \mathbb{Z}$, as any lattice basis should. Without loss of generality,

then, we can study the Lagrangian (2.10) in the special case that the potential depends only on θ_1'. Let us do so, dropping the $'$ labels:

$$\mathcal{L} = K_{ij}\partial_\mu\theta_i\partial^\mu\theta_j - V(\theta_1) + \frac{k_1\theta_1 + k_2\theta_2}{16\pi^2}F_{\mu\nu}\widetilde{F}^{\mu\nu}. \tag{2.14}$$

We could, equivalently, rewrite this in terms of two periodic, *dimensionful* axion fields a_i with period $2\pi F_i$, as in (2.1), with a dimensionless kinetic mixing parameter ε:

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{2}\partial_\mu a_i\partial^\mu a_i + \varepsilon\partial_\mu a_1\partial^\mu a_2 - V(a_1/F_1) + \frac{1}{16\pi^2}\left(k_1\frac{a_1}{F_1} + k_2\frac{a_2}{F_2}\right)F_{\mu\nu}\widetilde{F}^{\mu\nu}, \tag{2.15}$$

where

$$F_i := \sqrt{2K_{ii}} \quad \text{and} \quad \varepsilon := \frac{K_{12}}{\sqrt{K_{11}K_{22}}}. \tag{2.16}$$

Diagonalizing the propagating states

The Lagrangian (2.14) clearly describes one massive propagating field, θ_1, and another massless propagating field. For general K_{ij}, the massless field will be a general linear combination of θ_1 and θ_2, not necessarily aligned with any lattice vector. This means that it is not a periodic scalar, but rather winds around the torus in an irrational direction, never returning to its starting point. To identify this direction, we can diagonalize *both* the mass and the kinetic terms by performing a shift of the light field, i.e. by defining

$$a_L := a_2 + \varepsilon a_1. \tag{2.17}$$

This resembles the familiar diagonalization of massive dark photons kinetically mixing with the massless ordinary photon [744], which was further discussed in [745]. To canonically normalize the independently propagating fields, we can further introduce a rescaled heavy field

$$a_H := \sqrt{1-\varepsilon^2}a_1. \tag{2.18}$$

In terms of a_H and a_L, the Lagrangian takes the diagonalized form

$$\mathcal{L} = \frac{1}{2}\partial_\mu a_L\partial^\mu a_L + \frac{1}{2}\partial_\mu a_H\partial^\mu a_H - V(a_H/f_H) + \frac{1}{16\pi^2}\left(k_2\frac{a_L}{f_L} + (k_1 - \rho k_2)\frac{a_H}{f_H}\right)F_{\mu\nu}\widetilde{F}^{\mu\nu}, \tag{2.19}$$

where

$$f_H := \sqrt{1-\varepsilon^2}F_1, \quad f_L := F_2, \quad \text{and} \quad \rho := \varepsilon\frac{F_1}{F_2}. \tag{2.20}$$

We have denoted the suppression scale in the couplings by lowercase f rather than capital F to signal that, unlike in (2.1), they do not necessarily have an interpretation as the period of a compact boson. The quantity ρ is essentially a measure of how misaligned the basis of propagating fields is with the lattice basis.

This form of the effective Lagrangian has been derived several times before, e.g., [735,750,769]. However, there is an aspect of it that is, at first sight, puzzling and has not (to the best of our knowledge) been commented on. Specifically: the field a_H, being proportional to θ_1, is a periodic scalar, yet its couplings to gauge fields depend on the (generically) irrational number ρ and thus are not quantized. On the other hand, the field a_L is *not* a periodic scalar, but its couplings to gauge fields are quantized (proportional to the integer k_2).

Should this bother us? Our argument that periodic scalars have quantized couplings was based on requiring that exp(iS) be gauge-invariant when the scalars are shifted. Because (2.19) is fully equivalent to our manifestly gauge-invariant starting point (2.10), it must be the case that exp(iS) is well-defined despite the non-quantized coupling of the periodic scalar a_H. The reason is that a gauge transformation $\theta_1 \mapsto \theta_1 + 2\pi n$ does not *only* shift a_H, but also shifts a_L; our diagonalized Lagrangian (2.19) is, as it must be, invariant under the gauge transformations

$$\begin{aligned} &a_H \mapsto a_H + 2\pi n_1 f_H, \\ &a_L \mapsto a_L + 2\pi\,(n_2 + \rho n_1)\, f_L, \quad n_i \in \mathbb{Z}, \end{aligned} \tag{2.21}$$

which simply reflect the coordinates of the lattice in our new, misaligned basis. The lack of quantization of the a_H coupling leads to a shift in the action under a gauge transformation of a_H that is precisely compensated by the corresponding, ρ-dependent shift of a_L under the same gauge transformation. Everything is as it should be. However, one might wonder whether the lack of periodicity of a_L means that we can integrate out a_H and obtain a low-energy EFT of a_L that lacks the constraints that usually come from periodicity. Given that our Lagrangian has quantized couplings of a_L to $F\widetilde{F}$, it does not seem to be so easy to escape the constraints of periodicity. In fact, the non-periodicity of a_L is a red herring. Properly understood, the low-energy effective theory is a theory of a compact field, as we will now explain.

Periodicity in the low-energy EFT

We have noted that the light axion field a_L is not a simple periodic field, but it still has quantized couplings. We can understand this better by examining the two-axion field space, as shown in Fig. 2.1. The field space consists of periodic variables a_1 and a_2, whereas when we diagonalize the kinetic terms we find a light field a_L which is an irrational combination of the two, and which is constant on the blue diagonal lines in the

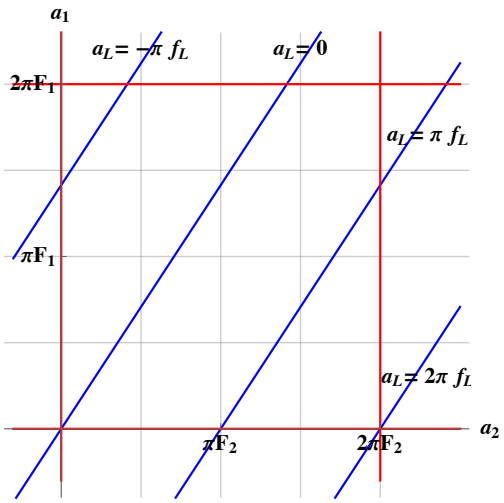


Figure 2.1: The space of two axions a_i with periods $2\pi F_i$. The red lines indicate two periodically identified values of the two axions. The blue lines are contours of constant $a_L = a_2 + \varepsilon a_1$, with $\varepsilon = -\frac{1}{\sqrt{2}}$ chosen as an example. Notice that a 2π shift of the heavy field a_1 shifts a_L by an irrational amount, but a 2π shift of a_2 at constant a_1 simply shifts $a_L \to a_L + 2\pi f_L$. This is the gauge symmetry of the EFT along the flat valley minimizing the potential $V(a_1)$.

plot.

The potential $V(a_1)$ is independent of a_2, and hence constant along horizontal lines in this plot. This means that there is a flat valley along the horizontal red line at $a_1 = 0$, which is repeated at $a_1 = 2\pi F_1$ and other gauge equivalent locations. The effective field theory of the light axion should be defined along this valley, since the field can move along it without incurring any potential energy cost. Notice that this statement is *independent* of the kinetic term for the axions, and in particular of the direction along which lines of constant a_L are oriented. The gauge symmetry $a_1 \mapsto a_1 + 2\pi F_1$ does not shift a_L by a quantized multiple of $2\pi f_L$. However, the only gauge symmetry that makes sense within the low-energy effective theory defined in a valley of fixed a_1, namely $a_2 \mapsto a_2 + 2\pi F_2$, *does* shift a_L by $2\pi f_L$: it is the horizontal translation that takes, for instance, the diagonal at $a_L = 0$ to that at $a_L = 2\pi f_L$. Furthermore, these facts are preserved by any lattice basis in which the potential depends only on θ_1; we could send $\theta_2 \mapsto \theta_2' := \theta_2 + n\theta_1$, and in the (θ_1, θ_2') basis it is still true that V is a function only of θ_1 and that the couplings of θ_2' are quantized. The EFT with θ_1 frozen at the minimum of its potential takes exactly the same form in the new basis.

From this point of view, there is very little mystery: the EFT is defined along the flat direction in field space, which is periodic. The couplings of the massless periodic axion should be quantized, and we have found that they are. A lesson to draw from this, which generalizes to other contexts, is that although diagonalizing the propagating states is a good way to proceed if you plan to do Feynman diagram calculations with multiple fields, it can be an unnecessarily confusing step in the process of understanding the correct way to think about the low-energy EFT.

A natural, straightforward approach to understanding the low-energy EFT is to obtain a theory of the periodic field θ_2 by directly integrating out θ_1 using its equation of motion,

$$2K_{11}\Box\theta_1 + 2K_{12}\Box\theta_2 + \mu^4\theta_1 - \frac{k_1}{16\pi^2}F_{\mu\nu}\widetilde{F}^{\mu\nu} + \ldots = 0, \tag{2.22}$$

where the omitted terms arise from higher orders in the Taylor expansion of $V(\theta_1)$. Solving this equation reveals that

$$\begin{aligned}\theta_1 &= -\frac{2K_{12}\Box\theta_2}{\mu^4} + \frac{4K_{11}K_{12}\Box^2\theta_2}{\mu^8} + \frac{k_1}{16\pi^2\mu^4}F_{\mu\nu}\widetilde{F}^{\mu\nu} + \ldots \\ &= \frac{1}{F_1}\left[-\frac{\varepsilon}{m_1^2}\Box a_2 + \frac{\varepsilon}{m_1^4}\Box^2 a_2 + \frac{k_1}{16\pi^2 m_1^2 F_1}F_{\mu\nu}\widetilde{F}^{\mu\nu} + \ldots\right],\end{aligned} \tag{2.23}$$

where in the second line we have rescaled to dimensionful fields and made the replacement $\mu^4 = m_1^2F_1^2$. This makes it apparent that we could equally well obtain such an expansion by working with Feynman diagrams defined in terms of the fields $\theta_{1,2}$ rather than the diagonalized fields. The kinetic mixing is then an insertion proportional to $\varepsilon\Box$ that flips a θ_2 propagator to a θ_1 propagator or vice versa, and leads to the $\Box a_2$ terms in the above equation.

In summary: the EFT of the light field is a theory of an axion θ_2, with couplings to gauge fields quantized as expected given its periodicity. All of the effects of kinetic mixing with the heavy field are encoded in manifestly shift symmetry-preserving derivative interactions.

2.2.2 An N-axion generalization

Above, we saw that if we had two axions (θ_1, θ_2) and a potential depending on one linear combination of the two, we could change to a new lattice basis in which the potential is independent of the periodic axion θ_2'. This allows us to integrate out the heavy field and obtain a theory of only the compact axion θ_2'. It is natural to generalize this to the case of N axions $(\theta_1, \ldots, \theta_N)$ with period 2π as follows: suppose that we have a potential that depends on k independent linear combinations of the N axions and respects their periodicity,

$$V = V(\vartheta_1, \ldots, \vartheta_k) \quad \text{where} \quad \vartheta_i = \sum_{j=1}^{N} Q_{ij}\theta_j, \quad Q_{ij} \in \mathbb{Z}. \tag{2.24}$$

Then we claim that there is a new lattice basis, $\theta_1', \ldots, \theta_N'$, in which the potential has the form $V(\theta_1', \ldots, \theta_k')$ and is independent of $\theta_{k+1}', \ldots, \theta_N'$. Hence, we can integrate out the massive modes $\theta_1', \ldots, \theta_k'$ to obtain an effective field theory of the $N - k$ massless, 2π-periodic axions $\theta_{k+1}', \ldots, \theta_N'$.

This fact follows from the existence of the *Smith normal form* [770] for matrices over a principal ideal domain (such as the integers): given the $k \times n$ integer matrix Q with entries Q_{ij}, there exist integer matrices

$S \in GL(k, \mathbb{Z}), T \in GL(n, \mathbb{Z})$ such that

$$R := SQT = \begin{pmatrix} r_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & r_k & 0 & \cdots & 0 \end{pmatrix}, \quad r_i \in \mathbb{Z}. \tag{2.25}$$

(The Smith decomposition also implies that we can arrange that r_i divides r_{i+1}, but we will not need this.) The definition of $GL(m, \mathbb{Z})$ means that S and T are invertible and their inverses have integer entries.

In our original basis, the span of the rows of Q defines the subspace of axions that obtain a potential. We can change to a new lattice basis by defining

$$\theta'_i = \sum_{j=1}^{N} T^{-1}_{ij} \theta_j, \tag{2.26}$$

where T^{-1}_{ij} are the entries in T^{-1}. In terms of this basis, the potential depends on the span of the rows of the matrix $QT = S^{-1}R$. We can read off immediately that the span of the rows of R contains only linear combinations of the first k basis vectors in the θ'_i basis. The form of R together with invertibility of S^{-1} guarantees that the rows of $S^{-1}R$ span the same subspace. Hence the potential is independent of $(\theta'_{k+1}, \ldots, \theta'_N)$.

This shows that our discussion of the 2-axion case can be fully generalized to N axions. When a potential gives a mass to $k < N$ axions, we can always find a lattice basis where it is manifest that $N - k$ axions with period 2π are flat directions. By the usual logic of effective field theory, then, we can integrate out all of the heavy axions, and obtain a theory of the $N - k$ light axions that respects all of the expected quantization rules for axion couplings. Any kinetic mixing with heavy axions, upon integrating them out, will produce only shift-symmetric terms involving $\Box$ acting on light axions, as we saw above.

2.2.3 Light axion obtaining a mass

So far we have discussed only cases in which light axions remain exactly massless, and have found that they are periodic fields with exactly quantized couplings. The quantization of axion couplings can be violated once the axions obtain a mass. One straightforward way to see this is by noting that within the effective field theory, we can use equations of motion to make the replacement

$$\Box a_L \mapsto -\frac{\partial V(a_L)}{\partial a_L} \approx -m_L^2 a_L + \ldots, \tag{2.27}$$

which exchanges a term that is manifestly invariant under continuous shift symmetries of θ_2 with one that is not. Although the linear term coupling a_L to $F\widetilde{F}$ is not necessarily quantized, if we keep higher-order terms in a_L this replacement does preserve the discrete shift symmetry (2.1) because $V(a_L)$ is a periodic function. One could also see this effect from the Feynman diagram approach; an external, on-shell light axion of mass m_L that kinetically mixes with the off-shell propagator of a heavier axion of mass m_H will obtain an insertion proportional to $\varepsilon p^2 = \varepsilon m_L^2$ followed by a propagator factor of $1/(m_H^2 - m_L^2)$, which agrees with the EFT result obtained by integrating out θ_1 using (2.23) and expanding order-by-order in $m_L^2/m_H^2 \ll 1$. Thus, the couplings of a massive axion field are not quantized, but to the extent that the mass of the axion is much smaller than all other mass scales in the problem, we expect the deviations from coupling quantization to be small.

It is instructive to compare this to the familiar non-quantized shift of the axion coupling to photons via its mixing with the neutral pion. As explained in [750], this does not violate the shift symmetry (2.1) of the axion because it is part of a set of terms that resum to a periodic function, similar to (2.27). Furthermore, the effect is suppressed by m_a^2, and is large only because the axion mass arises from the same strong dynamics as the pion mass, so that $m_a^2 F_a^2 \sim m_\pi^2 F_\pi^2$. In other words, in this case, the suppression factor m_a^2/m_π^2 that we have argued to exist on general EFT grounds is compensated by an enhancement factor of F_a^2/F_π^2. (One could, in principle, perform a field redefinition to discuss this example in the language of kinetic mixing rather than mass mixing, although because the kinetic mixing would then be nearly maximal, this is not a very useful viewpoint to take.) This example shows that some caution is in order when asserting that effective field theories of very light axions are expected to contain quantized couplings to gauge fields. On the other hand, it also reveals that one needs rather special circumstances to obtain a very large violation of this expectation, as arises when multiple periodic scalars obtain mass simultaneously from the same dynamics, as in QCD confinement.

2.3 Mixing with a Heavier Axion Eaten by a Spin-1 Field

As our next example, we again consider a theory with two axions, but with a linear combination obtaining a mass in a different way: by being eaten by a massive, spin-1 gauge field through the Higgs or Stückelberg mechanism [738, 739]. This type of theory has been considered in great detail in [740, 741]. A version of it in a Randall-Sundrum scenario was recently discussed in [763]. In this scenario, we will again see intermediate results that seem to break the expected connection between periodic scalar fields and quantized couplings. In this case, the pattern will be reversed from what we observed in §2.2.1: the heavy axion will be a *non*-periodic field, but will have quantized couplings; on the other hand, the light axion will be a periodic field, but will have non-quantized couplings. These non-quantized couplings have led to earlier claims that super-Planckian

field ranges can be obtained in models of this type [740, 741, 763]. However, our results do not support such claims. Once again, a careful assessment of the underlying gauge invariance of the theory will show that the proper understanding of the low-energy EFT is that of a periodic field with quantized couplings, despite initial appearances. Our conclusions are in accord with those of [742], which appeared while this work was being completed.

2.3.1 Diagonalizing the propagating states

Let's begin by looking at an effective theory with two axions, one combination of which is eaten to provide a mass to a spin-1 field via the Stückelberg mechanism. (It is possible to obtain this effective theory as a limiting case of a Higgs mechanism, so we expect our remarks to apply to both scenarios.) For the time being, we will neglect kinetic mixing, as the points we wish to illustrate do not depend on it. We begin by considering the action

$$\sum_{i=1}^{2} \frac{1}{2} F_i^2 (\partial_\mu \theta_i - q_i A_\mu)^2 - \frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4g^2} G_{\mu\nu} G^{\mu\nu} + \frac{k_1 \theta_1 + k_2 \theta_2}{16\pi^2} G_{\mu\nu} \widetilde{G}^{\mu\nu} + \mathcal{L}_{\rm con}, \tag{2.28}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength of the massive gauge field, whereas $G_{\mu\nu}$ is the field strength of a different, massless gauge field G_μ. We are interested in the quantization of the $\theta G\widetilde{G}$ coupling for the light axion. The term $\mathcal{L}_{\rm con}$ denotes additional couplings that, in some cases, may be necessary for consistency of the theory. We will discuss these couplings in more detail below.

The subtleties in this case, compared to our previous case, arise because we now must ensure invariance under *three* different gauge transformations that shift the axions. These are the two discrete shift symmetries $\theta_i \mapsto \theta_i + 2\pi$ associated with the periodicities of the axions, together with a continuous shift symmetry associated with the U(1) group gauged by A_μ:

$$A_\mu \mapsto A_\mu + \partial_\mu \alpha, \quad \theta_1 \mapsto \theta_1 + q_1 \alpha, \quad \theta_2 \mapsto \theta_2 + q_2 \alpha. \tag{2.29}$$

When studying the theory on a spacetime of nontrivial topology, $e^{i\alpha(x)} \in U(1)$ must be well-defined (single-valued) but α itself need not be. Because $e^{i\theta_i(x)}$ must also be single-valued, we see that the gauge transformation (2.29) makes sense only if $q_1, q_2 \in \mathbb{Z}$. This is consistent with our expectations if the axions θ_i arise as phases of complex fields of charge q_i that obtain a vacuum expectation value, in the case that (2.28) arises as a limit of the Higgs mechanism.

Consistency of the theory under the axion shift symmetries imposes that $k_1, k_2 \in \mathbb{Z}$ in the absence of additional interactions, just as in our earlier discussions. However, notice that in general the $\theta G\widetilde{G}$ terms are

not invariant under the U(1) gauge transformation (2.29), which shifts the Lagrangian by

$$\delta_\alpha \mathcal{L}_{\theta G\widetilde{G}} = \frac{k_1 q_1 + k_2 q_2}{16\pi^2} \alpha G_{\mu\nu} \widetilde{G}^{\mu\nu}. \tag{2.30}$$

Because α is a continuous quantity, the theory only respects the U(1) gauge symmetry if $k_1 q_1 + k_2 q_2 = 0$. Otherwise, it is necessary to add additional terms, indicated by $\mathcal{L}_{\text{con}}$ above, which are not gauge invariant on their own but which serve to cancel the gauge variation (2.30). Such terms could arise from fermions that carry G charge and transform anomalously under the U(1), or from generalized Chern-Simons terms proportional to $A_\mu K^\mu$ where $\partial_\mu K^\mu = G_{\mu\nu}\widetilde{G}^{\mu\nu}$ [739–741].

For the moment, let us leave $\mathcal{L}_{\text{con}}$ unspecified and proceed to diagonalize the propagating states in (2.28). We can change basis to diagonalize the kinetic terms,

$$\mathcal{L} \supset \frac{1}{2} m_A^2 (\partial_\mu a_H - A_\mu)(\partial^\mu a_H - A^\mu) + \frac{1}{2} m_A^2 \partial_\mu a_L \partial^\mu a_L, \tag{2.31}$$

where the "heavy axion," the linear combination eaten by the U(1) gauge boson with mass m_A, is

$$a_H := \frac{1}{m_A^2} \left[F_1^2 q_1 \theta_1 + F_2^2 q_2 \theta_2 \right], \quad \text{where} \quad m_A^2 := F_1^2 q_1^2 + F_2^2 q_2^2. \tag{2.32}$$

The orthogonal light combination is

$$a_L := \frac{F_1 F_2}{m_A^2} \left(q_2 \theta_1 - q_1 \theta_2 \right). \tag{2.33}$$

The proportionality of a_L to an integer linear combination of our original axion fields is no accident; it guarantees that a_L is invariant under the U(1) gauge transformation (2.29). In this basis, the couplings of the propagating axion eigenstates to the gauge field G are

$$\mathcal{L} \supset \frac{1}{16\pi^2} \left[\left(k_1 q_2 \frac{F_2}{F_1} - k_2 q_1 \frac{F_1}{F_2} \right) a_L + (k_1 q_1 + k_2 q_2)\, a_H \right] G_{\mu\nu} \widetilde{G}^{\mu\nu}. \tag{2.34}$$

These results have already been obtained in [740, 741], but let us discuss them from the point of view of gauge invariance, periodicity, and quantized couplings. The fact that the coupling of a_H is proportional to $k_1 q_1 + k_2 q_2$ follows from (2.30): when the Lagrangian is gauge invariant without further contributions, i.e., when $\mathcal{L}_{\text{con}} = 0$, U(1) gauge anomaly cancellation demands that the linear combination of axions that transforms under U(1) decouples from the other gauge fields.

Recall that in §2.2.1, we faced a puzzle: after diagonalizing the mass and kinetic mixings, we found a heavy propagating axion mass eigenstate that was a periodic scalar and yet had non-quantized couplings,

and a massless light axion that was not a periodic scalar and yet had quantized couplings. Here we seem to see exactly the opposite situation: the heavy linear combination a_H is not a periodic scalar, and yet its coupling to $G\widetilde{G}$ is proportional to the integer $k_1q_1 + k_2q_2$.[1] On the other hand, the light axion a_L is periodic; from (2.33) we can see that under a general shift of the underlying fields $\theta_i \mapsto \theta_i + 2\pi n_i$, the shift of a_L is proportional to the integer $q_2n_1 - q_1n_2$. Thus, a_L is periodic, with minimal period given by the identification

$$a_L \cong a_L + 2\pi \frac{F_1F_2}{m_A^2} \gcd(q_1, q_2). \tag{2.35}$$

The puzzle is that despite this periodicity, the couplings of a_L do not seem to be quantized, as the $G\widetilde{G}$ coupling depends not only on the integer charges k_i, q_i, but also on the ratio of decay constants F_1/F_2. In particular, if we define a scalar field θ_L of period 2π by rescaling a_L,

$$\theta_L := \frac{1}{\gcd(q_1, q_2)}(q_2\theta_1 - q_1\theta_2), \tag{2.36}$$

its coupling to $G\widetilde{G}$ is given by

$$\frac{1}{16\pi^2}\left[\gcd(q_1, q_2)\left(\frac{k_1q_2F_2^2 - k_2q_1F_1^2}{q_1^2F_1^2 + q_2^2F_2^2}\right)\right]\theta_L G_{\mu\nu}\widetilde{G}^{\mu\nu}, \tag{2.37}$$

where the term in brackets is, in general, *not* an integer. This appears to contradict the basic quantization (2.3) of a 2π-periodic axion.

We emphasize that the periodicity of θ_L is entirely determined by the original periodic lattice of identifications of (θ_1, θ_2) together with the charges q_i, which specify which linear combination of the fields remains uneaten. We can write the kinetic term of θ_L as

$$\frac{1}{2}F_L^2\partial_\mu\theta_L\partial^\mu\theta_L, \quad \text{where} \quad F_L = \frac{F_1F_2\gcd(q_1, q_2)}{m_A}. \tag{2.38}$$

The scale F_L is the most natural definition of the "decay constant" of the light axion, and determines the units in which the couplings of the canonically normalized axion to $G\widetilde{G}$ are expected to be quantized as well as the expected field range when an axion potential is generated. This should be contrasted with the approach of [740, 741, 763], which defines an effective decay constant F_{eff} which is inversely proportional to the factor in brackets in (2.37). In those studies, a small value of the factor $k_1q_2F_2^2 - k_2q_1F_1^2$ is argued to suppress the coupling and lead to a trans-Planckian F_{eff}. While it is interesting that the coupling in (2.37) allows for a very large F_{eff} defined in this way, the fact that it is not related to the period $2\pi F_L$ appearing

[1] This is a bit of an overstatement: in the presence of $\mathcal{L}_{\text{con}}$, as noted above, the k_i need not be integers, whereas in the absence of $\mathcal{L}_{\text{con}}$, this coupling is not just any integer, but zero.

in (2.38) should give us pause. In fact, the axion decay constant as we have defined it can only be smaller than the decay constants we started with:

$$\mathrm{F_L} < \min(\mathrm{F_1}, \mathrm{F_2}). \tag{2.39}$$

How do we reconcile this with claims of large $\mathrm{F_{eff}}$ extracted from (2.37)?

We have already laid the groundwork for the resolution of this puzzle: because the Lagrangian (2.28) is *not*, in general, gauge-invariant in the absence of additional terms $\mathcal{L}_{\mathrm{con}}$, we should not be surprised that it violates the expected periodicity properties. The physical coupling of an axion to gauge fields is quantized in units that allow us to read off the maximum field range of the axion potential, but in theories with $\mathcal{L}_{\mathrm{con}} \neq 0$, Lagrangian couplings like that in (2.37) do not determine the full amplitude, and consequently we do not expect that the scale $\mathrm{F_{eff}}$ extracted from such a term is related to a physical field range. The discussion in the introduction makes this clear: if we read off $\mathrm{F_{eff}} \propto 1/\mathrm{c_F}$ from (2.4), then because $\mathrm{c_F}$ shifts as in (2.6) under a field redefinition, we could obtain absolutely any value of $\mathrm{F_{eff}}$ by parametrizing our fields in a different way. The physical amplitude which is quantized, in the presence of anomalous fermions, depends on a combination of terms like (2.7). Only by first redefining the fermions to set $\mathrm{c_m} = 0$ (which, in this context, is the meaning of $\mathcal{L}_{\mathrm{con}} = 0$) do we obtain quantized $\mathrm{c_F}$, at which point we can read off the axion periodicity from this coupling. Hence, we cannot, in general, analyze the periodicity constraints on the effective action of θ_{L} without specifying the terms $\mathcal{L}_{\mathrm{con}}$, which we expect will always resolve the puzzle. The only case in which we can directly resolve the puzzle is in the case when it is consistent to set $\mathcal{L}_{\mathrm{con}} = 0$ because $\delta_\alpha \mathcal{L}_{\theta \mathrm{G}\widetilde{\mathrm{G}}}$ in (2.30) is identically zero, i.e., the case $\mathrm{k_1 q_1} + \mathrm{k_2 q_2} = 0$. In this case, the bracketed factor in (2.37) reduces to

$$\left[\gcd(\mathrm{q_1}, \mathrm{q_2}) \left(\frac{\mathrm{k_1 q_2 F_2^2} - \mathrm{k_2 q_1 F_1^2}}{\mathrm{q_1^2 F_1^2} + \mathrm{q_2^2 F_2^2}} \right) \right] \mapsto -\frac{\mathrm{k_2}}{\mathrm{q_1}} \gcd(\mathrm{q_1}, \mathrm{q_2}) \in \mathbb{Z}. \tag{2.40}$$

To justify the claim that this is an integer: given that $\mathrm{k_1 q_1} = -\mathrm{k_2 q_2}$, it follows that $\mathrm{q_1} | (\mathrm{k_2 q_2})$. In order for this to be true, $\mathrm{q_1}/\gcd(\mathrm{q_1}, \mathrm{q_2})$ must divide $\mathrm{k_2}$.

So far we have assumed the light axion to be exactly massless, and found that it has exactly quantized couplings. We could also consider a theory which has a potential that provides a mass for the light axion well below the mass of the heavy spin-1 field. Just as we discussed in §2.2.3, the effective field theory of the light axion allows for terms proportional to $\Box\theta_{\mathrm{L}}$ which, upon making use of the equations of motion, can appear as effectively non-quantized couplings proportional to the light axion mass squared.

Summing up: when we give one linear combination of the axions a mass through the Higgs or Stückelberg mechanism, the massless light axion is a periodic field, with smaller field range than our initial axions. In the case that the Lagrangian we have studied is gauge invariant in its own right, we have shown that the

couplings of this periodic field are quantized, just as we expect them to be. This is as it must be; if we integrate out the heavy fields, we obtain an effective field theory of a periodic axion, with all of the constraints that this entails. Nonetheless, to illustrate the point more generally, let us look at at an example in which $\mathcal{L}_{\rm con} \neq 0$. Specifically, we will consider a theory in which light fermions cancel the gauge variation (2.30).

2.3.2 Analyzing a 4d UV completion

To clarify the physics, it is useful to consider an explicit, 4d UV completion of the effective Lagrangian (2.28) in which the massive gauge field obtains a mass from the Higgs mechanism, and fermion fields supply a non-vanishing contribution to $\mathcal{L}_{\rm con}$. The goal of this model is simply to show a consistent example that generates the effective theory we are interested in, in which we can explicitly calculate the interactions and understand how the constraints of axion periodicity are respected. This model is not meant to be natural or aesthetically appealing, just to illustrate some points about the physics of axions. For this reason, we will freely assume hierarchies in the dimensionless couplings, and invoke global symmetries that are not necessarily accidental, with no need for further explanation.

In this model, the axions $\theta_{1,2}$ arise from the phases of two complex scalars $\varphi_{1,2}$ with U(1) gauge charges $q_{1,2}$. We also consider an SU(N) gauge group that will provide the $G\widetilde{G}$ couplings we are interested in. Each of the scalars will provide Dirac masses to some fermions $Q, \widetilde{Q}$ which transform in non-trivial, conjugate SU(N) representations, so that from the SU(N) point of view the theory is not chiral. However, these fields will have chiral couplings to U(1): Q carries charge and $\widetilde{Q}$ does not, or vice versa. To cancel the $U(1)^3$ and mixed U(1)–gravitational anomalies, we also introduce fermions $L, \widetilde{L}$ that have the opposite U(1) charge assignments but do *not* interact with SU(N) gauge fields (though they come in the appropriate number of copies to compensate for the anomalies of the $Q, \widetilde{Q}$ fields). By construction, this theory has no $SU(N)^3$, $U(1)^3$, or mixed gravitational anomaly, but the $SU(N)^2U(1)$ mixed anomaly still imposes a nontrivial constraint on the representations and charged assignments, to which we will return shortly. The field content of this model is summarized in Table 2.1.

The Lagrangian (2.28) can be obtained in a decoupling limit of this model. We begin with the complete theory, including Yukawa couplings

$$\mathcal{L}_{\rm Yuk} = \sum_{\rm i} \left(y_{1\rm Qi} \varphi_1^\dagger Q_{1\rm i} \widetilde{Q}_{1\rm i} + y_{1\rm Li} \varphi_1 L_{1\rm ik} \widetilde{L}_{1\rm ik} \right) + \sum_{\rm j} \left(y_{2\rm Qi} \varphi_2^\dagger Q_{2\rm j} \widetilde{Q}_{2\rm j} + y_{2\rm Lj} \varphi_2 L_{2\rm ik} \widetilde{L}_{2\rm ik} \right) + \text{h.c.}, \tag{2.41}$$

where sums over the copies k of the L fields are implicit. To generate an effective Lagrangian of the form (2.28), we suppose that there is a hierarchy among the Yukawa couplings so that some are much larger than others. Then below the symmetry breaking scale, we can integrate out the heavy fermions. In general,

	φ_1	φ_2	Q_{1i}	$\widetilde{Q}_{1i}$	L_{1ik}	$\widetilde{L}_{1ik}$	Q_{2j}	$\widetilde{Q}_{2j}$	L_{2jk}	$\widetilde{L}_{2jk}$
$U(1)_{\text{gauge}}$	q_1	q_2	q_1	0	$-q_1$	0	0	q_2	0	$-q_2$
$SU(N)$	1	1	R_{1i}	$\overline{R}_{1i}$	1	1	R_{2j}	$\overline{R}_{2j}$	1	1
$U(1)_{\text{global}}$	1	0	1	0	0	-1	0	0	0	0
N_{copies}	1	1	1		$\dim(R_{1i})$		1		$\dim(R_{2j})$	

Table 2.1: Matter field content in a potential UV completion of the two axion model. The integers $i \in \{1, \ldots, N_1\}$ and $j \in \{1, \ldots, N_2\}$ label the set of fields, while the subscripts 1 and 2 signal which Yukawa couplings provide mass to the fields, e.g., $\varphi_1 Q_{1i} \widetilde{Q}_{1i}$. The full set of Yukawa couplings is displayed in (2.41). The φ fields are scalars, whereas the $Q, \widetilde{Q}, L, \widetilde{L}$ fields are all left-handed Weyl fermions. The $L, \widetilde{L}$ fields come in multiple copies, $k \in \{1, \ldots, N_{\text{copies}}\}$, to ensure anomaly cancellation. With these charge assignments, the only anomaly cancellation condition that must be explicitly checked is the $SU(N)^2U(1)$ anomaly.

integrating out a term of the form $m(\varphi)\Psi\widetilde{\Psi} + \text{h.c.}$ produces a term of the form

$$\Delta\mathcal{L} = \frac{2\mu(R_\Psi)}{32\pi^2} \arg(m) G^a_{\mu\nu} \widetilde{G}^{a\mu\nu}, \tag{2.42}$$

where $\mu(R_\Psi)$ is the Dynkin index of the representation of Ψ under the group G. For concreteness, let us suppose that the fields with $i = 1$ and $j = 1$ are relatively heavy, whereas all of the others are much lighter (i.e., have much smaller Yukawa couplings to φ_1, φ_2). We further assume that the U(1) gauge coupling e is small enough that we can integrate out the heavy fermions *without* integrating out the massive gauge field, i.e., $eq_{1,2} \ll y_{1Q1}, y_{1L1}, y_{2Q1}, y_{2L1}$. We further assume that the fields $\varphi_{1,2}$ have a symmetry breaking potential which does not mix them, e.g.,

$$V_{\text{SSB}} = \frac{\lambda_1}{4}\left(|\varphi_1|^2 - v_1^2\right)^2 + \frac{\lambda_2}{4}\left(|\varphi_2|^2 - v_2^2\right)^2. \tag{2.43}$$

The structure of this potential ensures that, when we turn off the U(1) gauge interaction, we have two distinct Nambu-Goldstone bosons $\theta_{1,2}$ which are the phases of $\varphi_{1,2}$ respectively. An example of a U(1) global symmetry charge assignment that can be responsible for protecting the uneaten Nambu-Goldstone boson is given in the "$U(1)_{\text{global}}$" row of Table 2.1. We further assume that the radial modes of the φ fields are sufficiently heavy that we can integrate them out, i.e., $eq_{1,2} \ll \sqrt{\lambda_{1,2}}$. The choice of which fields to integrate out is not unique, but making this arbitrary choice suffices to illustrate our main points. We illustrate the various interesting ranges of energies, and corresponding effective field theories, in this model in Fig. 2.2.

Integrating out the heavy SU(N)-charged fermions $Q_{11}, \widetilde{Q}_{11}, Q_{21}, \widetilde{Q}_{21}$ will generate couplings

$$\mathcal{L}_{\theta G\widetilde{G}} = -\frac{1}{16\pi^2}\left[\mu(R_{11})\theta_1 + \mu(R_{21})\theta_2\right] G^a_{\mu\nu}\widetilde{G}^{a\mu\nu}. \tag{2.44}$$

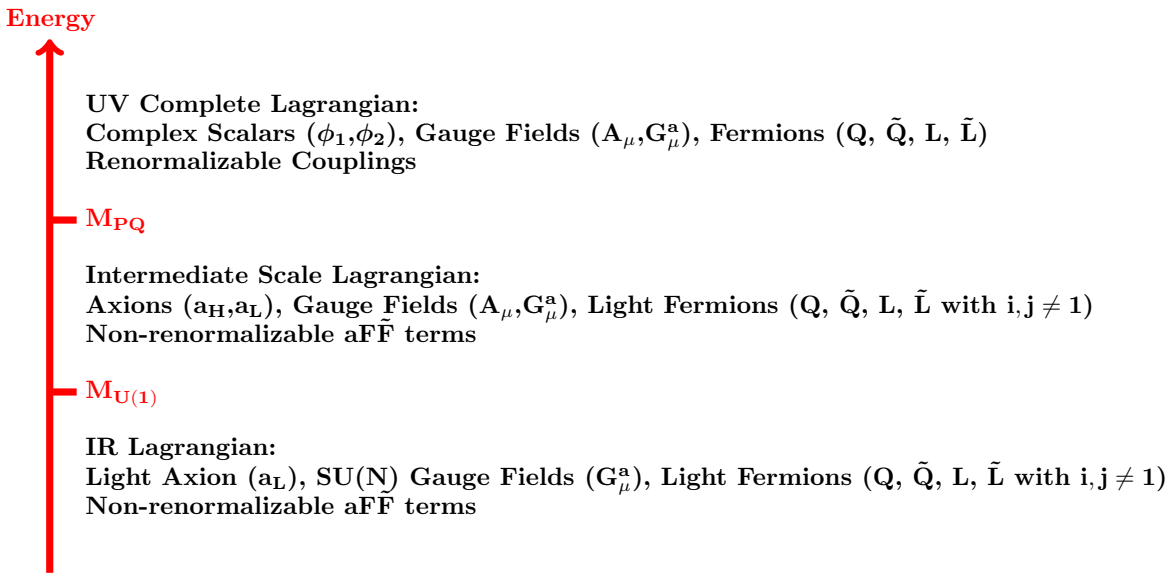


Figure 2.2: Schematic of important energy scales and effective field theories obtained from our UV completion. After integrating out the radial modes of the scalars and the heavy fermions at the Peccei-Quinn scale M_{PQ}, we obtain an effective field theory of two axions, one eaten by a spin-1 field, as considered in §2.3.1. Below the mass scale $M_{U(1)}$ of the spin-1 field, an effective field theory of a single light axion is obtained. We show that this axion is a periodic field with quantized couplings.

We follow the standard convention in particle physics that the Dynkin index of the fundamental representation of SU(N) is $\mu(\square) = 1/2$, in which case the Dynkin index of any representation R satisfies $2\mu(R) \in \mathbb{Z}$, which shows that the couplings of $\theta_{1,2}$ are quantized in the way that we expect. Changing basis as described in §2.3.1, this includes a coupling of the light axion θ_L of the form

$$-\frac{\gcd(q_1,q_2)}{16\pi^2}\frac{1}{m_A^2}\left[\mu(R_{11})q_2F_2^2-\mu(R_{21})q_1F_1^2\right]\theta_L G^a_{\mu\nu}\tilde{G}^{a\mu\nu}. \tag{2.45}$$

While it appears that we can choose this to be as small as we like by carefully choosing representations to impose relations among the q_l and $\mu(R_{mn})$, we have not yet taken into account gauge invariance. The condition for $SU(N)^2U(1)$ anomaly cancellation, given the field content in Table 2.1, is

$$\sum_{i=1}^{N_1}\mu(R_{1i})q_1+\sum_{j=1}^{N_2}\mu(R_{2j})q_2=0. \tag{2.46}$$

We can use this condition to eliminate $\mu(R_{11})$ from (2.45), obtaining a coupling

$$\begin{aligned}\mathcal{L}_{\theta_L G\tilde{G}} =& -\frac{\gcd(q_1,q_2)}{16\pi^2}\frac{1}{m_A^2}\left[-\sum_{i=2}^{N_1}\mu(R_{1i})q_2F_2^2-\sum_{j=1}^{N_2}\mu(R_{2j})\frac{q_2^2}{q_1}F_2^2-\mu(R_{21})q_1F_1^2\right]\theta_L G^a_{\mu\nu}\tilde{G}^{a\mu\nu}\\ =&\frac{\gcd(q_1,q_2)}{16\pi^2}\left[\frac{\mu(R_{21})}{q_1}+\sum_{i=2}^{N_1}\mu(R_{1i})q_2\frac{F_2^2}{m_A^2}+\sum_{j=2}^{N_2}\mu(R_{2j})\frac{q_2^2}{q_1}\frac{F_2^2}{m_A^2}\right]\theta_L G^a_{\mu\nu}\tilde{G}^{a\mu\nu}.\end{aligned} \tag{2.47}$$

Notice that we have now written the coupling in terms of a single term that depends on one of the heavy fields, together with a sum over *only* the light fields (i.e., the sums omit $i = 1$ and $j = 1$). The first term in brackets in (2.47) is a rational number, while the others are, in general, irrational. However, recall that this is not unexpected: there are additional light fermions in the theory, labeled by $i \in \{2, \ldots, N_1\}$ and $j \in \{2, \ldots, N_2\}$. The quantization condition applies only to a combination of coefficients like (2.7), which depends on how the light fermions couple to the axion. By either performing an anomalous field redefinition to eliminate the $\theta_L \Psi \widetilde{\Psi}$ couplings, or computing the one-loop triangle diagram contribution to the $\theta_L G\widetilde{G}$ amplitude, we find that the light fermion contributions cancel the irrational pieces of the terms in (2.47) that arise from the sum over i and j. To compute these contributions, we note that the masses of the Q_{1i} and Q_{2j} fields are proportional to

$$\begin{aligned}
\varphi_1^\dagger &\sim \exp(-i\theta_1) = \exp\left[-i\left(q_1 a_H + \frac{q_2 \gcd(q_1, q_2) F_2^2}{m_A^2}\theta_L\right)\right], \\
\varphi_2^\dagger &\sim \exp(-i\theta_2) = \exp\left[-i\left(q_2 a_H - \frac{q_1 \gcd(q_1, q_2) F_1^2}{m_A^2}\theta_L\right)\right],
\end{aligned} \tag{2.48}$$

where we have changed to the basis of heavy and light fields. As a result, if we eliminate the θ_L couplings to the light fermions, we produce new contributions to the $\theta_L G\widetilde{G}$ coupling,

$$\Delta\mathcal{L}_{\theta_L G\widetilde{G}} = -\frac{\gcd(q_1, q_2)}{16\pi^2}\frac{1}{m_A^2}\left[\sum_{i=2}^{N_1}\mu(R_{1i}) q_2 F_2^2 - \sum_{j=2}^{N_2}\mu(R_{2j}) q_1 F_1^2\right]\theta_L G^a_{\mu\nu}\widetilde{G}^{a\mu\nu}. \tag{2.49}$$

The first of the new terms cancels the middle term in brackets in (2.47), while the second term combines with the last term in brackets in (2.47) and simplifies:

$$\begin{aligned}
\mathcal{L}_{\theta_L G\widetilde{G}} + \Delta\mathcal{L}_{\theta_L G\widetilde{G}} &= \frac{\gcd(q_1, q_2)}{16\pi^2}\left[\frac{\mu(R_{21})}{q_1} + \sum_{j=2}^{N_2}\mu(R_{2j})\left(\frac{q_2^2}{q_1}\frac{F_2^2}{m_A^2} + \frac{q_1 F_1^2}{m_A^2}\right)\right]\theta_L G^a_{\mu\nu}\widetilde{G}^{a\mu\nu} \\
&= \frac{1}{32\pi^2}\left[\frac{2\gcd(q_1, q_2)}{q_1}\sum_{j=1}^{N_2}\mu(R_{2j})\right]\theta_L G^a_{\mu\nu}\widetilde{G}^{a\mu\nu}.
\end{aligned} \tag{2.50}$$

Now we have finally obtained a manifestly quantized coupling, as we expect for a periodic axion. We can argue that the term in brackets is an integer in precisely the same way that we argued following (2.40), once we make use of (2.46) and the aforementioned integer quantization of $2\mu(R)$.

2.4 Mixing with a Heavier Non-compact Scalar

In this section, we will study examples in which an axion mixes with a non-compact scalar. As in our previous examples, our purpose is to study the periodicity of the light axion after decoupling the heavy field.

In our first example, we consider mixing of the light axion with a radial mode of the same complex field. In the second example, we consider mixing of an ordinary axion with a monodromy axion. For concreteness, we consider an extra-dimensional realization of monodromy in which the two axions are the Wilson loop phases of two different five dimensional gauge fields obtained after compactification on $\mathbb{R}^{3,1} \times S^1$. One has been Higgsed ($H^\mu$), the other remains massless (A^μ), and both couple to the same charged bulk scalar. While there is extensive literature on both the one-loop potential (e.g., [771–777]) and axion monodromy (e.g., [733, 734, 743, 778–781]), we highlight features of their interplay which have not previously been emphasized in the literature and use them to show our broader conclusions still hold in a more general setting. In both of the examples we consider, if the non-compact field is much heavier than the axion, we find that we can integrate it out to obtain a typical EFT of the light axion. In the case of mixing with a monodromy axion, we find that in the limit where the monodromy potential is subdominant to a periodic potential for a linear combination of the ordinary and monodromy axions, the monodromy is effectively "realigned" to the surviving light axion in the EFT, which has a larger decay constant than the original monodromy axion. In every case, we find that deviations of $\theta G\widetilde{G}$ couplings from their quantized values are, as before, proportional to the mass squared of the axion field.

The case of mixing with a monodromy axion that we discuss is related to an earlier discussion in [765], in which certain axions obtain masses via fluxes (which makes them monodromy axions) and other axions remain light. That paper emphasized that the light axions can have enhanced field ranges, providing an implementation of alignment [736] in which the heavy mode is decoupled by fluxes rather than a periodic potential. Our claims are in accord with theirs, but we consider an extended range of possibilities including the scenario when a periodic potential provides a larger mass term than a monodromy potential.

2.4.1 Mixing with a Radial Mode

As our first example of mixing with a non-compact scalar, we consider a simple KSVZ UV completion of a single axion [592, 593] and add at least one PQ-breaking term:

$$\mathcal{L} = \lambda(|\varphi|^2 - v^2)^2 + \left(y\varphi Q\widetilde{Q} + \text{h.c.}\right) + \left(\frac{z\varphi^N}{\Lambda^{N-4}} + \text{h.c.}\right). \tag{2.51}$$

The presence of the PQ-breaking term allows, when perturbing around a generic point in field space, for the radial and angular modes of φ to mix with each other. (The Yukawa term also allows this, after confinement.) An example of this potential for a particular choice of parameters is shown in Fig. 2.3. Our purpose in studying this theory is to understand whether it can produce a non-compact scalar field after integrating out the radial mode. We find that the answer is no, because even before integrating out the radial mode, we see that there is a nearly-flat, periodic valley at the minimum of the potential.

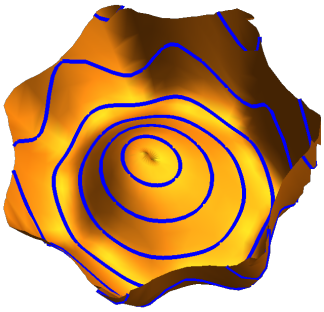


Figure 2.3: An example potential from mixing an axion with a radial Higgs mode through a KSVZ-like model, including additional PQ breaking terms. Lines of constant r (defined to be the radial mode of φ) are shown in blue. Both the quark couplings and the PQ-breaking terms generate a potential along the θ (the phase of φ) direction, but are unable to produce a non-compact valley at the minimum of the potential.

The potential, within the UV theory of two real scalar fields, has a periodic valley because of the form of the radial dependence of each of the contributions to the potential. In order for the valley to unwind into a non-compact flat direction, a cross section of the potential at fixed theta would have to oscillate as a function of the radial mode. However, in each of the contributions to the potential in this example, the radial dependence is polynomial. In general, models of this form will generate potentials that are a sum of periodic functions of theta each multiplied by an envelope function that is a polynomial in the radial model. This means that while PQ-breaking terms can generate complicated radial dependence for the precise location of the minimum of the potential, they cannot make the valley unwind into a non-compact direction without fine-tuning the coefficients of the radial polynomials to approximate a periodic function.

2.4.2 Mixing with a Monodromy Axion

The simple four-dimensional theory that we considered in the previous subsection is not sufficient to allow the valley at the base of the potential to become a non-compact direction. Our discussion suggests that this is more likely to occur in a potential that mixes the radial and angular modes *inside* a periodic function. To generate this type of potential, we will consider the case where a monodromy axion mixes with an ordinary, compact axion. In this model, the nearly-flat valley at the base of the potential is non-compact because of mixing with the non-compact monodromy axion. On the other hand, there is still an underlying periodicity, which is reflected in quantized $a G\widetilde{G}$ couplings. The mixing between one monodromy axion and a second compact axion can lead to a low energy EFT of a monodromy axion which is lighter and has a larger field range than the original monodromy axion on its own. This type of potential and string theory completions have been discussed previously in [743]. Instead of a string theory construction, we will instead illustrate the core concepts using a simpler mechanism for generating the potential from dimensional reduction of a

higher dimensional QFT.

We consider the simplest toy example that has this effect: a five dimensional theory of two U(1) gauge fields, where one has been Higgsed (H^μ) and the other remains massless (A^μ). The use of massive U(1) gauge fields in higher dimensions to produce monodromy axions in a compactified theory has been discussed previously in [781–783]. After compactifying the fifth dimension on a circle of radius R, we obtain axions as the Wilson loop phases $\theta_i := \oint dx^5 G_{5i}$ of the fifth component of each gauge field around the compactified dimension. We can see that the axion is normalized here to have period 2π because G_{5i} and $G_{5i} + \frac{1}{R}$ are related by a large gauge transformation. The Higgsed field alone will generate a monodromy potential for θ_H. To generate a periodic potential that mixes θ_H and θ_A, we couple both gauge fields to the same form of matter. For simplicity, we take this matter to be a massless scalar and take the 5d spin-1 field to have a simple Stückelberg mass term, but these choices do not qualitatively change our results. (In particular, our qualitative conclusions should carry over to the other shapes of monodromy potentials that are known to arise in string models, e.g., [733, 779, 784].) The action in this theory takes the form

$$S = \int d^5x \left(-\frac{1}{4g_{5H}^2} H_{MN}(x) H^{MN}(x) - \frac{m^2}{2g_{5H}^2} \mathcal{H}_\mu \mathcal{H}^\mu - \frac{1}{4g_{5A}^2} A_{MN}(x) A^{MN}(x) + D_M \chi^\dagger(x) D^M \chi(x) \right) \tag{2.52}$$

where the covariant derivative is

$$D_M \chi(x) := \mathbf{p}_M \chi(x) - i q_A A_M(x) \chi(x) - i q_H H_M(x) \chi(x) \tag{2.53}$$

and following [782] we have defined

$$\mathcal{H}_M(x) := H_M(x) - i e^{i\theta(x)} \mathbf{p}_M e^{-i\theta(x)}, \tag{2.54}$$

where the Stückelberg field $\theta(x)$ is a periodic scalar. Since θ is an angular variable it can have nontrivial winding around the extra dimension, $\frac{w x^5}{R}$ for integer w, which is responsible for the monodromy after compactification.

The potential obtained after compactification contains two distinct contributions. At tree level, we only see the monodromy potential of the Higgsed gauge field from the mass terms [782]

$$\mathcal{L}_4 \supset -V_{mon}(\theta_H) := -\frac{m^2}{2g_{4H}^2 R^2} \left(\frac{\theta_H}{2\pi} - w \right)^2 = -\frac{1}{2} m^2 F_H^2 \left(\theta_H - 2\pi w \right)^2, \tag{2.55}$$

where we have defined the 4d gauge couplings $g_{4i} = g_{5i}/\sqrt{2\pi R}$ as well as the decay constants of the 4d axion fields, $F_i = 1/(2\pi g_{4i} R)$. Since the kinetic terms are $\frac{1}{2} F_i^2 (\partial \theta_i)^2$, we see that m is the canonically normalized mass of θ_H. As is typical with monodromy, for the Lagrangian to remain invariant under a shift by the axion period, we must also shift w. On any given branch of fixed w, the potential is effectively not periodic, so

θ_H behaves as a non-compact scalar. In addition to the tree level potential, both gauge fields get one-loop potentials from their couplings to matter. Since they are coupled to the same form of matter, the one loop potential will be a periodic potential that mixes the monodromy axion with the ordinary axion. In particular, the potential generated by integrating out the mass terms for the tower of scalar Fourier modes

$$\chi^{\dagger(n)} \frac{1}{R^2} \left(n - q_A \frac{\theta_A}{2\pi} - q_H \frac{\theta_H}{2\pi} \right)^2 \chi^{(n)} \tag{2.56}$$

will simply be a sum of cosines in the case where χ is massless (e.g., [772]),

$$V_{per}(\theta_A, \theta_H) = -\frac{3}{64\pi^6 R^4} \sum_{n=1}^{\infty} \frac{\cos(n q_A \theta_A + n q_H \theta_H)}{n^5}. \tag{2.57}$$

In the case where χ is massive the exact form the potential is more complicated [771, 773, 774], but will still be periodic and produce qualitatively the same effect.

Although V_{per} is a one-loop effect and V_{mon} is a tree-level effect in this model, it need not be the case that the monodromy potential dominates. This is because V_{mon} originates from spontaneous breaking of the discrete shift symmetry, which is preserved by V_{per}, so it is of parametrically different (potentially much smaller) size. It is interesting to consider two different limits, one in which $\partial_H^2 V_{mon} > |\partial_H^2 V_{per}|$ throughout the field space, and one with the opposite inequality. (Here ∂_H denotes $\partial/\partial\theta_H$.) These are depicted in the left- and right-hand panels of Fig. 2.4, respectively. The left panel shows the case where the monodromy potential dominates over the periodic potential. The periodic potential creates a small perturbation, but there is no obstruction to any nonzero value of θ_H rolling down the potential toward $\theta_H = 0$. The right panel shows the more interesting case, in which $|\partial_H^2 V_{per}| > \partial_H^2 V_{mon}$. This creates a series of ridges in the potential; it is conceivable that the field could be localized (for instance, during inflation) in a valley between ridges far up the potential, and will evolve toward the minimum by following the winding path down the valley rather than moving directly in the θ_H direction.

The phenomenon exhibited in the case with a ridged potential might be thought of as "monodromy realignment." In the effective theory containing both θ_H and θ_A, it is θ_H that carries the monodromy. This is because the Stückelberg field that produced the monodromy shifted only under shifts of H_μ. Nonetheless, the low-energy effective theory is that of a monodromy axion that is a nontrivial linear combination of θ_H and θ_A. One way to see this is by noting that we could first integrate out the linear combination of fields that obtains a mass from V_{per}. As in §2.2.1, we could choose an alternative lattice basis (θ_1, θ_2) in which this field is θ_1. Specifically, we can find integers r_A, r_H such that

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} p_A & p_H \\ r_A & r_H \end{pmatrix} \begin{pmatrix} \theta_A \\ \theta_H \end{pmatrix}, \tag{2.58}$$

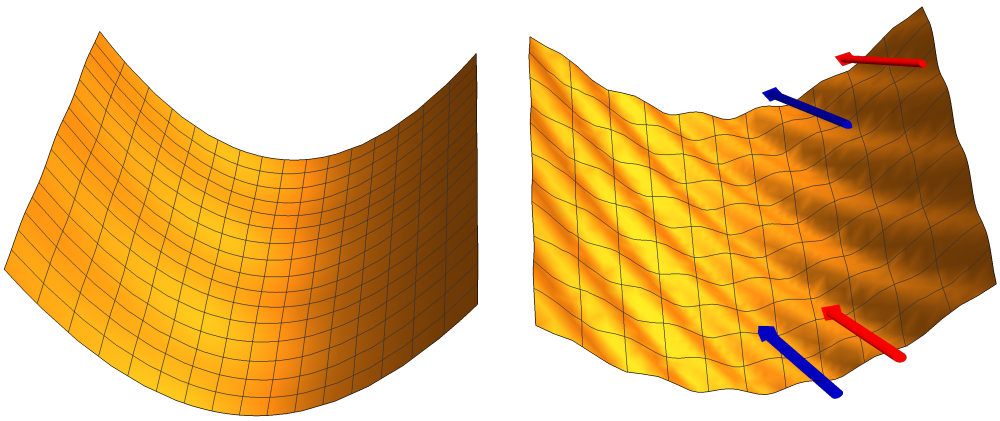


Figure 2.4: The two-axion potential in the case where a monodromy axion θ_H (horizontal axis) mixes with a compact axion θ_A (vertical axis). The front and back edges of the surface correspond to $\theta_A = 0$ and $\theta_A = 2\pi$ and are identified. In the left plot, we have chosen parameters so that $\partial_H^2 V_{mon} > |\partial_H^2 V_{per}|$. In this case, the monodromy axion is heavy and can be integrated out; the light axion θ_A has a smaller cosine potential along the periodic valley at the base of the monodromy axion's potential. In the right plot, the opposite limit $|\partial_H^2 V_{per}| > \partial_H^2 V_{mon}$ is taken. In this case, the cosine potential is large enough to produce a series of ridges. The light axion mode is neither θ_A nor θ_H, but the mode that traverses the valley in between ridges, along which $\theta_A = -q_H\theta_H/q_A \,(\mathrm{mod}\, 2\pi)$. The colored arrows show the path of the minimum down the potential, where arrows of a particular color should be identified together.

where

$$p_A := \frac{q_A}{\gcd(q_A, q_H)}, \quad p_H := \frac{q_H}{\gcd(q_A, q_H)}, \quad \text{and} \quad p_A r_H - p_H r_A = 1. \tag{2.59}$$

In this basis, the potential is (choosing the branch where $w = 0$)

$$V(\theta_1, \theta_2) = V_{per}(\theta_1) + \frac{1}{2} m^2 F_H^2 \left(p_A \theta_2 - r_A \theta_1\right)^2. \tag{2.60}$$

The effective theory along the valley in the potential is obtained by taking $\theta_1 = 0$ (or a 2π shift thereof), so that we can integrate it out to obtain an effective theory of the light field θ_2,

$$\frac{1}{2} F_2^2 \partial_\mu \theta_2 \partial^\mu \theta_2 - \frac{1}{2} m^2 p_A^2 F_H^2 \theta_2^2 + (\text{terms proportional to } \Box\theta_2), \tag{2.61}$$

where, using (2.13), the kinetic term of θ_2 is proportional to

$$F_2^2 = p_A^2 F_H^2 + p_H^2 F_A^2. \tag{2.62}$$

From this we can read off the canonically normalized mass of the light field,

$$m_2^2 = m^2 \frac{p_A^2 F_H^2}{p_A^2 F_H^2 + p_H^2 F_A^2}. \tag{2.63}$$

The nonperiodic potential for θ_2 indicates that, in the low-energy effective theory, it is a monodromy axion; we say that the monodromy has *realigned* from θ_H to $\theta_2 = r_A\theta_A + r_H\theta_H$. Monodromy realignment has both *increased* the effective decay constant and, correspondingly, *decreased* the mass of the monodromy axion. Both of these features are intuitively apparent from the winding valley in Fig. 2.4.

We could also ask if couplings to external gauge fields are quantized the way that we expect them to be. To study this we consider adding Chern-Simons terms to the theory,

$$\mathcal{L}_{CS} = \frac{c_A}{16\pi^2}\varepsilon^{MNPQR}A_M\,\mathrm{Tr}[G_{NP}G_{QR}] + \frac{c_H}{16\pi^2}\varepsilon^{MNPQR}H_M\,\mathrm{Tr}[G_{NP}G_{QR}], \tag{2.64}$$

where G is an arbitrary gauge field (which could be one of the two already in the theory). Gauge invariance requires that the coefficients c_i be integers. After dimensionally reducing, these Chern-Simons terms will contain $\theta_i G_\mu \widetilde{G}^\mu$ couplings of the axions to the four dimensional gauge fields with quantized couplings: gauge invariance required us to start with c_i quantized, and dimensionally reducing won't change that. Just as in earlier sections, the change of lattice basis from (θ_A, θ_H) to (θ_1, θ_2) does not change the quantization of the $\theta G\widetilde{G}$ couplings. However, even though we chose our (θ_A, θ_H) basis to have diagonal kinetic terms (which need not be true, in general), the kinetic terms in the (θ_1, θ_2) basis are generally not diagonal. As in §2.2.1, when we integrate out θ_1, we will generally obtain terms $\propto (\Box\theta_2)G\widetilde{G}$ in the low-energy EFT. When we consider the mass that θ_2 obtains from V_{mod}, these will appear as effectively non-quantized couplings. Just as in our earlier discussion, these contributions are all proportional to the mass parameter m_2^2 of the light axion.

2.5 Non-compact Symmetries Should Not Emerge in the IR

We can summarize our results by saying that if we start with a theory of several axions and, in one way or another, decouple some linear combinations of them while leaving others massless, the massless fields will still be axions, i.e., their field space will be compact and their couplings will be quantized accordingly. In cases where we found non-quantized couplings of a light axion field, we found that the field also obtained a mass, and the deviation of the axion's couplings from their quantized values were proportional to the mass squared of the axion. As we noted in the introduction, this has the same flavor as a well-known fact about gauge theory: if we begin with a compact gauge group and then Higgs it, the surviving infrared gauge group will be compact (and hence will have quantized charges). Such a result is known to hold in many different contexts with compact gauge groups in the UV, in cases where we decouple gauge fields via confinement, via Chern-Simons mass terms in (2+1)d gauge theory, or even when we alter the gauge group entirely in the infrared, as in Seiberg duality. It is also known to be robust against kinetic mixing [744, 745].

Our observations about axions and the corresponding observations about gauge fields are linked in more than a vague qualitative manner. In the case of (2+1)d theories, they are identical, because a massless axion field θ in (2+1)d is Hodge dual to a gauge field A_μ defined by $dA = 2\pi e F_\theta \star d\theta$, where $2\pi F_\theta$ is the distance in field space around the θ circle and e is the gauge field coupling. The scenario discussed in §2.3.1, where θ is eaten to provide a Stückelberg mass to another gauge field B, maps to precisely the case where the gauge field A dual to θ obtains a mass through a mixed Chern-Simons term $B \wedge dA$. The low-energy theory contains a massless gauge field for a compact gauge group with finite coupling, which is dual to a compact axion field.

One reason to expect that a theory with a compact gauge group in the UV flows to a theory with a compact gauge group in the IR is that any effective field theory that contains a non-compact gauge group, such as $\mathbb{R}$, is believed to be inconsistent when coupled to gravity. In such theories, one can generally construct black holes of irrational charge [602], which violate entropy bounds that are believed to be true in all theories of quantum gravity [785]. If it were possible to construct UV theories with compact gauge groups that flow to IR theories with non-compact gauge groups, the UV theory would lie in the Swampland [786]. This would be an interesting new Swampland constraint, but we are unaware of any examples that realize such RG flows.

One possible reason why such RG flows do not exist in general is that they lead to IR theories with a continuum of operators that did not exist in the UV. In theories with a compact gauge group that has an associated p-form gauge field A_p, Wilson line or surface operators of the form $\exp\left(iq \int_\Sigma A_p\right)$, where Σ is a p-dimensional submanifold of spacetime, are defined for discrete choices of charge $q \in \mathbb{Z}$. If the gauge group is $\mathbb{R}$, then there is a *continuum* of well-defined operators with arbitrary q. A similar statement holds for axions: if θ is a 2π-periodic boson, then θ itself is not a well-defined operator, but $\exp(iq\theta)$ for $q \in \mathbb{Z}$ is a sensible local operator. On the other hand, in the non-compact limit, there is no obstruction to constructing such operators for arbitrary $q \in \mathbb{R}$. This suggests a possible general argument against the emergence of either non-compact gauge groups or non-compact bosons from theories with compact gauge groups and axions in the UV: this would be an RG flow from a UV theory with a discrete operator spectrum to an IR theory with a continuous operator spectrum. It seems plausible that such RG flows are forbidden in sensible theories.

In this chapter, we will not go further in attempting to make these suggestions rigorous, but we believe that they point toward a deeper understanding of why our results hold. The properties that arise in many different effective field theories of axions are very closely akin to properties arising in gauge theories, and are likely to be enforced by very general principles of quantum field theory.

2.6 Discussion

Periodicity imposes strong constraints on the axion couplings and field ranges, even in cases where axions mix with other axions or a non-compact scalar. Given our results, it appears the options for generating significantly different axion couplings or field ranges than naively expected are: generating a large integer in the effective theory of a single light axion, as in the clockwork scenario [736, 746–748, 764, 765]; building an effective theory that intrinsically involves multiple axions (e.g., kinetically mixing the axion of interest with an even lighter one); or relaxing these constraints through effects proportional to the mass of the light axion (e.g., realignment of monodromy). While the clockwork scenario has been explored extensively, further studying kinetic mixing with a lighter axion and realignment of monodromy could have potentially interesting phenomenological prospects.

3

Axion Mass from Magnetic Monopole Loops

3.1 Introduction

It is well known that instanton effects can generate a potential for an axion θ [585,586] when it is coupled to a nonabelian gauge field via the topological coupling $\theta\,\mathrm{tr}(F \wedge F)$. Even in the absence of axion interactions with gauge fields, it is known that Euclidean branes can give rise to axion potentials [787–791]. Here, we argue that axions coupled to *abelian* gauge fields through a $\theta F \wedge F$ coupling acquire a potential through an instanton effect whenever there are monopoles magnetically charged under F, due to the Witten effect [792]. Like nonabelian instantons, these effects are associated with 4d gauge theory dynamics. Like Euclidean branes, they occur within a well-behaved semiclassical expansion free of infrared divergences. In fact, we expect that our instantons are continuously connected to, or a limiting case of, known instanton effects in specific UV completions [793]. The virtue of our approach is that, by working from the bottom up, we deduce that such effects must exist [794] even when we do not know the UV theory.[1]

The Completeness Hypothesis postulates that any UV-complete theory of an interacting U(1) gauge field (which has quantized charge) contains magnetic monopoles [796], which break a would-be 1-form global symmetry [458]. This is, in particular, believed to be true of all theories of quantum gravity [602,604,753,797].

[1] The existence of the instantons we discuss here has been noted previously by Jake McNamara [795] and communicated to MR in the course of writing [793], although neither considered computing an axion potential from them at the time.

Assuming the validity of the Completeness Hypothesis, the instanton effect that we discuss will give rise to an effective potential for any axion interacting with photons. This is of great phenomenological interest, since the $\theta F \wedge F$ interaction is the primary target of experimental searches for axions [588, 798–800].

We consider an effective theory of a periodic axion field $\theta \cong \theta + 2\pi$ coupled to a gauge field A normalized such that the allowed Wilson lines $P[\exp(iq\oint A)]$ have integer charge $q \in \mathbb{Z}$:

$$S = \int \left[\frac{1}{2} f^2 d\theta \wedge \star d\theta - \frac{1}{2e^2} F \wedge \star F + \frac{k\theta}{8\pi^2} F \wedge F \right]. \tag{3.1}$$

The axion-gauge field coupling is of Chern-Simons type, with quantized coefficient $k \in \mathbb{Z}$. Through the Witten effect, a magnetic monopole in the presence of a nonzero background θ acquires an electric charge $-k\theta/(2\pi)$. A consistent description of this effect requires that the effective theory on the magnetic monopole worldvolume contains, in addition to the usual translational zero modes x^μ, a collective coordinate interacting with the field θ. This takes the form of a compact scalar boson $\sigma \cong \sigma + 2\pi$, with an action that (expanding around a monopole worldline extended in time) contains [801]:

$$S = \int_\gamma \left[\frac{1}{2} l_\sigma d_A\sigma \wedge \star d_A\sigma + \frac{\theta}{2\pi} d_A\sigma \right], \tag{3.2}$$

where the gauge-covariant derivative $d_A\sigma \equiv d\sigma + kA$ respects a shift of σ under A gauge transformations. The mode σ behaves as a quantum particle on a circle (see, e.g., App. D.1 of [802]). Its energy eigenstates, labeled by integers $n \in \mathbb{Z}$, correspond to dyonic states of the monopole with electric charge $k\,(n - \theta/2\pi)$ and energy

$$E_n = \frac{1}{2l_\sigma} \left(n - \frac{\theta}{2\pi} \right)^2. \tag{3.3}$$

There is a monodromy $n \mapsto n + 1$ when $\theta \mapsto \theta + 2\pi$ that ensures the spectrum of the theory is periodic.

We can estimate l_σ by comparing (3.3) to the energy of the classical field configuration outside a monopole in an axion background, following [803], from which we obtain:

$$l_\sigma \sim \frac{4\pi}{e^2 k^2} r_*, \quad r_* = \max(r_c, r_0), \tag{3.4}$$

where $r_c = \pi/(e^2 m_m)$ is the classical radius of the magnetic monopole (of mass m_m) and $r_0 = ke/(8\pi^2 f)$ is the length scale over which the axion field is screened near the monopole core. In the special case of critical 't Hooft-Polyakov monopoles [804, 805], we begin with an SU(2) gauge theory with coupling g. Matching to (3.1) gives $e = g/2$ and $k = 2$, while matching to (3.2) (when $r_c \gg r_0$) gives $l_\sigma = m_m/m_w^2$ where $m_m = 4\pi v/g$ and $m_w = gv$ is the W boson mass. (We have chosen the order-one coefficient in (3.4) to be accurate for this case, but it will differ in general theories.)

Because the dyon energy spectrum (3.3) is θ-dependent, we can integrate out the dyons and obtain an effective potential for θ. This can be understood either as a sum of Coleman-Weinberg-type potentials [806] from each dyon mode n, or as a sum over loops with nontrivial winding of σ around the loop. These two calculations are related by Poisson resummation, as we explain below. Although there is prior work on the θ potential generated by a gas of (non-virtual) monopoles and antimonopoles (see [803, 807–809] and follow-ups) and similar ideas have been applied to Seiberg-Witten partition functions [810], the effect of monopole loops on the vacuum θ potential is, as far as we know, absent from the prior literature.

3.2 Monopole Loops

We would like to compute the vacuum energy in the presence of "fundamental" magnetic monopoles. Schematically, the vacuum energy should be derived by computing a Euclidean path integral of the form

$$\mathcal{Z}(\theta) = \sum_{\text{worldlines}} \int \mathcal{D}(\text{fields})\, e^{-S_e[\text{fields},\text{worldlines},\theta]}\,, \tag{3.5}$$

and taking the limit of infinite spacetime volume $\mathcal{V}$,

$$V_{\text{eff}}(\theta) = -\lim_{\mathcal{V}\to\infty} \frac{1}{\mathcal{V}} \log \mathcal{Z}(\theta)\,. \tag{3.6}$$

The worldline formalism has previously been applied to other physical processes involving monopoles, e.g., to pair production in magnetic fields [811].

In the limit where interactions between the configurations are small, we expect the partition function to be dominated by disconnected vacuum paths characterized by the transition amplitude $Z_{S^1}(\theta)$, the Feynman-weighted sum over all paths that are topologically a circle S^1. These contributions exponentiate:

$$\mathcal{Z}(\theta) = \sum_{n=0}^{\infty} \frac{1}{n!} (Z_{S^1})^n = \exp\big(Z_{S^1}(\theta)\big)\,. \tag{3.7}$$

Hence $V_{\text{eff}}(\theta) = -\frac{1}{\mathcal{V}} Z_{S^1}(\theta)$; we work in the first-quantized picture to compute the amplitude $Z_{S^1}(\theta)$ [812]. We sum over all trajectories that return to the same configuration. This includes an integral over the invariant length (Schwinger proper time) τ, weighted with a $1/2\tau$ to account for overcounting trajectories related by translations and reflections. So,

$$Z_{S^1} = \int_0^\infty \frac{d\tau}{2\tau}\, Z(\tau,\theta)\,, \tag{3.8}$$

with $Z(\tau,\theta)$ the sum over transition amplitudes at fixed θ of all trajectories with invariant length τ.

There are two ways we can compute Z_{S^1}. For a free particle of mass m, the gauge fixed transition

amplitude for a trajectory of length τ from point x to point x$'$ is

$$\langle \mathrm{x}'|\mathrm{x}\rangle_\tau = \frac{1}{2(2\pi\tau)^2}\exp\left(-\frac{1}{2\tau}(\mathrm{x}-\mathrm{x}')^2 - \mathrm{m}^2\tau\right) \tag{3.9}$$

After integrating over all trajectories that begin and end at the same point and canceling off a factor of the spacetime volume from the measure with the factor in the definition of the effective potential, we obtain

$$\mathrm{V}_{\text{eff}} = -\int_0^\infty \frac{\mathrm{d}\tau}{2\tau}\frac{1}{2(2\pi\tau)^2}\exp\left(-\frac{\mathrm{m}^2\tau}{2}\right). \tag{3.10}$$

We will sum over all dyon modes, labeled by $\mathrm{n}\in\mathbb{Z}$. To simplify the computation, we assume that the dyon mass spectrum takes the form

$$\mathrm{m}_\mathrm{n}^2 = \mathrm{m}_\mathrm{m}^2 + \mathrm{m}_\Delta^2\left(\mathrm{n}-\frac{\theta}{2\pi}\right)^2, \quad \mathrm{m}_\Delta^2 = \frac{\mathrm{m}_\mathrm{m}}{\mathrm{l}_\sigma}. \tag{3.11}$$

This agrees with (3.3) to order $1/\mathrm{l}_\sigma$, and in certain cases is an exact consequence of a BPS condition. In general, there may be power corrections in $(\mathrm{m}_\mathrm{m}\mathrm{l}_\sigma)^{-1}$. Summing over the tower of states, we obtain the effective potential

$$-\sum_{\mathrm{n}\in\mathbb{Z}}\int_0^\infty \frac{\mathrm{d}\tau}{4\tau\,(2\pi\tau)^2}\exp\left(-\frac{\mathrm{m}_\mathrm{m}^2\tau}{2}-\frac{\mathrm{m}_\Delta^2\tau}{2}\left(\mathrm{n}-\frac{\theta}{2\pi}\right)^2\right). \tag{3.12}$$

Periodicity in θ, arising from the sum over n, is manifest after Poisson resummation:

$$\sum_{\mathrm{n}\in\mathbb{Z}}\mathrm{e}^{-\frac{1}{2}\mathrm{m}_\Delta^2\tau\left(\mathrm{n}-\frac{\theta}{2\pi}\right)^2} = \sum_{\ell\in\mathbb{Z}}\sqrt{\frac{2\pi}{\mathrm{m}_\Delta^2\tau}}\exp\left(-\frac{2\pi^2\ell^2}{\mathrm{m}_\Delta^2\tau}+\mathrm{i}\ell\theta\right). \tag{3.13}$$

The effective potential then becomes

$$-\frac{\pi^2}{\mathrm{m}_\Delta}\sum_{\ell\in\mathbb{Z}}\int_0^\infty \frac{\mathrm{d}\tau\,\mathrm{e}^{\mathrm{i}\ell\theta}}{(2\pi\tau)^{7/2}}\exp\left(-\frac{\mathrm{m}_\mathrm{m}^2\tau}{2}-\frac{2\pi^2\ell^2}{\mathrm{m}_\Delta^2\tau}\right). \tag{3.14}$$

After integrating, the result is

$$\mathrm{V}_{\text{eff}}(\theta) = -\sum_{\ell=1}^\infty \frac{\mathrm{m}_\Delta^2\mathrm{m}_\mathrm{m}^2}{32\pi^4\ell^3}\mathrm{e}^{-2\pi\ell\mathrm{m}_\mathrm{m}/\mathrm{m}_\Delta}\cos(\ell\theta)\left(1+\frac{3\mathrm{m}_\Delta}{2\pi\ell\mathrm{m}_\mathrm{m}}+\frac{3\mathrm{m}_\Delta^2}{(2\pi\ell\mathrm{m}_\mathrm{m})^2}\right), \tag{3.15}$$

where we have ignored the irrelevant constant from the divergent $\ell=0$ integral.

We can think of the integer ℓ as the number of times the coordinate σ winds around itself for a particular configuration, and so we expect that we can interpret the effective potential (3.14) in terms of the monopole wordline action. Indeed, if we consider the relativistic completion of (3.2) with the dyon collective coordinate

σ treated as another (compact) spatial direction in which the monopole propagates, analogous to the DBI action:

$$S_m = m_m \int_\gamma d\lambda \sqrt{\frac{dx_\mu}{d\lambda}\frac{dx^\mu}{d\lambda} + \frac{l_\sigma}{m_m}\left(\frac{d_A\sigma}{d\lambda}\right)^2} + \int_\gamma \frac{\theta}{2\pi} d_A\sigma\,, \tag{3.16}$$

then we can compute the transition amplitude for a trajectory of length τ from point (x,σ) to point (x',σ'),

$$\langle x',\sigma'|x,\sigma\rangle_\tau = \frac{1}{2(2\pi\tau)^{5/2}} \exp\left(-\frac{1}{2\tau}(x'-x)^2 - \frac{l_\sigma}{2m_m\tau}(\sigma'-\sigma)^2 - \frac{m_m^2\tau}{2} + \frac{i\theta}{2\pi}(\sigma'-\sigma)\right). \tag{3.17}$$

Again, we integrate over all trajectories that begin and end at the same point, this time getting a contribution from the sum over windings $\sigma'-\sigma = 2\pi\ell$, which nicely reproduces (3.15). These calculations are identical to those of loop effects of Kaluza-Klein modes propagating in a circular dimension [773,775,813,814], but the physical interpretation is not; here the extra dimension is an internal one, visible only to the monopole.

We can understand the exponential factor in (3.15) via a saddle point approximation for each ℓ, corresponding to a classical Euclidean instanton solution that winds ℓ times in the σ coordinate while remaining at constant x^μ. The saddle is at Schwinger proper time $\tau_* = 2\pi\ell/(m_m m_\Delta)$. The instanton action, which controls the convergence of the Fourier expansion (3.15), is

$$S = \frac{2\pi m_m}{m_\Delta} \sim \frac{4\pi^2}{ke^2}\sqrt{\frac{\max(r_c, r_0)}{r_c}}. \tag{3.18}$$

Remarkably, for the critical 't Hooft-Polyakov monopole, the instanton action is $S = 8\pi^2/g^2$, precisely that of the classical BPST instanton in Yang-Mills theory [815,816]!

3.2.1 Light and Massless Fermions

As is familiar from standard instanton physics, the presence of light, charged fermions can dramatically alter a theory's θ-dependence. In particular, any dependence on θ should vanish as we take any charged fermion's mass to zero and thus restore a chiral symmetry.

While a full analysis of this effect—and the inclusion of multiple light fermions—is reserved for future work, we can easily understand how it impacts the dyon mass spectrum on dimensional grounds. Since the fermion dilutes the induced electric charge over a region roughly the size of its Compton wavelength, we expect that $r_* \sim m^{-1}$ in the estimate (3.4), and so the dyonic mass spacing becomes of order $m_\Delta^2 \sim m_m m$. Since this spacing vanishes as $m \to 0$, so does the θ-dependence of the dyon tower.

3.2.2 Higher-Derivative Corrections and Validity

Our calculation assumed the dyon mass spectrum presented in (3.11), which we expect to receive corrections in effective field theory when monopoles are not BPS. We should check that our result is robust against such corrections. These corrections can arise from higher derivative operators in the bulk effective Lagrangian, like $(F_{\mu\nu}F^{\mu\nu})^2$ or $\left(F_{\mu\nu}\tilde{F}^{\mu\nu}\right)^2$, or higher powers of $(\partial_\mu\sigma + kA_\mu)$ in the worldline Lagrangian. These are related: the former add $\mathbf{B}^4, (\mathbf{E}\cdot\mathbf{B})^2, \mathbf{B}^2\mathbf{E}^2$ and $\mathbf{E}^4$ terms to the energy density ρ. Integrating ρ outside the monopole core, similarly to the logic that led us to (3.4), implies that these terms modify the dyon mass spectrum. A series of terms of the form $c_{2j}\mathbf{E}^{2j}/\Lambda^{4(j-1)}$ in ρ generates corrections to the mass spectrum in even powers of $(n - \theta/2\pi)$:

$$m_n^2 = m_m^2 + m_\Delta^2 \sum_{j=1}^{\infty} \lambda_{2j} \left(n - \frac{\theta}{2\pi}\right)^{2j}, \tag{3.19}$$

where $\lambda_{2j} \sim c_{2j}\left[e^2k^2/\left(16\pi^2(r_*\Lambda)^4\right)\right]^{j-1}$ is small when $j > 1$ (and $\lambda_1 \equiv 1$, by the definition of m_Δ^2). Terms involving powers of both $\mathbf{B}$ and $\mathbf{E}$ give subleading shifts to the definitions of m_m^2, m_Δ^2, and the λ_j.

Repeating our earlier logic, we can sum the loop corrections (3.10) using the mass spectrum (3.19). Poisson resummation and relabeling $n - \theta/2\pi \to n$ then gives

$$\begin{aligned} V_{\text{eff}}(\theta) &= -\sum_{\ell\in\mathbb{Z}} \int_0^\infty \frac{d\tau}{4\tau} \frac{1}{(2\pi\tau)^2}\, e^{-\frac{1}{2}m_m^2\tau + i\ell\theta}\, \mathcal{Z}(\ell,\tau), \\ \mathcal{Z}(\ell,\tau) &\equiv \int_{-\infty}^{\infty} dn\, e^{-2\pi i n\ell - \frac{1}{2}m_\Delta^2\tau(n^2 + \lambda_4 n^4 + \cdots)} \end{aligned} \tag{3.20}$$

To evaluate the integral over n, we work in a saddle point approximation: defining $S_\ell(n)$ to be the function inside the exponent in (3.20), we ask that $dS_\ell(n_*)/dn = 0$. Treating λ_4 as a perturbation, we find that

$$n_* = -\frac{2\pi i\ell}{m_\Delta^2\tau} - 2i\lambda_4 \left(\frac{2\pi\ell}{m_\Delta^2\tau}\right)^3 + \mathcal{O}(\lambda_4^2, \lambda_6). \tag{3.21}$$

We require a small correction to the subsequent integral over τ, dominated by the saddle at $\tau_* = 2\pi\ell/(m_m m_\Delta)$. In particular, $m_\Delta^2\tau_* \ll 1$ (for small ℓ), which calls for caution: the semiclassical approximation requires sufficiently small λ_{2j}. In particular, (3.21) implies that a small correction to n_* at the saddle τ_* requires

$$|\lambda_4| \ll \frac{1}{2}\left(\frac{m_\Delta^2\tau_*}{2\pi\ell}\right)^2 \sim \frac{1}{2}\left(\frac{m_\Delta}{m_m}\right)^2 \lesssim \frac{e^4k^2}{8\pi^2}\frac{r_c}{r_*}, \tag{3.22}$$

where the last inequality can be derived from (3.4) and the surrounding discussion. Along similar lines, we require $|\lambda_{2j}| \ll (m_\Delta/m_m)^{2(j-1)}$. Focusing only on power counting in e, this requires that $|c_{2j}| \lesssim e^{2(j-1)}$. This will always hold when the operator $\mathbf{E}^{2j}$ is generated through loops of charged particles, as in the Euler-Heisenberg Lagrangian, where $|c_{2j}| \sim e^{2j}/(16\pi^2)$. In the case with $r_* = r_c$, this verifies that our semiclassical

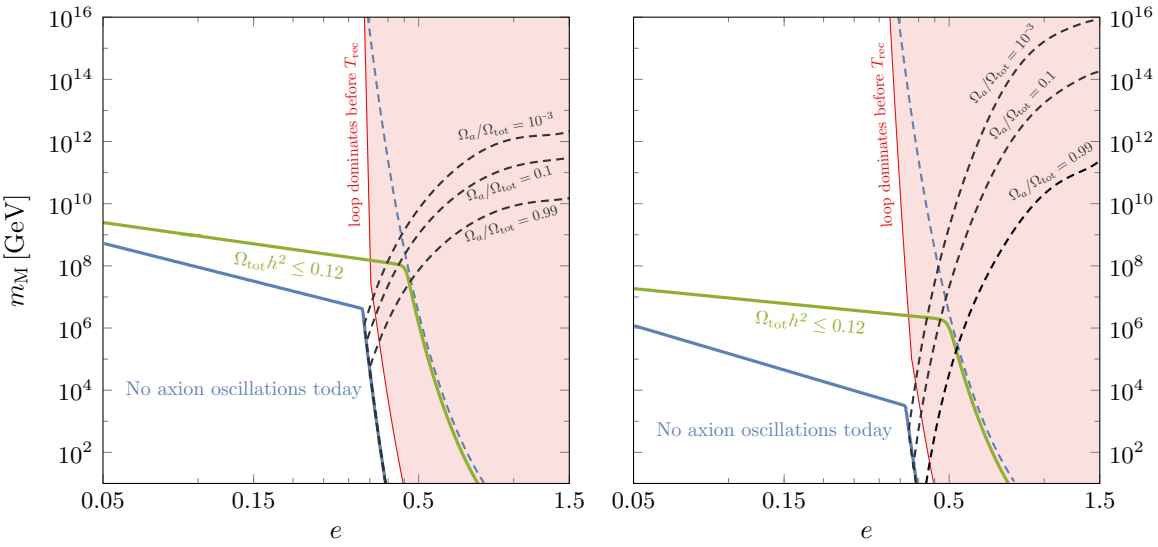


Figure 3.1: Pink regions: axion oscillates with the monopole loop-induced mass dominating over the monopole background-induced mass before CMB formation at temperature T_{rec}. Green solid line: the sum of axion abundance ($\Omega_a h^2$) and monopole abundance ($\Omega_m h^2$) today is $\Omega_{tot}h^2 = 0.12$, above which the abundance overcloses the Universe. Black dashed lines indicate the fraction of axion dark matter in the total abundance of axion and monopole today. Below the solid blue line, the axion's mass is so small that it never oscillates, while it is greater than 10^{-13} eV to the right of the dashed blue line. When the axion is the dominant component of dark matter today, its mass is around 10^{-13} eV, safely above bounds from structure formation. Left panel: monopole yield saturates the Kibble bound. Right panel: monopole yield is from a second order phase transition with a critical exponent $\nu = 0.5$. We fix $m_m r_c = \pi/e^2$, the critical temperature to be $T_c = 1/r_c$, and $f = 10^{15}$ GeV.

calculation can be performed within the context of a sensible effective field theory in which higher derivative operators produce controllably small corrections. The case $r_* = r_0$ requires somewhat more care regarding the allowed range of the scale Λ, which we will not delve into here.

3.3 Phenomenological Applications

There could potentially be many interesting phenomenological implications of this new monopole contribution to the axion potential, which calls for future work. Here we will only consider a hidden sector with a gauged $U(1)_d$ symmetry and a gauge coupling e as a minimal example to show that this new contribution could play an important role in the cosmological evolution of axion-like particles. In the hidden sector, there is also a global $U(1)_{pq}$ symmetry which is broken spontaneously at the scale f and results in a Goldstone boson, the axion a.

In the presence of monopoles carrying magnetic charge under the $U(1)_d$, the axion obtains a mass from both the temperature-independent monopole loop, which is the new finding of this chapter, and the temperature-dependent contribution from a monopole background as discussed in [803, 807–809]. The

monopole background could be generated via the Kibble-Zurek mechanism in a phase transition happening at a critical temperature T_c in the early Universe [622, 623]. The monopole yield satisfies the Kibble lower bound [622] and could be significantly above the bound if the phase transition is second order [623, 817]. In this model, both axions and monopoles could be components of dark matter. The relevant parameter space is shown in Fig. 3.1, assuming that the visible and dark sectors share a common temperature at early times. We see that when the gauge coupling e is large, $e \gtrsim 0.5$, the monopole loop-induced axion mass would dominate over the contribution from the monopole background before the CMB formation. In addition, the axion abundance is negligible when $e \lesssim 0.5$, but it could take over that of the monopoles at larger values $e \gtrsim 0.5$. In order not to overclose the Universe, the monopoles must not be very heavy [818, 819]. Fig. 3.1 establishes that the new effect we discuss can modify the cosmology of axions and monopoles; it would be interesting to incorporate it in a wider range of models in the future.

3.4 Discussion

In this chapter, we have presented and computed a new contribution to the vacuum axion potential from magnetic monopole loops, when the axion is coupled to an abelian gauge field. Much more remains to be studied, both in developing the formalism and exploring the phenomenological and cosmological implications. We briefly comment on some possible future directions: *1)* We have assumed $V_{\text{eff}}(\theta)$ is dominated by a single monopole loop, but there are long-range Coulomb interactions between the monopoles. Their effect on the semiclassical expansion should be explored. *2)* We found that the action of the monopole-loop instanton in the critical 't Hooft-Polyakov case is that of a BPST instanton, $8\pi^2/g^2$. This may be a harbinger of a stronger statement: we expect that the monopole-loop instantons can be continuously deformed into nonabelian instantons. If not, the theory would have an unbroken global (-1)-form U(1) symmetry in 4d, and a $(d-5)$-form symmetry in higher dimensions [793]. Similarly, in cases where U(1) gauge fields arise on $D(n+3)$-branes wrapped on n-cycles in extra dimensions, one obtains axion potentials either from wrapped Euclidean $D(n-1)$-branes or from magnetic monopoles, which are $D(n+1)$-branes ending on the $D(n+3)$-branes. The winding of σ on the monopole worldline in 4d arises from a nonvanishing field strength on the $D(n+3)$-brane, which via the worldvolume Chern-Simons term, is equivalent to $D(n-1)$-brane charge. Again, we expect that the monopole-loop instanton can be continuously deformed into a Euclidean brane instanton in this context. These deformations between instantons should be constructed more explicitly. *3)* We demonstrated that this new contribution could be important in a hidden sector model with the axion coupling to a dark $U(1)_d$. Consider an axion coupling to the standard model photon instead. Does this imply a minimum mass of the axion, even without nonabelian instantons? What are the effects of multiple fermions, present in the standard model? Future work answering these questions will directly connect the effect we have presented with ongoing experiments.

4

Zero Modes of Massive Fermions Delocalize from Axion Strings

4.1 Introduction

In addition to the monopole potential discussed in Chapter 3, there are other interesting connections between axions and topological defects. For example, axion strings can be important phenomenologically. One example is in post-inflationary axion models, where the axion abundance is determined by radiation from a network of cosmic axion strings [624, 820–835]. Other examples include recent work on potential signatures of axion string networks [836, 837].

Recently, there has been renewed interest in the fact that axion strings can be *superconducting*: they support charged zero modes localized to the string core, which lead to a current on the string proportional to an applied electric field. That axion strings are superconducting has been understood for many years, as it is intimately related to the phenomenon of anomaly inflow elucidated by Callan and Harvey [838], which we will review in this chapter. A great deal of the physics of anomaly inflow and axion strings was worked out in subsequent years: see [839–847] for a collection of early and important works on the subject. The contemporary interest in axion string superconductivity is in part due to the realization that these strings could interact with a primordial magnetic field, leading to striking signatures due to the formation of vortons [848, 849] or bound states [850].

The existence of the charged zero modes is often explained in the context of simple models with a classical

PQ symmetry. In these models, the fermions acquire mass from the vacuum expectation value of the radial mode of the scalar field whose phase is the axion. In classical string configurations, the vacuum expectation value goes to zero at the core of the string solution and intuition strongly suggests that massless modes will exist in a region localized at the string core. An analogous argument successfully explains the existence of zero modes on domain walls.

While this simple picture is intuitive, for axion strings, it is clearly incomplete. For one, it pays no heed to the crucial fact that the zero modes on axion strings are *chiral*—a property which distinguishes them from zero modes on e.g., Witten strings [851]. Moreover, in more complicated models (for instance, the DFSZ axion [594, 595]), there can exist string configurations in which the vacuum expectation value of the scalar does not vanish at the string core [852]. Despite this, there are arguments that these strings can be superconducting as well.

This picture is also related to another puzzle. What happens to the zero mode when this classical PQ symmetry is badly and explicitly broken? For instance, if the fermions have a very large mass m, we should have no trouble completely integrating them out and the axion string should be blind to their existence. In this case, there is no anomaly to inflow, and so these zero modes should not exist in the limit $\mathrm{m} \to \infty$. In contrast, since they are chiral we still expect them to exist for small but non-zero m. So, there must be some critical value of the mass m at which they cease to exist. What happens to the zero modes near this critical mass?

The goal of this chapter is to shed some light on these puzzles. We will do so by studying a simple model of axion electrodynamics in which the classical PQ symmetry is *explicitly* broken by a Dirac mass m for the fermion. We demonstrate explicitly that there exist zero mode solutions to the equations of motion in the axion string background, and numerically solve for their profile. We find that when the Dirac mass is roughly equal in size to the mass μ induced by the scalar field which spontaneously breaks the PQ symmetry, the profile functions change dramatically: the zero mode becomes completely delocalized from the string and onto a semi-infinite wedge. We also revisit the anomaly inflow story in the presence of the bulk mass term, clarifying the topological origin of the zero modes even in the absence of a "classical" PQ symmetry. Finally, we derive the low-energy two-dimensional effective theory for the zero mode and calculate the leading higher-derivative interactions with the bulk gauge field. As one would expect, we show that this effective theory completely breaks down as $\mathrm{m} \to \mu$. While this simple model is not relevant phenomenologically, our hope is that this work can be applied to more realistic models, with potential astrophysical or cosmological effects that can be studied in future work.

It is worth emphasizing that these considerations are entirely distinct from situations in which the zero modes localized on cosmic (non-axionic) strings can acquire mass from the pairing of left- and right-moving

modes [853], which can be induced e.g., by finite temperature effects [854]. The mass term we introduce is for the "full" theory of fermions propagating in four-dimensional spacetime. The zero modes localized to the string remain massless on topological grounds, as in the original example [838]. The zero modes we discuss are also distinct from those found in the background of "Z-strings" [855–859]. Those string configurations are not topologically stable and there is no mechanism that protects the zero modes from pairing up and acquiring a mass [860, 861].

The rest of this chapter is structured as follows. In §4.2, we review the axion string solution in the usual case, with a global PQ symmetry, and solve for the massless zero modes explicitly. In §4.3, we break the PQ symmetry with a Dirac mass and demonstrate that the zero modes still exist. We discuss how this can be understood by analogy to the existence of zero modes on domain walls in $2+1$ dimensions, present numerical results for the profile of these zero modes on the string, and discuss the behavior in the "critical" mass case. In §4.4, we recap the original anomaly inflow story, with appropriate modifications to account for the mass term. Finally, in §4.5, we derive the low-energy theory of the zero modes on the string, and discuss how it is impacted by the Dirac mass term. We conclude in §4.6. Appendix A provides more detail on the numerical techniques used to solve for the zero modes with Chebyshev interpolation, and on the derivation of the low-energy effective action.

4.2 Axion Strings

In this section, we will review the original ultraviolet completion of an axion string studied by Callan and Harvey [838], with Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}i\not{D}\psi + |\partial_\mu\Phi|^2 + y\bar{\psi}\left(\Phi_1 + i\gamma^5\Phi_2\right)\psi - V(\Phi)\,. \tag{4.1}$$

This theory lives in four-dimensional Minkowski space $\mathcal{M}_4$ and consists of an abelian gauge field A_μ with field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, an uncharged complex scalar field $\Phi \equiv \Phi_1 + i\Phi_2 \equiv f(x)e^{i\theta(x)}$ with potential

$$V(\Phi) = \lambda\left(|\Phi|^2 - v^2\right)^2\,, \tag{4.2}$$

and a single charged Dirac fermion ψ, which chirally couples to Φ with strength y. We use $D_\mu = \partial_\mu - ieA_\mu$ to denote the gauge covariant derivative, where $\not{D} \equiv \gamma^\mu D_\mu$, and γ^μ are the standard Dirac gamma matrices with $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ and $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$. This theory enjoys a $U(1)_{PQ}$ global symmetry, commonly called the Peccei-Quinn symmetry, under which the complex scalar and fermion transforms as $\Phi \to e^{i\alpha}\Phi$ and $\psi \to e^{-i\gamma^5\alpha/2}\psi$, respectively.

4.2.1 The Axion String

The potential (4.2) forces the scalar Φ to acquire a vacuum expectation value $\langle\Phi\rangle = \mathrm{v}$. It also allows for the existence of static axionic string configurations,

$$\Phi_{\mathrm{n}}(\mathrm{x}) = \mathrm{f}(\mathrm{r})\mathrm{e}^{\mathrm{in}\phi}\,, \tag{4.3}$$

characterized by an integer topological charge $\mathrm{n} \in \mathbb{Z}$. We work in standard cylindrical coordinates $(\mathrm{x}, \mathrm{y}, \mathrm{z}) = (\mathrm{r}\cos\phi, \mathrm{r}\sin\phi, \mathrm{z})$, with the string oriented along the z-axis. In these configurations, the phase of the complex scalar field, i.e. the "axion" $\theta(\mathrm{x}) \equiv \arg\Phi(\mathrm{x})$, winds n times around its field space $\theta \sim \theta + 2\pi$ as we move around the string. Said differently,

$$\mathrm{n} \equiv \frac{1}{2\pi}\oint_{\gamma} \mathrm{d}\theta\,, \tag{4.4}$$

where γ is a closed contour that encircles the string at $\mathrm{r} = 0$. The "radial profile" of this string is determined by the real function $\mathrm{f}(\mathrm{r})$. Far from the string, the potential (4.2) forces the scalar to sit in the minimum of its potential and so $\mathrm{f}(\mathrm{r}) \to \mathrm{v}$ as $\mathrm{r} \to \infty$. Likewise, regularity of the solution forces $\mathrm{f}(\mathrm{r})$ to vanish in the core of the string: $\mathrm{f}(\mathrm{r}) \to 0$ as $\mathrm{r} \to 0$. These solutions are also called "global vortices" in the literature, since they are vortex solutions that are charged under the *global* $\mathrm{U}(1)_{\mathrm{pq}}$ symmetry, as opposed to a gauge symmetry.

The tension of an axion string with charge n is given by

$$\mathrm{T}_{\mathrm{n}} = 2\pi\int_0^{\mathrm{L}} \mathrm{dr}\,\mathrm{r}\left[\left(\frac{\mathrm{df}}{\mathrm{dr}}\right)^2 + \frac{\mathrm{n}^2}{\mathrm{r}^2}\mathrm{f}^2 + \lambda(\mathrm{f}^2 - \mathrm{v}^2)^2\right]. \tag{4.5}$$

Since $\mathrm{f}(\mathrm{r}) \to \mathrm{v}$ as $\mathrm{r} \to \infty$, the tension diverges logarithmically with the size of the system or IR cutoff L,

$$\mathrm{T}_{\mathrm{n}} \sim 2\pi\mathrm{v}^2\mathrm{n}^2\log\left(\frac{\mathrm{L}}{\mathrm{r}_{\mathrm{core}}}\right), \qquad \mathrm{L} \to \infty\,. \tag{4.6}$$

Here, we have introduced the "size" of the string $\mathrm{r}_{\mathrm{core}}$, which is of order $\mathrm{r}_{\mathrm{core}} = \left(\sqrt{\lambda}\mathrm{v}\right)^{-1}$. This IR divergent total energy is common for global strings and is typically regulated by either placing the theory in a box or by positing that there is another string of opposite charge a distance L away, as is often the case in, e.g. cosmological simulations of axion string networks.

Minimizing the tension yields the equation of motion

$$\frac{1}{\mathrm{r}}\frac{\mathrm{d}}{\mathrm{dr}}\left(\mathrm{r}\frac{\mathrm{df}}{\mathrm{dr}}\right) - \left[\frac{\mathrm{n}^2}{\mathrm{r}^2} + 2\lambda(\mathrm{f}^2 - \mathrm{v}^2)\right]\mathrm{f} = 0\,. \tag{4.7}$$

In general, this equation cannot be solved analytically but is amenable to numerics. We will restrict our

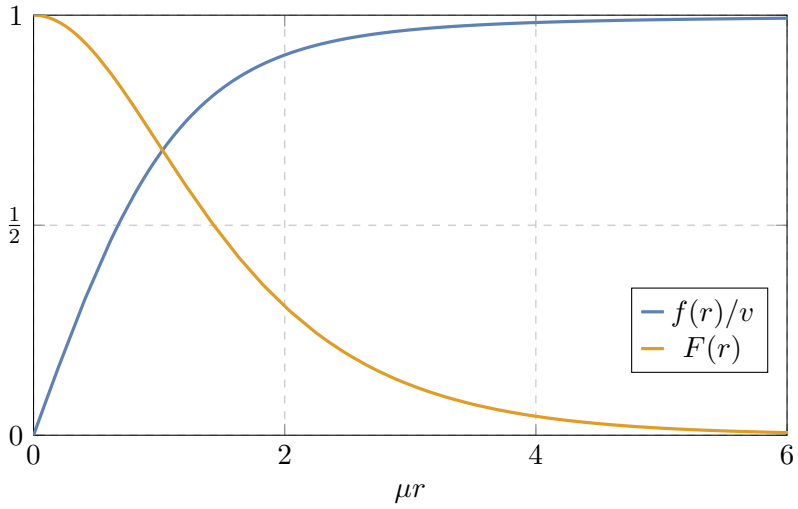


Figure 4.1: The radial profile f(r) [blue], in units of v, for the axion string of charge ±1 and corresponding (non-normalized) zero mode profile F(r) [orange] with $y = \sqrt{\lambda}$.

attention throughout to strings of charge $n = +1$. An asymptotic analysis of (4.7) shows that the radial profile f(r) behaves as

$$f(r) \sim \begin{cases} v C_1 \left(\sqrt{\lambda} v r\right) + \cdots & r \to 0 \\ v\left(1 - \dfrac{1}{4\lambda v^2 r^2}\right) + \cdots & r \to \infty \end{cases}, \tag{4.8}$$

with C_1 an overall constant that can be analytically determined by matching the two solutions in an intermediate region, or by numerics. It is clear from the behavior as $r \to \infty$ that the string solution varies over length scales $r_{core} = \left(\sqrt{\lambda} v\right)^{-1}$. We may then use the techniques outlined in Appendix A.1 to numerically search for the solution which varies over scales set by r_{core} with the correct asymptotics (4.8). We show this profile in Fig. 4.1.

4.2.2 Fermionic Zero Modes

We are interested in the low energy dynamics of (4.1) in the background of the axion string. Absent the string, the fermion acquires a Dirac mass $\mu = yv$—which we call the "Yukawa mass" since it arises from the spontaneous breaking of the $U(1)_{pq}$ symmetry—and so the low-energy theory only contains the massless axion and the abelian gauge field. One might think that the low-energy theory in the presence of the axionic string is similar; far from the string, the fermions again get a mass μ and the low-energy theory should be unchanged. However, as argued in [838] this reasoning is incomplete: there are instead massless chiral zero modes which are localized to the string. In this section, we will review the direct construction of these zero modes. A common intuitive explanation is that these massless modes exist because the radially-dependent

"mass" of the bulk fermions $\mathrm{yf}(\rho)$ vanishes in the core of the string. However, we argue in §4.3 that this is not a necessary condition, while in the §4.4 we discuss why these modes are necessary from the perspective of anomaly inflow.

In the presence of the string of charge $\mathrm{n} = +1$ and no background gauge field, the fermions obey the equation of motion

$$\left(\mathrm{i}\not{\partial} + \mathrm{yf}(\mathrm{r})\mathrm{e}^{\mathrm{i}\gamma^5\phi}\right)\psi = 0\,. \tag{4.9}$$

For simplicity, we will restrict to solutions of the form

$$\psi(\mathrm{x}) = \psi(\mathrm{r},\phi)\,\mathrm{e}^{-\mathrm{ip}(\mathrm{t}+\mathrm{z})}\,, \tag{4.10}$$

where $\psi(\mathrm{r},\phi) = \psi_\alpha(\mathrm{r},\phi)$ is a four-component spinor function with $\alpha = 0,\ldots,3$. This ansatz describes solutions that travel in the (–z)-direction at the speed of light. In the Weyl representation, (4.9) reduces to the set of coupled equations

$$\begin{aligned}
0 &= 2\mathrm{p}\psi_2 + \mathrm{e}^{-\mathrm{i}\phi}\left(\mathrm{yf}(\mathrm{r})\psi_0 + \mathrm{i}\partial_\mathrm{r}\psi_3 + \mathrm{r}^{-1}\partial_\phi\psi_3\right)\\
0 &= \mathrm{yf}(\mathrm{r})\mathrm{e}^{-\mathrm{i}\phi}\psi_1 + \mathrm{e}^{\mathrm{i}\phi}\left(\mathrm{i}\partial_\mathrm{r}\psi_2 - \mathrm{r}^{-1}\partial_\phi\psi_2\right)\\
0 &= \mathrm{yf}(\mathrm{r})\mathrm{e}^{\mathrm{i}\phi}\psi_2 - \mathrm{e}^{-\mathrm{i}\phi}\left(\mathrm{i}\partial_\mathrm{r}\psi_1 + \mathrm{r}^{-1}\partial_\phi\psi_1\right)\\
0 &= 2\mathrm{p}\psi_1 + \mathrm{e}^{\mathrm{i}\phi}\left(\mathrm{yf}(\mathrm{r})\psi_3 - \mathrm{i}\partial_\mathrm{r}\psi_0 + \mathrm{r}^{-1}\partial_\phi\psi_0\right)
\end{aligned}\,. \tag{4.11}$$

Such solutions should have definite helicity, and so we may set $\psi_1 = \psi_2 = 0$. We then see that $\psi_0 = \psi_0(\mathrm{r})$ and $\psi_3 = \psi_3(\mathrm{r})$ are purely radial functions that satisfy the coupled system of equations

$$\begin{aligned}
0 &= \mathrm{yf}(\mathrm{r})\psi_0 + \mathrm{i}\partial_\mathrm{r}\psi_3\\
0 &= \mathrm{yf}(\mathrm{r})\psi_3 - \mathrm{i}\partial_\mathrm{r}\psi_0
\end{aligned}\,. \tag{4.12}$$

These equations can be easily solved to find two solutions of the form

$$\psi_\mathrm{p}(\mathrm{x}) = \sqrt{\frac{\mathrm{p}}{2}}\begin{pmatrix}1\\0\\0\\\mp\mathrm{i}\end{pmatrix}\mathrm{e}^{-\mathrm{ip}(\mathrm{t}+\mathrm{z})}\mathrm{F}(\mathrm{r})\,, \tag{4.13}$$

where $\mathrm{F}'(\mathrm{r}) = \pm\mathrm{yf}(\mathrm{r})\mathrm{F}(\mathrm{r})$ or

$$\mathrm{F}(\mathrm{r}) = \mathcal{A}\exp\left(\pm\mathrm{y}\int_0^\mathrm{r}\mathrm{dr}'\,\mathrm{f}(\mathrm{r}')\right), \tag{4.14}$$

with $\mathcal{A}$ an overall constant of integration. For this solution to be appropriately normalizable,

$$\int_{\mathbb{R}^3}\mathrm{d}^3\mathrm{x}\,\psi^\dagger_\mathrm{p}(\mathrm{x})\psi_{\mathrm{p}'}(\mathrm{x}) = 2\pi|\mathrm{p}|\,\delta(\mathrm{p}-\mathrm{p}')\,, \tag{4.15}$$

or specifically

$$2\pi \int_0^\infty \mathrm{dr}\, \mathrm{r} \mathrm{F}^2(\mathrm{r}) = 1\,, \tag{4.16}$$

so we must select the negative sign in (4.14). We thus find that the axion string with charge +1 supports a chiral fermionic zero mode that travels at the speed of light in the (−z)-direction, with form

$$\psi(\mathrm{x}) = \mathcal{A}\sqrt{\frac{\mathrm{p}}{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -\mathrm{i} \end{pmatrix} \mathrm{e}^{-\mathrm{ip}(\mathrm{t}+\mathrm{z})} \exp\left(-\mathrm{y}\int_0^\mathrm{r} \mathrm{dr}'\, \mathrm{f}(\mathrm{r}')\right)\,. \tag{4.17}$$

If we were to repeat this analysis with the charge −1 axion string, then it would support a normalizable chiral zero mode that instead travels at the speed of light in the (+z)-direction,

$$\psi(\mathrm{x}) = \mathcal{A}\sqrt{\frac{\mathrm{p}}{2}} \begin{pmatrix} 0 \\ 1 \\ -\mathrm{i} \\ 0 \end{pmatrix} \mathrm{e}^{-\mathrm{ip}(\mathrm{t}-\mathrm{z})} \exp\left(-\mathrm{y}\int_0^\mathrm{r} \mathrm{dr}'\, \mathrm{f}(\mathrm{r}')\right)\,. \tag{4.18}$$

In Fig. 4.1, we show the radial profile F(r) for $\mathrm{y} = \sqrt{\lambda}$. These zero modes are localized in a region about the string of size $\mathrm{r}_{\mathrm{zm}} = \mathrm{r}_{\mathrm{core}}/\mathrm{y}$.

Before we move on, it will be useful to understand why the presence of the axion string permits these normalizable modes to exist, solely from the perspective of the differential equations. We will see that, contrary to naive expectation, it is not because the fermion is "massless" in the core of the string. With (4.10), we can decouple the equations in (4.12) to find two copies of the second order equation

$$\left[\frac{\mathrm{d}^2}{\mathrm{dr}^2} - \frac{\mathrm{f}'(\mathrm{r})}{\mathrm{f}(\mathrm{r})}\frac{\mathrm{d}}{\mathrm{dr}} - \mathrm{y}^2\mathrm{f}^2(\mathrm{r})\right]\mathrm{F}(\mathrm{r}) = 0\,, \tag{4.19}$$

with $\psi_0(\mathrm{r}) = \mathrm{F}(\mathrm{r})$ and $\psi_3(\mathrm{r}) = -\mathrm{iF}(\mathrm{r})$. As is clear from Fig. 4.1, the function f(r) is smooth and non-singular for all $\mathrm{r} \in (0, \infty)$. This implies that the solutions to (4.19) are regular everywhere, except possibly as $\mathrm{r} \to 0$ and $\mathrm{r} \to \infty$, which are regular and irregular singular points, respectively. A normalizable solution must be regular at both of these points and generally fails to exist because the solution that is regular about one singularity is not regular about the other.

Using the asymptotics (4.8) of the background f(r), far from the string (4.19) reduces to

$$\left[\frac{\mathrm{d}^2}{\mathrm{dr}^2} - \mu^2 + \cdots\right]\mathrm{F}(\mathrm{r}) = 0\,, \tag{4.20}$$

which has solutions that behave as

$$\mathrm{F}(\mathrm{r}) = \mathrm{C}_-\mathrm{F}_-(\mathrm{r}) + \mathrm{C}_+\mathrm{F}_+(\mathrm{r}) \sim \mathrm{C}_-\,\mathrm{e}^{-\mu\mathrm{r}} + \mathrm{C}_+\mathrm{e}^{\mu\mathrm{r}}\,, \qquad \mathrm{r} \to \infty\,, \tag{4.21}$$

where we have defined two linearly-independent solutions $\mathrm{F}_\pm(\mathrm{r})$ with definite scaling as $\mathrm{r} \to \infty$. In this limit,

the fermion zero modes do not "see" the string, but merely feel their acquired mass $\mu = yv$. However, near the core of the string, (4.19) instead reduces to

$$\left[\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \cdots\right] F(r) = 0\,. \tag{4.22}$$

As we move towards the core of string $r \to 0$, there is a balance between the first and second terms in (4.19), the latter of which only appears due to the existence of the string. The third term, due to the Yukawa interaction, can be neglected. Solutions then behave as

$$F(r) = C_1 F_1(r) + C_2 F_2(r) \sim C_1 + C_2 r^2\,, \qquad r \to 0\,, \tag{4.23}$$

where we have again identified two solutions $F_{1,2}(r)$ which have definite scaling as $r \to 0$. Importantly, both of these solutions are regular near the origin.

We can understand why the axion string allows a normalizable zero mode to exist as follows. Since (4.19) is a second-order differential equation, it has only two linearly-independent solutions. Above, we found that a general solution can either be expressed as a linear combination of $F_{1,2}(r)$ or $F_{\pm}(r)$. Since these two sets of solutions are not linearly independent, there always exists a linear map between them. For a solution to be normalizable, it must decay as $r \to \infty$ and so we must have that $C_+ = 0$; therefore we must be able to write this solution as $F(r) = C_- F_- = C_{-,1} F_1(r) + C_{-,2} F_2(r)$. It is often the case that only one of the solutions $F_1(r)$ or $F_2(r)$ is well-behaved at the origin $r = 0$ and, unless there is some special structure that ensures it, the solution that is well-behaved as $r \to \infty$ will not be well-behaved as $r \to 0$. Fortunately, the presence of the string ensures that *both* linearly-independent solutions are regular as $r \to 0$. Since $f(r)$ is a smooth function, we are also guaranteed that the solutions are regular for positive r. Consequently, a normalizable solution to (4.19) exists, regardless of its exact form.

An analogous situation occurs for the unbound states of the quantum mechanical hydrogen atom, whose radial modes obey

$$\left[-\frac{1}{2r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}\right) + \frac{\ell(\ell+1)}{2r^2} - \frac{\alpha}{r} - E\right]\psi(r) = 0\,. \tag{4.24}$$

Due to the centrifugal barrier, solutions behave as $\psi(r) \sim C_1 r^{\ell} + C_2 r^{-(\ell+1)}$ as $r \to 0$ and $\psi(r) \sim C_- e^{-\sqrt{-2E}r} + C_+ e^{\sqrt{-2E}r}$ as $r \to \infty$. Bound states, $E < 0$, have a discrete spectrum because it is impossible to simultaneously impose that the wavefunction be both regular at the origin *and* exponentially decay at spatial infinity, except at a discrete set of energy eigenvalues. For all other energies, solutions with $C_+ = 0$ necessarily have $C_2 \neq 0$. For unbound states, $E > 0$, there is no restriction on the behavior as $r \to \infty$, and so there exists a continuum of regular solutions with $C_2 = 0$, for all positive energies.

Having reviewed how axion strings, and their associated fermionic zero modes, arise in the Callan–Harvey

model (4.1), we are now in a position to understand how these zero modes behave as we deform (4.1). In the next section, we will study what happens to these zero modes in the presence of a Dirac mass m for the four-dimensional fermion. Even though the four-dimensional fermion is everywhere massive, these fermionic zero modes still exist as long as $|m| < \mu$. Furthermore, we argue that there is an interesting "phase structure" in which the zero modes become increasingly unbound from the string as $|m| \to \mu$, eventually disappearing for $|m| > \mu$.

4.3 Adding a Mass

In the previous section, we reviewed how fermionic zero modes could arise along an axion string in the context of a simple toy model. In that specific UV completion, the axion string is a solitonic object in which the $U(1)_{pq}$ symmetry that is spontaneously broken far from the string is restored at the core. This causes the four-dimensional fermion to see effectively zero mass near the core, and a popular refrain is that this why we should expect to see massless fermionic zero modes localized on the string. The goal of this section is to explain why this is not a necessary condition and see that fermionic zero modes can arise even in theories where bulk four-dimensional fermions have a nowhere vanishing mass.

We will deform (4.1) by adding a mass m to the Dirac fermion ψ,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi + |\partial_\mu \Phi|^2 + y\bar{\psi}\left(\Phi_1 + i\gamma^5\Phi_2\right)\psi - V(\Phi)\,, \tag{4.25}$$

again with $V(\Phi) = \lambda(|\Phi|^2 - v^2)$. We restrict to the case where the amplitude of ψ is small. This theory permits the same axion string solution (4.3) as before. However, the fermion now obeys a modified equation,

$$\left(i\not{\partial} - M(r,\phi)e^{i\gamma^5\alpha(r,\phi)}\right)\psi = 0\,, \tag{4.26}$$

where $M(r,\phi)e^{i\gamma^5\alpha(r,\phi)} = m - yf(r)\,e^{i\gamma^5\phi}$ or, explicitly,

$$\begin{aligned} M(r,\phi) &= \sqrt{\left(m - yf(r)\cos\phi\right)^2 + y^2 f^2(r)\sin^2\phi} \\ \alpha(r,\phi) &= \arg\left(m - yf(r)e^{-i\phi}\right) = i\log\left(\frac{m - yf(r)e^{-i\phi}}{M(r,\phi)}\right) \end{aligned}\,. \tag{4.27}$$

In this case, the fermion mass $M(r,\phi)$ no longer goes to zero at the core of the string even though $f(r) \to 0$, but instead approaches the "core mass" $M(r=0,\phi) = m$. Even so, we will still find that this string supports fermionic zero modes as long as this core mass is less than the Yukawa mass, $|m| < \mu$.

To find these zero modes, we again restrict to axion strings with charge +1 and search for zero modes of the form (4.10). The mass deformation $m\bar{\psi}\psi$ explicitly breaks the chiral symmetry or, analogously, the

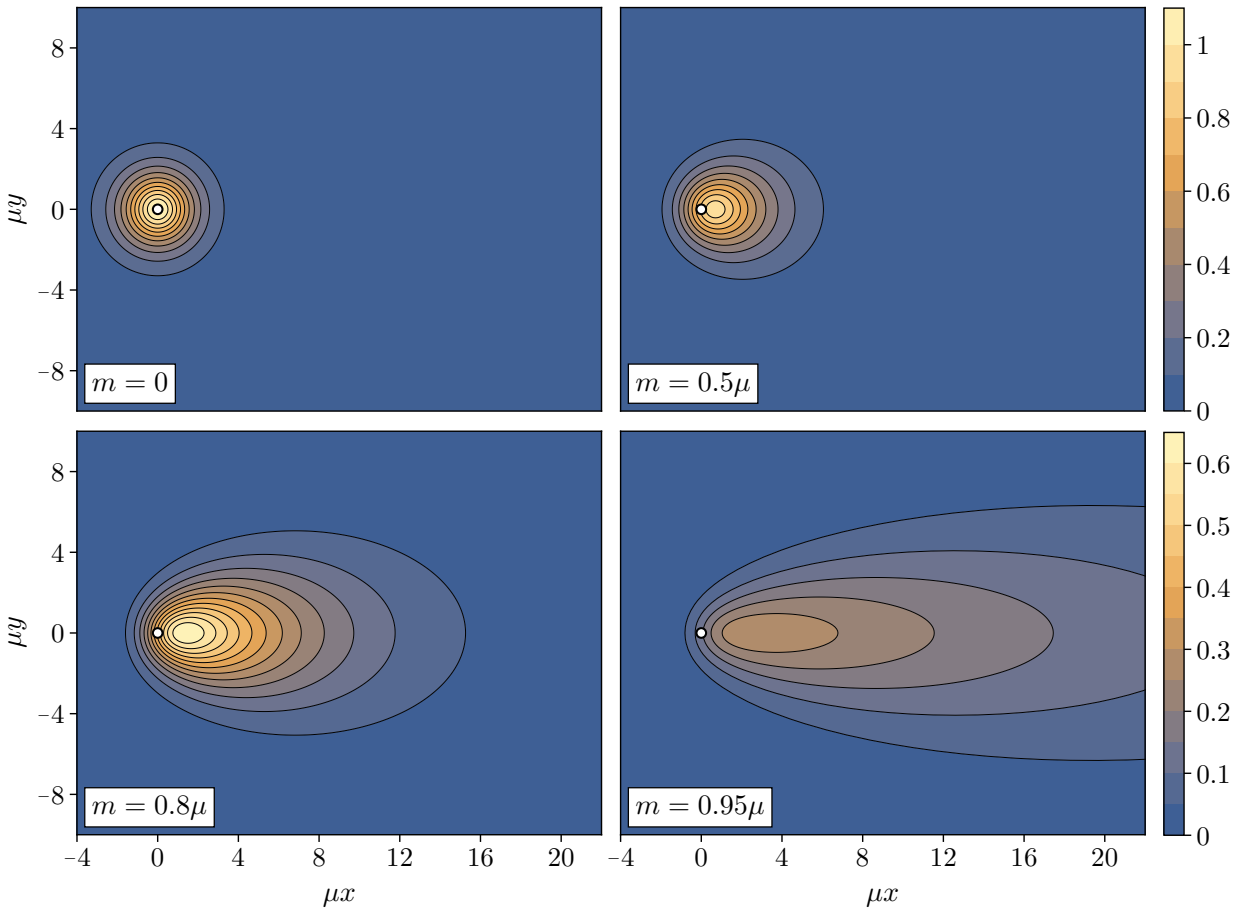


Figure 4.2: The profile functions of the zero mode wavefunction $|\psi_0(\mathrm{r},\phi)|$ in the presence of a Dirac core mass m and normalized according to (4.15), for various values of m/μ and $\sqrt{\lambda} = \mathrm{y} = \sqrt{0.1}$, with contours in units of $\sqrt{\omega}\mu$. The other component satisfies $\psi_3(\mathrm{r},\phi) = -\mathrm{i}\psi_0(\mathrm{r},-\phi)$ and is thus identical in structure. The white dot denotes the origin of the string, $\mathrm{r} = 0$. As $\mathrm{m} \to \mu$, the zero mode profiles spread out and become unbound from the string—note the difference in scales of the contours between the two rows. For $\mathrm{m} > \mu$, these zero modes cease to exist. Qualitatively similar results apply when $\sqrt{\lambda} \neq \mathrm{y}$.

continuous axion shift symmetry $\phi \to \phi + \alpha$, and thus changes the form of the axion string solution. However, classically this potential is controlled by fermion's number density $\bar{\psi}\gamma^0\psi$, and in the limit of small amplitude—or, quantum mechanically, small occupation numbers—this effect is subleading and can be ignored. Similarly, we will not consider the electromagnetic field generated by the charged fermion, setting the vector potential $\mathrm{A}_\mu = 0$, as this effect is also subleading in the limit of small amplitudes.

It is worth noting that integrating out the fermion generates a potential for the scalar that depends on ϕ. The leading Coleman–Weinberg estimate $\propto \mathrm{M}^4(\mathrm{r},\phi)\log(\mathrm{M}(\mathrm{r},\phi)/\mu)$ [806, 862] is minimized along $\phi = 0$ and leads to a domain wall for the axion that emanates off the string and extends in this direction. In what follows, we are explicitly assuming that this contribution is fine-tuned away, though we will see that there is a remnant of the axion domain wall in the orientation of the zero mode profiles.

By Lorentz symmetry, we can set $\psi_1(r,\phi) = \psi_2(r,\phi) = 0$ since the modes we are searching for propagate in the (–z)-direction at the speed of light, and so in the Weyl representation (4.26) reduces to

$$\begin{aligned} 0 &= -m\psi_0 + e^{-i\phi}\big(yf(r)\psi_0 + i\partial_r\psi_3 + r^{-1}\partial_\phi\psi_3\big) \\ 0 &= -m\psi_3 + e^{i\phi}\big(yf(r)\psi_3 - i\partial_r\psi_0 + r^{-1}\partial_\phi\psi_0\big) \end{aligned} . \tag{4.28}$$

We can solve (4.28) numerically using Chebyshev interpolation, whose details we describe in Appendix A.1. The resulting profile functions are shown in Fig. 4.2 for various values of m/μ. In the plot, we see the normalized zero mode as a function of dimensionless x and y variables, with a white dot corresponding to the location of the origin and hence the axion string. Surprisingly, we find that this zero mode becomes less localized to the string the larger the core mass m is. For large $m \lesssim \mu$, it is no longer centered at the core of the axion string and instead stretches out along the positive x-axis. This direction is determined by the orientation of the string and specifically the axis along which $\phi = 0$, corresponding to the minimum of $M(r,\phi)$. As $m \to \mu$, we find that there is a sort of phase transition in which case the zero mode completely delocalizes onto a two-dimensional wedge. Finally, the mode completely disappears for $m > \mu$.

To better understand the counterintuitive results shown in Fig. 4.2, in §4.3.1 we first solve (4.28) by treating m as a small perturbation. As in §4.2.2, we will see that it is the presence of the axion string, and not a vanishing mass at the core, which allows for this zero mode to exist. In §4.3.2, we describe a simple analog of this mechanism for Dirac fermions in the three-dimensional half-space $\mathbb{R}^{2,1}_+$. In §4.3.3, we analyze the zero mode for the critical case $m = \mu$.

4.3.1 The Small Mass Limit

In this section, we will argue that a normalizable zero mode solution to (4.28) exists in the small Dirac mass limit by analyzing the structure of (4.28) when $m \ll \mu$. Expanding each spinor into Fourier modes,

$$\psi_\alpha(r,\phi) = \sum_{\ell\in\mathbb{Z}} \psi_{\alpha,\ell}(r) e^{i\ell\phi} , \tag{4.29}$$

the zero mode equation (4.28) reduces to an infinite set of coupled equations

$$\begin{aligned} 0 &= -m\psi_{0,\ell-1} + yf(r)\psi_{0,\ell} + i\partial_r\psi_{3,\ell} + i\ell r^{-1}\psi_{3,\ell} \\ 0 &= -m\psi_{3,\ell+1} + yf(r)\psi_{3,\ell} - i\partial_r\psi_{0,\ell} + i\ell r^{-1}\psi_{0,\ell} . \end{aligned} \tag{4.30}$$

When $m = 0$, the length scale over which the zero mode varies is set by the Yukawa mass $\mu = yv$, which dominates the system of equations (4.30) as $r \to \infty$. The mass m will also be relevant as $r \to \infty$, and so our perturbative expansion will be in powers of m/μ.

We will thus search for solutions of the form

$$\psi(x) = \sqrt{\frac{p}{2}}\left[\begin{pmatrix}1\\0\\0\\-i\end{pmatrix}F(r) + \delta\psi(r,\phi)\right]e^{-ip(t+z)}\,, \tag{4.31}$$

and work to first order in $\delta\psi = \mathcal{O}(m/\mu)$. In this limit, (4.30) reduces to two identical sets of differential equations,

$$\begin{aligned}\left(\frac{d}{dr} - \frac{1}{r}\right)G(r) &= yf(r)H(r)\\ \left(\frac{d}{dr} + \frac{1}{r}\right)H(r) &= yf(r)G(r) - mF(r)\end{aligned}\,, \tag{4.32}$$

where $(G,H) = (\delta\psi_{0,1}, -i\delta\psi_{3,1})$ or $(G,H) = (i\delta\psi_{3,-1}, -\delta\psi_{0,-1})$. All other equations in (4.30) are trivially satisfied by setting the other $\delta\psi_{\alpha,\ell}(r) = 0$. The function $G(r)$ satisfies the inhomogeneous equation

$$\left[\frac{d^2}{dr^2} - \frac{f'(r)}{f(r)}\frac{d}{dr} + \frac{f'(r)}{rf(r)} - y^2f^2(r)\right]G(r) = -myf(r)F(r)\,, \tag{4.33}$$

which can be solved using solutions to the homogeneous equation (with $m = 0$) via variation of parameters [863]. These homogeneous solutions will determine how the inhomogeneous solution behaves, and so it will be helpful to understand their asymptotic behavior.

Far from the string, at $r \to \infty$, the homogeneous equation reduces to

$$\left[\frac{d^2}{dr^2} - \mu^2 + \cdots\right]G(r) = 0\,, \tag{4.34}$$

in which case the general solution behaves as

$$G(r) = C_-G_-(r) + C_+G_+(r) \sim C_-e^{-\mu r} + C_+e^{\mu r}\,, \qquad r \to \infty\,, \tag{4.35}$$

where we have again introduced the two linearly-independent solutions $G_\pm(r)$ with definite scaling as $r \to \infty$. We will choose their normalization such that the Wronskian is

$$W(r) = 2yf(r)\,. \tag{4.36}$$

Likewise, as $r \to 0$, the homogeneous equation simplifies to

$$\left[\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{1}{r^2} + \cdots\right]G(r) = 0\,. \tag{4.37}$$

As before, the presence of the axion string introduces a regular singularity at $r = 0$, and solutions take the form

$$G_\pm(r) = C_{\pm,1}\sum_{k=0}^{\infty} a_k^\pm r^{k+1} + C_{\pm,2}\sum_{k=0}^{\infty} b_k^\pm r^{k+1}\log r\,. \tag{4.38}$$

Crucially, both of the linearly-independent homogeneous solutions are regular as $r \to 0$, regardless of the overall coefficients $C_{\pm,1}$ or $C_{\pm,2}$. The most general solution to (4.33) that decays as $r \to \infty$ is

$$G(r) = \frac{1}{2} m C_- G_-(r) + \frac{1}{2} m\, G_+(r) \int_r^\infty dr'\, F(r') G_-(r') + \frac{1}{2} m\, G_-(r) \int_0^r dr'\, F(r') G_+(r')\,, \tag{4.39}$$

for arbitrary values of C_-. Since $F(r)$ and $G_-(r)$ both decay as $e^{-\mu r}$, while $G_+(r')$ grows as $e^{\mu r}$, all three terms in (4.39) decay as $e^{-\mu r}$ as $r \to \infty$. Likewise, both $F(r)$ and the homogeneous solutions $G_\pm(r)$ are regular as $r \to 0$, and thus so is (4.39). This solution is thus normalizable.

Given a solution for $G(r)$, we can solve for the other profile function $H(r)$ via

$$H(r) = \frac{1}{y f(r)} \left(\frac{d}{dr} - \frac{1}{r} \right) G(r)\,. \tag{4.40}$$

This equation should cause some alarm: since $f(r) \propto \sqrt{\lambda} v r$ diverges as $r \to 0$, we might worry that $H(r)$ is *not* regular as $r \to 0$ even though $G(r)$ is. However, the free coefficient C_- in (4.39) can generally be chosen to yield a normalizable $H(r)$. Writing (4.40) as

$$\begin{aligned} H(r) = \frac{m}{2 y f(r)} \Bigg[& C_- \, r \frac{d}{dr} \left(\frac{G_-(r)}{r} \right) + r \frac{d}{dr} \left(\frac{G_+(r)}{r} \right) \int_0^\infty dr'\, F(r') G_-(r') \\ & - r \frac{d}{dr} \left(\frac{G_+(r)}{r} \right) \int_0^r dr'\, F(r') G_-(r') + r \frac{d}{dr} \left(\frac{G_-(r)}{r} \right) \int_0^r dr'\, F(r') G_+(r') \Bigg]\,, \end{aligned} \tag{4.41}$$

the combinations

$$r \frac{d}{dr} \left(\frac{G_\pm(r)}{r} \right) = C_{\pm,1} \sum_{k=0}^\infty a_k^\pm k r^k + C_{\pm,2} \sum_{k=0}^\infty b_k^\pm (1 + k \log r) r^k\,, \tag{4.42}$$

are crucially regular as $r \to 0$, approaching a constant $b_0^\pm$. Furthermore, since each of the integrals in the second line of (4.41) decay as $\propto r^2 \log r$ as $r \to 0$, each of those terms vanish as $r \to 0$ and thus do not generate a non-normalizable contribution to $H(r)$.

In contrast, the terms in the first line are dangerous: they can contain a constant piece which, when divided by $f(r)$, would cause $H(r) \sim H_0/r$ as $r \to 0$ for some constant H_0. Note, however, that we can always[1] choose C_- so that the overall constant term in the square braces of (4.41) cancels. In this case, $H(r) \sim H_0 \log r$ as $r \to 0$, for some constant H_0. But, this is still square-integrable, and so we conclude that we can construct, at least perturbatively, a normalizable zero mode solution in the presence of a small, non-zero Dirac mass m. As in the unperturbed case reviewed in §4.2, the axion string modifies the wave

[1]The only way this strategy could fail is if both $G_\pm(r)$ also have definite scaling behavior as $r \to 0$, such that either $a_k^+ = b_k^- = 0$. In this case, we cannot tune C_- to cancel the constant and the solution is non-normalizable. While this behavior is extremely non-generic, we would need the connection formulae for (4.33) to prove it does not happen. We take our numerical solutions to be proof that it does not.

equation so that both linearly-independent solutions are regular as $r \to 0$, allowing for a normalizable zero mode.

The numerical solutions presented in Fig. 4.2 demonstrate that these solutions exist non-perturbatively as well, as long as $m < \mu$. As $m \to \mu$, these solutions become unbound from the string and occupy a two-dimensional "wedge" in the plane orthogonal to the string. To better understand the critical case $m = \mu$, and why these solutions exist at all, it will be helpful to first study a simpler, albeit analogous, system in which normalizable zero modes exist even in the presence of a nowhere vanishing gap.

4.3.2 Zero Modes on Domain Walls

That a massless field can emerge from one that is everywhere massive is counterintuitive, so it will be helpful to consider a simpler system where it also occurs: the Dirac fermion on the half-plane $\mathbb{R}^{2,1}_{+}$ [838, 864–866]. Studying this system[2] will also help us understand why these zero modes completely delocalize onto a two-dimensional wedge when $m = \mu$, which we will discuss in detail in §4.3.3.

Let us first consider a single Dirac fermion in the full space $\mathbb{R}^{2,1}$ with a spatially-varying mass, and Lagrangian

$$\mathcal{L} = \bar{\psi}\big(i\partial\!\!\!/ - m(x)\big)\psi\,. \tag{4.43}$$

Here, $(\gamma^0, \gamma^1, \gamma^2) = (\sigma_1, i\sigma_3, -i\sigma_2)$ are the three-dimensional γ-matrices and we use the Cartesian coordinates (t, x, z). We take the mass term $m(x)$ to only depend on the coordinate x. We can search for zero modes that move at the speed of light in the $(+z)$-direction by assuming that there exists a solution of the form

$$\psi(x) = \sqrt{p}\, e^{-ip(t-z)} \begin{pmatrix} \eta(x) \\ \chi(x) \end{pmatrix}\,. \tag{4.44}$$

with $\eta(x)$ and $\chi(x)$ functions of x. With this ansatz, (4.43) reduces to

$$\begin{aligned} 0 &= \eta' + m(x)\eta - 2p\chi \\ 0 &= \chi' - m(x)\chi \end{aligned}\,, \tag{4.45}$$

where the $'$ denotes differentiation with respect to x. This solution must have definite helicity, with $\chi(x) = 0$, so that solutions take the form

$$\eta(x) = \mathcal{A} \exp\left(-\int_0^x dx'\, m(x')\right)\,, \tag{4.46}$$

where the coefficient $\mathcal{A}$ is determined by a normalization condition analogous to (4.15).

For (4.46) to yield a normalizable zero mode propagating in the $(+z)$-direction, the mass $m(x)$ must be

[2] This system naturally appears, for instance, in the study of the quantum anomalous hall effect [867] along interfaces between topological and regular insulators [866].

positive as $x \to \infty$ and negative as $x \to -\infty$, and thus *vanish* for some value of x. Likewise, if the mass is instead negative as $x \to \infty$ and positive as $x \to -\infty$, there is a normalizable zero mode propagating in the $(-z)$-direction. Thus, if there exists a *domain wall* in which the fermion mass crosses through zero, there will be chiral fermionic zero modes—often called chiral domain wall fermions—localized and propagating along it.[3] We see that in $\mathbb{R}^{2,1}$, the fermion cannot be everywhere gapped and still yield a normalizable zero mode. This is one case in which vanishing mass and the existence of fermionic zero modes are inextricably linked.

However, the situation changes if we restrict to the half-space $\mathbb{R}^{2,1}_+$, with $x \geq 0$. In this case, constant mass $m(x) = \mu$ *does* yield a normalizable zero mode which is exponentially localized to the boundary at $x = 0$. For example, by decoupling (4.45) we find that η obeys

$$\eta'' + \big(m'(x) - m^2(x)\big)\eta = 0\,. \tag{4.47}$$

To match onto the analysis of §4.2.2 and §4.3.1, we note that, since $m(x)$ approaches a constant μ at infinity, then $\eta(x) \sim C_- e^{-\mu x} + C_+ e^{\mu x}$ as $x \to \infty$. Likewise, as long as $m(x)$ is regular as $x \to 0$, both linearly-independent solutions will be regular, and so we will always be able to construct a normalizable solution to (4.47). When $m(x) = \mu$ is constant everywhere on $\mathbb{R}^{2,1}_+$, this solution is just $\eta(x) = \mathcal{C}_- e^{-\mu x}$. In contrast, this fails on the full-space $\mathbb{R}^{2,1}$ because the solution that exponentially decays as $x \to \infty$ never matches onto the one which decays as $x \to -\infty$, unless the mass $m(x)$ switches sign for some x. We note in passing that the existence of this edge mode in the half-plane depends also on the choice of boundary conditions for the fermion; see e.g., [869].

To connect this simple system to the one we are interested in (4.26), we can picture the polar radial coordinate in $\mathbb{R}^{3,1}$ as the analog of the x coordinate on the half-space $\mathbb{R}^{2,1}_+$, while the axion string itself provides an effective "mass" $m(r)$ which is regular as $r \to 0$, as are the solutions to (4.26). The same mechanism is at work for both cases: both linearly-independent solutions are regular at the core of the string or wall, and thus the solution that decays at spatial infinity is necessarily normalizable. As we discuss in the next section, this simple system is also useful for understanding the critical case in which core and Yukawa masses are equal, $m = \mu$, and the zero modes become completely delocalized from the string.

4.3.3 The Critical Mass Case

From the numerical results presented in Fig. 4.2, we found that there is a phase transition that occurs as we tune the core mass to the Yukawa mass, $m \to \mu$, wherein the fermionic zero mode seemingly delocalizes from

[3] This is a well-known fact which has been exploited [864] to simulate chiral fermions on the lattice. Many of the phenomena we find in this chapter have analogs there. For instance, these chiral modes only exist for momenta $\mathbf{p}$ in a particular region of the Brillouin zone [868], which we can denote as $\mathbf{p} \in \mathcal{C}$. For $\mathbf{p} \notin \mathcal{C}$, the chiral modes cease to exist. In analogy with Fig. 4.2, the chiral modes have wavefunctions that are well-localized to the domain wall for momenta comfortably inside $\mathcal{C}$, while they completely delocalize from the domain wall as $\mathbf{p}$ approaches the boundary $\partial\mathcal{C}$, and are no longer normalizable for $\mathbf{p} \notin \mathcal{C}$.

the axion string. Beyond this critical mass, $m > \mu$, the zero mode ceases to exist. Since this critical case turns out to be very difficult to study numerically, it will instead be helpful to attack this case analytically to understand exactly how these modes behave when $m = \mu$.

First, however, it will be useful to qualitatively understand how solutions to (4.26) should behave in the limit $m \to \mu$. In the previous section, we described how a fermionic zero mode can arise whenever a fermion's spatially-dependent mass crosses through zero. At criticality and far from the string, the mass (4.27) approaches

$$\lim_{r\to\infty} M(r,\phi) = \sqrt{(m-\mu\cos\phi)^2 + \mu^2\sin^2\phi}\,, \tag{4.48}$$

which crosses through zero at $\phi = 0$ when (and only when) $m = \mu$. From the logic of the previous section, we should then expect that the zero modes are no longer localized along the axion string, but are instead allowed to propagate freely along—and are localized to—the plane defined by $\phi = 0$. Thus, we expect that these zero modes may have momentum along both the x- and z-directions, which we denote p_x and p_z, respectively, with frequency $\omega = \sqrt{p_x^2 + p_z^2}$.

We can exhibit these solutions in the critical case by searching for solutions to the equation of motion (4.26) far from the string along the positive x-axis. In this limit, the equations of motion become approximately translationally invariant in both the x- and z-directions, and so we may search for solutions of the form

$$\psi(x) = \psi(x,y)\,e^{-i\omega t + ip_x x + ip_z z}\,, \tag{4.49}$$

where $\psi(x,y)$ varies slowly along the x-direction. Since these modes must have definite helicity, we will assume an ansatz of the form

$$\psi(x,y) \approx \sqrt{\frac{\omega}{2}}\begin{pmatrix}\cos\alpha\\ \sin\alpha\\ i\sin\alpha\\ -i\cos\alpha\end{pmatrix} F(x,y) \tag{4.50}$$

with $\cos 2\alpha = -p_z/\omega$, and $\sin 2\alpha = -p_x/\omega$. With this ansatz, when far from the string (4.26) reduces to

$$\partial_y F(x,y) \sim -\mu(y/x)F(x,y)\,, \qquad x \gg |y|\,, \tag{4.51}$$

where we have dropped derivatives with respect to x and approximated $e^{i\phi} \approx 1 + i(y/x)$. This can be solved to find

$$F(x,y) \approx \mathcal{A}(x)\,\exp\left(-\int_0^{|y|} dy'\,M(x,y)\right) \sim \left(\frac{\mu}{\pi x}\right)^{\frac{1}{4}}\exp\left(-\frac{\mu y^2}{2x}\right), \tag{4.52}$$

for $x \gg |y|$, where $M(x,y) = \mu\sqrt{(1-\cos\phi)^2 + \sin^2\phi} \approx \mu(y/x)$ is the mass (4.48) far from the string, where $\mu x \gg 1$. The overall amplitude $\mathcal{A}(x)$ is determined by imposing the normalization condition,

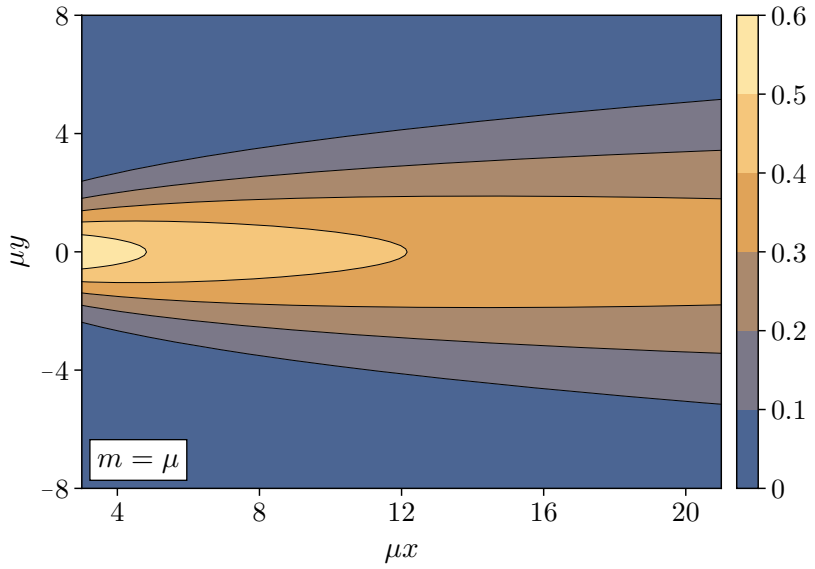


Figure 4.3: Plot of the approximate critical wavefunction $|F(x, y)|$, trustable for large positive x, with contours displayed in units of $\sqrt{\mu}$. When $p_x = 0$ and $p_z = -\omega$, $|F(x, y)| = |\psi_0(x, y)|$, and so this figure can be directly compared to the bottom left panel of Fig. 4.2.

$$\int d^3x\, \psi^\dagger_{\mathbf{p}}(x)\psi_{\mathbf{p}'}(x) = (2\pi)^2\omega\, \delta^{(2)}(\mathbf{p} - \mathbf{p}') \tag{4.53}$$

at any constant t, the direct analog of (4.15). For the ansatz (4.50), this translates into the requirement that $\int dy\, |F(x, y)|^2 = 1$.

We plot this approximate critical wavefunction (4.52) in Fig. 4.3 and find that it matches with the $m \to \mu$ behavior we observe in the numerical results shown in Fig. 4.2. As our qualitative arguments suggested, we find that the zero modes delocalize from the string once $m = \mu$ and instead are free to move along the half-plane defined by $\phi = 0$. They are relatively well-localized along $y = \phi = 0$, but become more and more spread out the further we get from the string, with an approximate width of $\Delta y \sim \sqrt{x}/\mu$.

To summarize, in this section we studied how fermionic zero modes along an axion string respond to the addition of a non-zero Dirac mass m. Absent this core mass, the four-dimensional fermion sees effectively zero mass at the core of the string, and so one might expect that this vanishing mass explains why the axion string supports fermionic zero modes. However, we showed that these zero modes exist even when the four-dimensional fermion has a core mass m and is thus everywhere massive, and we explained what it is about the string that allows these modes to exist. Furthermore, we found that these zero modes completely delocalize from the string as $m \to \mu$, and cease to exist for $m > \mu$. In the next section, we explain this phase structure from the perspective of anomaly inflow, in which these zero modes are necessary to render the low-energy effective theory consistent.

4.4 Zero Modes from Anomaly Inflow

The existence of these zero modes can also be inferred on topological grounds, based on the logic of anomaly inflow [838]. Under an infinitesimal U(1) gauge transformation, the fields in (4.25) transform as

$$\delta_\Lambda \psi = ie\Lambda(x)\psi\,, \qquad \delta_\Lambda A_\mu = \partial_\mu \Lambda(x)\,, \tag{4.54}$$

where $\Lambda(x)$ is the U(1) gauge parameter. In a topologically trivial background, the action (4.25) is manifestly invariant under such a transformation. In the presence of the axion string, however, gauge invariance is more subtle. This symmetry is *anomalous* in the presence of the string, such that electric charge is not conserved in a region localized to the string. The theory is then inconsistent unless there are some anomalous degrees of freedom, i.e. charged chiral excitations, that are localized to the string and can cancel this anomaly and carry away electric charge. These zero modes then allow electric charge and the anomaly to "flow" out of the bulk spacetime and onto the string, ensuring that the full theory remains consistent.

The general strategy is as follows. We will attempt to construct a low-energy effective field theory of the axion and abelian gauge field by integrating out both the fermion $\psi(x)$ and radial mode $f(x)$ of the complex scalar. We do this by first *ignoring* the contribution from possible fermionic zero modes and by carefully considering the effect of a gauge transformation on the effective theory. Our arguments are similar to the original story, which can be found in [838,840,845], except that the $U(1)_{pq}$ breaking mass leads to a subtlety in the identification of the phase which should be rotated away. Depending on the relative size of m and μ, we find a gauge anomaly localized to the string. This inconsistency then forces us to include an additional contribution from fermionic zero modes localized to the string which cancels this anomaly.

We are interested in the gauge invariance of the low-energy effective theory for the axion field after integrating out the fermions. In the string background (4.3), the fermion sector of the theory can be written

$$\mathcal{L} \supset \bar{\psi}\big(i\not{D} - m + yf(r)e^{i\gamma^5\theta(x)}\big)\psi, \tag{4.55}$$

with $\theta(x)$ the axion field, and so we wish to compute

$$\mathcal{Z}_\psi(\theta) = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi\, \exp\left[i\int d^4x\, \bar{\psi}\big(i\not{D} - M(r,\theta)\, e^{i\gamma^5\alpha(r,\theta)}\big)\psi\right], \tag{4.56}$$

with $M(r,\theta)$ and $\alpha(r,\theta)$ defined in (4.27). Performing the path integral by introducing a Pauli-Villars regulator with mass $\tilde{M}$, we find

$$\mathcal{Z}_\psi(\theta) = \frac{\det\Big(i\not{D} - M(r,\theta)e^{i\gamma^5\alpha(r,\theta)}\Big)}{\det\Big(i\not{D} - \tilde{M}\Big)}\,. \tag{4.57}$$

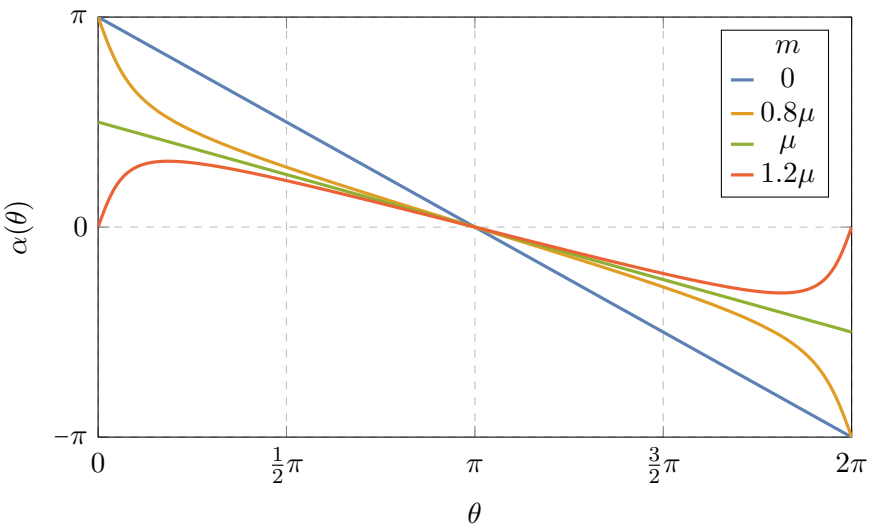


Figure 4.4: The phase of fermion "mass" (4.59) far from the string. The phase is double-valued at $\theta = 2\pi \sim 0$ for $m \leq \mu$ [blue, orange, green]. The zero mode solutions vanish as $m > \mu$ [red] once the phase becomes single-valued.

Now we perform a spatially-dependent chiral field redefinition, $\psi \to e^{-i\gamma^5\alpha(r,\theta)/2}\psi$ to try and remove $\mathcal{Z}_\psi(\theta)$'s dependence on $\alpha(r,\theta)$. However, due to the chiral anomaly, this transformation introduces a Jacobian factor and we thus have

$$\mathcal{Z}_\psi(\theta) = \frac{\det\left(i\not{D} - M(r,\theta)\right)}{\det\left(i\not{D} - \tilde{M}\right)} \exp\left[\frac{i}{8\pi^2}\int \alpha(r,\theta) F \wedge F\right], \tag{4.58}$$

which can be written in the form of an effective action for the axion θ interacting with the gauge field A_μ. Note that in the $m \to 0$ limit, in which there is a classical $U(1)_{pq}$ symmetry, $\alpha(r,\theta) \to \theta$ and this Jacobian factor reduces to the usual quantized coupling of the axion to the gauge field, $\propto \theta F \wedge F$. Furthermore, integrating out the fermions leads to a nontrivial effective potential for the axion via the θ-dependence in $M(r,\theta)$.

It is this additional $\alpha(r,\theta)F \wedge F$ term in the effective action that leads to a potential issue with gauge invariance. Whether or not this term is gauge invariant is determined by whether the phase far from the string,

$$\alpha(\theta) = \lim_{r\to\infty} \alpha(r,\theta) = \arg(m - \mu e^{-i\theta}), \tag{4.59}$$

is single- or multi-valued as a function of θ, and this depends on the relative size of m and μ as illustrated in Fig. 4.4. For instance, in the limit that $m \ll \mu$, we find

$$\alpha(\theta) \approx \pi - \theta, \tag{4.60}$$

which is clearly not single-valued at $\theta = 2\pi \sim 0$, and this applies more generally for all $m < \mu$. In this case,

we must integrate the effective action by parts to make it well-defined,

$$\frac{1}{8\pi^2}\int \alpha(\theta)\mathrm{F}\wedge\mathrm{F} \to -\frac{1}{8\pi^2}\int \mathrm{d}\alpha(\theta)\wedge\mathrm{A}\wedge\mathrm{F}\,. \tag{4.61}$$

However, under the gauge transformation (4.54), this transforms as

$$\delta_\Lambda\Big[-\frac{1}{8\pi^2}\int \mathrm{d}\alpha(\theta)\wedge\mathrm{A}\wedge\mathrm{F}\Big] = -\frac{1}{8\pi^2}\int \mathrm{d}\alpha(\theta)\wedge\mathrm{d}(\Lambda\mathrm{F}) = +\frac{1}{8\pi^2}\int \mathrm{d}^2\alpha(\theta)\wedge(\Lambda\mathrm{F})\,, \tag{4.62}$$

where we have integrated by parts again in the last equality, neglecting any potential surface terms at the string core. Since $\alpha(\theta)$ is not single-valued, $\mathrm{d}^2\alpha(\theta)\neq 0$ but is instead localized along the string, and we find an apparent gauge anomaly.

The reason for this apparent anomaly is that we have heretofore incorrectly assumed that there are no zero modes at the core of the string. Their existence modifies the chiral transformation in (4.58) and, with a careful treatment of the variation at the core,[4] we find a violation of gauge invariance precisely equal and opposite to the gauge anomaly due to a massless chiral fermion localized to the string worldsheet,

$$\delta_\Lambda\Big[-\frac{1}{8\pi^2}\int \mathrm{d}\alpha(\theta)\wedge\mathrm{A}\wedge\mathrm{F}\Big] = \frac{1}{4\pi}\Lambda\,\mathrm{F}\,. \tag{4.63}$$

On the other hand, it is precisely when $\mathrm{m}>\mu$ that the phase $\alpha(\theta)$ becomes single-valued since, near the end points at $\theta=0$ and 2π, we have $\alpha(0)=\alpha(2\pi)=\arg(\mathrm{m}-\mu)=0$. In this case the interaction $\propto\alpha(\theta)\,\mathrm{F}\wedge\mathrm{F}$ is both well-defined and single-valued, and gauge invariance is maintained without the need for additional degrees of freedom along the string. This analysis breaks down as $\mathrm{m}\to\mu$ because it no longer makes sense to define an effective field theory for the axion and gauge field "far" from the string since, as is evident from our numerical results in Fig. 4.2 and is discussed in §4.3.3, the fermionic zero modes are no longer localized at small r.

It is worth noting that, from the IR perspective, these anomaly inflow arguments do not entirely determine the theory on the string. While we have focussed throughout on a particular UV model, in which the zero modes are unambiguously identified with chiral modes of the fermion, in general, the anomaly could be cancelled by a different theory on the string worldsheet. For strings with an even winding number, for instance, the anomaly can be cancelled by compact bosons living on the worldsheet. Anomaly inflow guarantees that the IR theory has a U(1) symmetry with a particular anomaly, while the precise theory on the

[4]Somewhat famously, arriving at the exact value of the gauge anomaly, which matches the contribution from two-dimensional chiral fermions localized on the string worldsheet, requires some care. A proper treatment involves introducing a "bump-form" which extends the validity of the effective theory in (4.58) to all of spacetime and properly includes the zero modes on the string core so that all surface contributions vanish [845]. This bump form can be computed explicitly in terms of the zero modes profiles, and the calculation of [845] is unchanged in the case $\mathrm{m}\neq 0$.

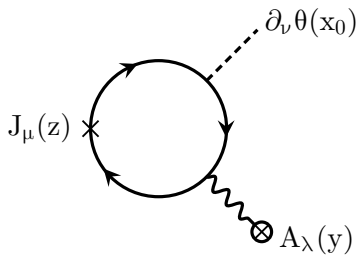


Figure 4.5: Leading order diagram for the current $\langle J^\mu(z)\rangle$ produced by a spatial variation of the axion $\partial_\nu\theta(x_0)$ in the presence of a background electromagnetic field A_λ, used by Goldstone and Wilczek to compute the effective action.

string depends on the details of the model in the UV.

Finally, it is illuminating to compare our effective action to the effective action computed via the method of Goldstone and Wilczek [870], as in [838, 840]. In that case, one computes the current at a point x_0 far from the string core, in response to a background electromagnetic field, in the long wavelength limit. We choose $\theta(x_0) = 0$, so that the fermion-scalar interaction can be expanded as

$$y\bar{\psi}f(r)e^{i\gamma_5\theta(x)}\psi \simeq i\mu\,\partial_\lambda\theta(x_0)(w-x_0)^\lambda\bar{\psi}\gamma^5\psi. \tag{4.64}$$

The leading contribution comes from the diagram in Fig. 4.5. For the theory (4.25), this yields

$$\begin{aligned}\langle J^\mu(z)\rangle = e^2\mu\,\partial_\lambda\theta(x_0)\int\frac{d^4p}{(2\pi)^4}\frac{d^4q}{(2\pi)^4}\frac{d^4k}{(2\pi)^4}\,d^4y\,d^4w\,(w-x_0)^\lambda A_\nu(y)e^{-i[p(z-y)+(p+q)(y-w)+(p+q+k)(w-z)]}\\ \times\mathrm{Tr}\left[\frac{1}{\not p-M}\gamma^\nu\frac{1}{\not p+\not q-M}\gamma^5\frac{1}{\not p+\not q+\not k-M}\gamma^\mu+\frac{1}{\not p-M}\gamma^5\frac{1}{\not p+\not k-M}\gamma^\nu\frac{1}{\not p+\not k+\not q-M}\gamma^\mu\right]\end{aligned} \tag{4.65}$$

and evaluates to

$$\langle J^\mu(z)\rangle = \frac{e^2}{8\pi}\frac{\mu}{\mu-m}\varepsilon^{\mu\nu\kappa\lambda}\partial_\nu\theta(x_0)F_{\kappa\lambda}\,. \tag{4.66}$$

Following [840], we can posit an effective action (valid far from the string),

$$S_{\text{eff}} = \frac{1}{8\pi^2}\frac{\mu}{\mu-m}\int d\theta\wedge A\wedge F\,, \tag{4.67}$$

from which the current (4.66) can be computed. When $m = 0$, integrating (4.67) by parts yields the usual, properly quantized, coupling of the axion to the gauge field. When $m \neq 0$, the $\mu/(\mu-m)$ prefactor would appear to violate the quantization condition for the axion–gauge field coupling. However, expanding for small values of the axion field,

$$d\alpha(\theta) = \left[\frac{\mu}{\mu-m}+\mathcal{O}(\theta^2)\right]d\theta\,, \tag{4.68}$$

and we see that (4.67) is precisely the leading term of (4.61), which is properly quantized.

Let us note that there is an ambiguity in which field one calls the "axion," θ or $\alpha(\theta)$. Both are compact fields, $\theta \sim \theta + 2\pi$ and $\alpha \sim \alpha + 2\pi$. If one works in the "charge-quantized" basis in which the coefficient of the topological coupling (4.61) is an integer, the axion α has a (possibly highly) non-trivial metric on its field space given by

$$g_{\alpha\alpha} = v^2(\theta'(\alpha))^2 = v^2 \left[1 + \frac{m}{\mu} \frac{\cos\alpha}{\sqrt{1 - (m/\mu)^2 \sin^2\alpha}}\right]^2 , \tag{4.69}$$

and so around $\theta \approx 0$ or $\alpha \approx \pi$ the decay constant is effectively reduced by a factor of $(\mu - m)/\mu$, which is identical to the factor that appears in (4.67). If one instead works with θ, the (classical) metric on field space is flat but the coefficient of the topological coupling (4.67) is not an integer. This story is similar in spirit to the reason for the non-quantized coupling of the axion to photons, due to mixing with the neutral pion.

To summarize, we have shown that—with a careful treatment of the axion's periodicity—the arguments for anomaly inflow, and the resulting necessity of chiral zero modes on the string, persist in the presence of a field-independent mass, as long as this "bulk" mass is smaller than the mass generated by the chirally coupled scalar field. This makes it clear that it is the topology of the axion string, and not the dynamics of the scalar field which spontaneously breaks the classical $U(1)_{pq}$ symmetry, which is responsible for the existence of the fermion zero modes.

4.5 Low-Energy Effective Theory

In the previous section, we found that the full four-dimensional effective theory was rendered consistent when $m < \mu$ by including chiral fermionic zero mode localized to the axion string, while they disappeared for $m > \mu$. Our numerical results in Fig. 4.2 show that they disappear because the zero mode completely delocalizes from the string as $m \to \mu$. It will be useful to understand how this process appears from the zero mode's two-dimensional effective theory on the string worldsheet. At leading order in μ^{-1}, this action is completely constrained by symmetry and does not depend on the ratio m/μ. However, we will show that the Wilson coefficients of the higher derivative interactions between the zero mode and the gauge field depend sensitively on m/μ and diverge as $m \to \mu$, causing the two-dimensional effective theory to break down.

We again will specialize to an axion string with charge $+1$, and start by introducing the (quantized) mode expansion for the fermion

$$\psi(x) = \int_0^\infty \frac{dp}{2\pi} \frac{1}{\sqrt{2|p|}} \hat{a}_p \psi_p(r,\phi) e^{-ip(t+z)} + \int_{-\infty}^0 \frac{dp}{2\pi} \frac{1}{\sqrt{2|p|}} \hat{b}^\dagger_{-p} \psi_p(r,\phi) e^{-ip(t+z)} + \cdots \tag{4.70}$$

where the $\cdots$ denote the non-zero modes of the massive four-dimensional fermion about the axion string.

Here, $\psi_{\rm p}(\mathrm{r},\phi)$ is the zero mode wavefunction (4.10) found in §4.3, subject to the normalization condition (4.15). The creation and annihilation operators, $\hat{\mathrm{b}}^\dagger_{\rm p}$, $\hat{\mathrm{a}}_{\rm p}$ satisfy the canonical commutation relations appropriate for a *two-dimensional* fermion,

$$\{\hat{\mathrm{a}}_{\rm p},\hat{\mathrm{a}}^\dagger_{\rm p'}\}=2\pi\,\delta(\mathrm{p}-\mathrm{p}'),\qquad \{\hat{\mathrm{b}}_{\rm p},\hat{\mathrm{b}}^\dagger_{\rm p'}\}=2\pi\,\delta(\mathrm{p}-\mathrm{p}'). \tag{4.71}$$

Since the only place the momentum p enters into $\psi_{\rm p}(\mathrm{r},\phi)$ is in its overall normalization, we can define the spinor $\mathcal{F}(\mathrm{r},\phi)=\psi_{\rm p}(\mathrm{r},\phi)/\sqrt{\mathrm{p}}$ and write (4.70) as $\psi(\mathrm{x})=\chi_-(\mathrm{t},\mathrm{z})\mathcal{F}(\mathrm{r},\phi)+\cdots$, where

$$\chi_-(\mathrm{t},\mathrm{z})=\int_0^\infty\frac{\mathrm{dp}}{2\pi}\left[\hat{\mathrm{a}}_{\rm p}\,\mathrm{e}^{-\mathrm{ip(t+z)}}+\hat{\mathrm{b}}^\dagger_{\rm p}\,\mathrm{e}^{\mathrm{ip(t+z)}}\right] \tag{4.72}$$

is a canonically normalized two-dimensional fermion field operator with negative chirality.

It will be convenient to repackage $\chi_-(\mathrm{t},\mathrm{z})$ into a two-dimensional Dirac fermion $\chi(\mathrm{t},\mathrm{z})$ living on the axion string worldsheet

$$\chi(\mathrm{t},\mathrm{z})=\begin{pmatrix}\chi_-(\mathrm{t},\mathrm{z})\\ \chi_+(\mathrm{t},\mathrm{z})\end{pmatrix}, \tag{4.73}$$

by grouping it with a positive chirality fermion $\chi_+(\mathrm{t},\mathrm{z})$, which we later set to zero. We will use $\mathrm{a},\mathrm{b},\ldots=0,3$ to denote worldsheet indices, and define our worldsheet γ-matrices as $\gamma^{\rm a}=(\tilde\gamma^0,\tilde\gamma^3)=(\sigma_1,\mathrm{i}\sigma_2)$[5] with chiral projectors $\gamma_\pm=\frac{1}{2}(1\pm\tilde\gamma^0\tilde\gamma^3)$ such that $\gamma_\pm\chi=\chi_\pm$. We will thus impose the constraint $\gamma_+\chi=0$ to remove χ_+ from the theory. The kinetic term for the two-dimensional fermion then takes the standard Dirac form,

$$\mathrm{S}_\chi=\int\mathrm{d}^2\sigma\,\mathrm{i}\bar\chi\gamma^{\rm a}\partial_{\rm a}\chi+\cdots=\int\mathrm{d}^2\mathrm{z}\,\mathrm{i}\chi_-^\dagger\partial_+\chi_-+\cdots \tag{4.74}$$

where we denote $\partial_\pm=\partial_{\rm t}\mp\partial_{\rm z}$ and use $\mathrm{d}^2\sigma=\mathrm{dt}\,\mathrm{dz}$ to denote the volume element on the string worldsheet, while we will use $\mathrm{d}^2\mathrm{r}=\mathrm{r}\,\mathrm{dr}\,\mathrm{d}\phi=\mathrm{dx}\,\mathrm{dy}$ to denote the volume element orthogonal to the string. The $\cdots$ denote terms in the effective action that encode how the chiral zero mode χ_- interacts with the gauge field A_μ, which we now derive.

The zero mode generates a four-dimensional current that points along the axion string

$$\mathrm{j}^\mu=\mathrm{e}\bar\psi\gamma^\mu\psi=\begin{cases}\mathrm{e}|\mathcal{F}(\mathrm{r},\phi)|^2(\bar\chi\gamma^{\rm a}\chi)(\mathrm{t},\mathrm{z}) & \mu=\mathrm{a}=0,3\\ 0 & \mu=1,2\end{cases}, \tag{4.75}$$

where $|\mathcal{F}(\mathrm{r},\phi)|^2=(\mathcal{F}^\dagger\mathcal{F})(\mathrm{r},\phi)$, and so its interaction with the gauge field A_μ is determined by

[5] We use $\tilde\gamma^0$ and $\tilde\gamma^3$ to distinguish these two-dimensional γ-matrices from the four-dimensional γ^0 and γ^3, used for the four-dimensional Dirac fermion $\psi(\mathrm{x})$.

$$S_{em} = -\int d^4x\, j^\mu A_\mu = -\int d^4x\, e|\mathcal{F}(r,\phi)|^2(\bar{\chi}\gamma^a\chi)(t,z)A_\mu(x)\,. \tag{4.76}$$

Our goal is to encode the information contained in (4.76) into a series of effective interactions in the two-dimensional effective theory between χ and (derivatives of) the gauge field pulled back onto the string worldsheet. This will encode the zero mode wavefunction's multipolar structure into a set of effective two-dimensional interactions and will define a set of multipole moments for the current (4.75).

To derive these effective interactions, we perform a long-wavelength expansion of the gauge field about the axion string $x = (t, 0, 0, z)$ following a procedure similar to the one outlined in [871] and explained in more detail in Appendix A.2. After using current conservation and integration by parts, (4.76) can be written as

$$\begin{aligned} S_{em} = &-\int d^2\sigma \big[e(\bar{\chi}\gamma^a\chi)A_a\big](t,z) \\ &-\sum_{n=1}^{\infty}\frac{1}{n!}\left[\int d^2r\, x^{k_1}\cdots x^{k_n}\,|\mathcal{F}(r,\phi)|^2\right]\int d^2\sigma\big[e(\bar{\chi}\gamma^a\chi)\partial_{k_2}\cdots\partial_{k_n}F_{k_1 a}\big](t,z), \end{aligned} \tag{4.77}$$

where we use $i, j, k_1, \cdots = 1, 2$ indices to denote directions orthogonal to the string, i.e. $x^i = (x, y)$. It is more illuminating to rewrite the higher order terms in this expansion in terms of irreducible SO(2) tensors. This introduces terms $\propto \partial^k F_{ka}$, which can be rewritten using the equations of motion that follow from (4.58),

$$\partial^k F_{ka} = -\partial^b F_{ba} + \frac{e^2}{16\pi^2}\partial^\mu(\alpha(\theta)\tilde{F}_{a\mu}) + j_a\,, \tag{4.78}$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$ is the dual field strength. To second order in the derivative expansion, we find that the two-dimensional current $\tilde{\jmath}^a \equiv \bar{\chi}\gamma^a\chi$ couples to

$$S_{em} \supset \int d^2\sigma\left[-\tilde{\jmath}^a A_a + I_1^i\,\tilde{\jmath}^a F_{ia} + \frac{1}{4}I_2^{ij}\tilde{\jmath}^a\partial_{(i}F_{j)a} - \frac{1}{4}I_2^{tr}\tilde{\jmath}^a\partial^b F_{ba} + \frac{e^2}{64\pi^2}I_2^{tr}\tilde{\jmath}^a\partial^\rho(\theta\tilde{F}_{a\rho})\right] \tag{4.79}$$

where

$$I_1^i = \int d^2r\, x^i|\mathcal{F}(r,\phi)|^2 \tag{4.80}$$

is the zero mode current's "dipole," while

$$I_2^{ij} = \int d^2r\,(x^i x^j - \tfrac{1}{2}\delta^{ij}x_k x^k)|\mathcal{F}(r,\phi)|^2, \qquad I_2^{tr} = \int d^2r\, r^2|\mathcal{F}(r,\phi)|^2\,, \tag{4.81}$$

are the current's "quadrupole" and variance. Each of the operators in (4.79) are operators on the string worldsheet, while the dimensionless coefficients $c_1^i \equiv \mu I_1^i$, $c_2^{ij} \equiv \mu^2 I_2^{ij}$, $c_2^{tr} \equiv \mu^2 I_2^{tr}$ can be understood as Wilson coefficients in the effective theory on the string worldsheet.[6]

[6]There are only three independent Wilson coefficients that appear at each order in this multipole expansion.

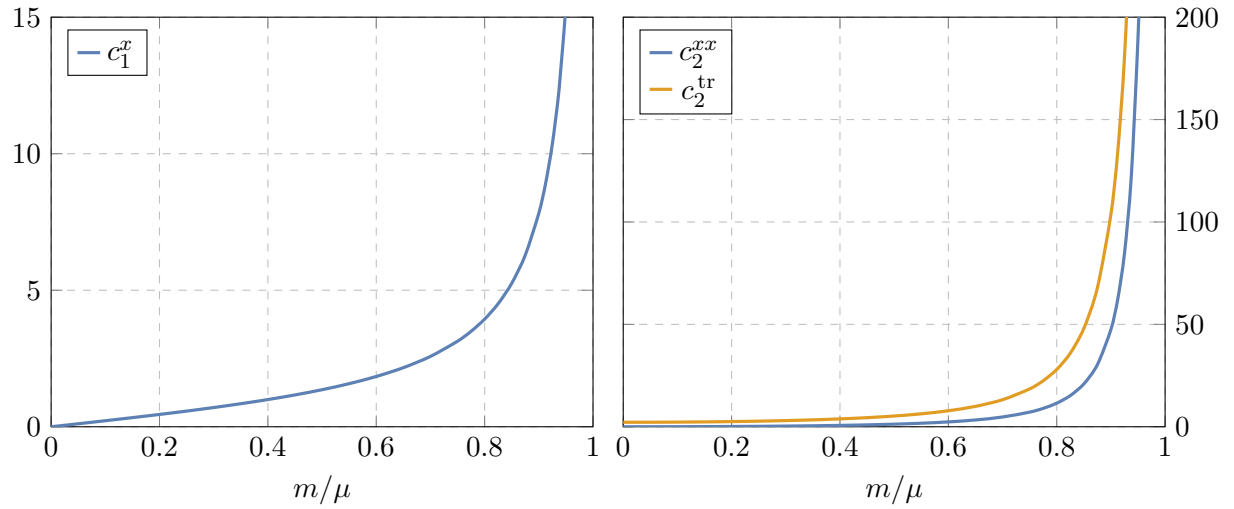


Figure 4.6: Plots of the Wilson coefficients c_1^x, c_2^{xx}, and c_2^{tr} vs. m/μ. The coefficients c_1^y and $c_2^{xy} = c_2^{yx}$ vanish due to the zero mode's reflection symmetry about the x-axis, as evident in Fig. 4.2, while the other coefficient c_2^{yy} is completely determined by $c_2^{yy} = -c_2^{xx}$.

We plot these Wilson coefficients in Fig. 4.6 as functions of m/μ. Two features are immediately obvious: at $m = 0$, the dipole and quadrupole moments vanish due to the cylindrical symmetry of the zero mode solution, as expected. The trace term c_2^{tr} does not vanish but instead approaches a value $c_2^{tr} \to 2.19$ as $m \to 0$. Such trace terms also appear in the generic multipole expansion of a point-like source [871] where, as here, they multiply *time* derivatives of the field strength and thus do not appear in the static multipole expansion.

As m approaches the critical value, $m = \mu$, all three Wilson coefficients quickly blow up, reflecting the breakdown of the two-dimensional effective theory. This is also expected since, as seen in Fig. 4.2, the zero modes delocalize from the string as $m \to \mu$ and must completely disappear for $m > \mu$, and so the worldsheet effective theory must fail. As is usual, the effective theory signals its own demise through Wilson coefficients that become uncontrollably large.

So, far from a charge +1 axion string the theory (4.25) is well-described by the effective action

$$\begin{aligned} S = &\int d^4x \left[-\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + v^2(\partial\theta)^2 + \frac{\alpha(\theta)}{16\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu} + \cdots \right] \\ &+ \int d^2\sigma \left[\bar{\chi} i\not{D}\chi + \frac{c_1^i}{\mu}\tilde{\jmath}^a F_{ia} + \frac{c_2^{ij}}{4\mu^2}\tilde{\jmath}^a \partial_{(i}F_{j)a} - \frac{c_2^{tr}}{4\mu^2}\tilde{\jmath}^a \partial^b F_{ba} + \frac{e^2 c_2^{tr}}{64\pi^2\mu^2}\tilde{\jmath}^a \partial^\rho(\theta\tilde{F}_{a\rho}) + \cdots \right] . \end{aligned} \tag{4.82}$$

The first line describes the dynamics of the gauge field and axion, where the $\cdots$ denote interactions that

This can be most easily seen by remembering that the irreducible tensors of SO(2) with weight ℓ are in one-to-one correspondence with the two Fourier modes $e^{\pm i\ell\phi}$ and so, in general, correspond to two complex (or four real) degrees of freedom, which we may denote as f_ℓ and $f_{-\ell}$. Since the distribution is purely real, these coefficients obey $f_\ell = f_{-\ell}^*$, and so the distribution is described by only two real degrees of freedom at each weight ℓ. To this, we must add the pure trace term at each ℓ, e.g. I_2^{tr}, which increases the number of moments at each order to three.

are induced by, for instance, integrating out the scalar's radial mode f and the non-zero mode fluctuations of the fermions and include a potential for the axion θ. The second line describes the chiral zero mode living on the string and its interactions with the gauge field and axion. Here, the $\cdots$ denote terms that are either subleading in the derivative expansion or are nonlinear in the current j^a, the latter of which are also generated by the long-wavelength expansion (described above and in Appendix A.2) and upon integrating out the heavy scalar and fermionic modes.

4.6 Discussion

In this chapter, we revisited the superconductivity of string solutions in a simple model of axion electrodynamics in which the $U(1)_{pq}$ symmetry is explicitly broken by a mass term for the fermions. When the PQ-breaking mass m is smaller than the asymptotic mass μ acquired from the radial mode of the scalar whose phase is the axion, the fermionic zero modes responsible for superconductivity persist. We demonstrated the existence of these zero modes both by studying the asymptotic behavior of the equations of motion and by solving for their profiles numerically. We also demonstrated how their existence can be understood from anomaly inflow, with some modifications to the original arguments by Callan and Harvey.

For $m > \mu$, the zero modes cease to exist. As m approaches μ, however, the zero modes exhibit an interesting critical behavior, in which they delocalize from the core of the string and propagate along a two-dimensional wedge.[7] We further studied the effective theory of the zero modes on the worldsheet and their interactions with external gauge fields. As m approaches its critical value, we demonstrated explicitly how this effective theory breaks down as a result of the zero modes delocalizing.

While we have focused on a simple "minimal" model of axion superconductivity here, there are several aspects which may be phenomenologically relevant in more realistic theories that warrant further exploration. First, as mentioned in the introduction, DFSZ axion models have multiple distinct, topologically stable string configurations and at least one of these (the "Type-C strings" in [852]) has a scalar profile which does not restore electroweak symmetry at the string core. It is possible that the nonzero vacuum expectation value at the core in these configurations may play a similar role to the explicit PQ-breaking mass term we have considered, which helps explain how these solutions may still be superconducting (as anomaly inflow arguments suggest they must be). This should be studied in more detail.

It is also worth recalling that in a high-temperature background, fermions acquire a Debye mass $\sim gT$ which has the same effect as the explicit mass m. In an expanding universe, if the initial temperature is large

[7] Amusingly, this behavior is very similar to how light fermions behave in the presence of a magnetic monopole [872], but in reverse. There, the light fermionic modes become more and more delocalized from the monopole core as their mass *decreases*, ultimately explaining why the θ-angle becomes a redundant parameter for dyon physics once $m = 0$. Here, counterintuitively, the light fermionic modes delocalize as their mass *increases*.

enough that the Debye mass is larger than the Yukawa mass, a phase transition to the superconducting case may actually occur. It may thus be of interest to study if the delocalization of the zero modes found in the simple model studied here can have consequences for the evolution of string networks in cosmology.

Finally, we should emphasize that our numerical results were derived under the assumption that the fermion amplitude is small. It would be worthwhile to understand how the profiles of the zero modes change beyond the limit of small fluctuations. In the same vein, it would be enlightening to understand the higher-order source terms in the worldsheet effective action that we have dropped. We hope to return to some of these topics in the future.

5

A Closer Look at CP-Violating Higgs Portal Dark Matter

5.1 Introduction

Dark matter constitutes the majority of the mass in our universe, but its properties remain largely unknown. Over the years, there have been tantalizing hints in various experiments; while many of these signals have vanished due to increased statistics or a better understanding of systematic uncertainties, some signals, such as the Galactic Center Excess (GCE), have persisted for over a decade.

In astrophysical settings, the Galactic Center is expected to have some of the largest dark matter densities, and is therefore one of the most promising targets for indirect searches. The GCE is a statistically significant excess of gamma rays at energies of $\sim 2-3$ GeV observed in the Galactic Center by the Fermi Gamma Ray Space Telescope [873]. As pointed out by [874–880], the GCE could be explained by a thermal WIMP annihilating to Standard Model particles. To truly confirm such a hypothesis, it is crucial to observe a signal in other indirect channels. In fact, it is possible that AMS-02 is observing an antiproton excess [881] at a concordant energy range [882–885], though the existence of this excess is not as well established [725, 726]. While promising, it has also been suggested that the GCE signal could be generated by millisecond pulsars [886, 887]. In recent years, the debate surrounding the origin of the GCE has intensified [888–903]. New measurements in the coming decade and a better theoretical understanding of Galactic diffuse emission models will help settle this debate, but until then, the origin of the GCE remains unknown and dark matter

annihilation remains a viable explanation.

As discussed in [879,880], the GCE can be well described by dark matter annihilations, particularly to $b\bar{b}$. This has fostered the development of many dark matter models with WIMP-like annihilation mechanisms, which are too numerous to review here (see [502] for a review). Of these, models with pseudoscalar s-channel mediators are particularly well-motivated because they are neutral and can evade direct detection constraints. In particular, if the dark matter lives close to resonance, the annihilation cross section can be boosted enough to explain the GCE [904–911]. While much of the previous work relies on the introduction of a new pseudoscalar mediator, the authors of [912] proposed an interesting alternative. In their setup, the dark sector is connected to the visible sector via a CP-violating coupling to the Higgs, which allows annihilation and spin-independent scattering to be governed by different parameters. In principle, the CP-violating coupling can generate a viable thermal relic candidate even away from the resonance, by suppressing the scattering rather than enhancing the annihilation. However, in [912], the authors consider specific model realizations within the context of supersymmetry where the benchmark best fit model still has the dark matter mass very close to half the Higgs mass.

In this work, we extract the key ingredients of their model, namely a Majorana dark matter candidate with CP-violating coupling to the Higgs, and explore the extent of freedom away from the mass resonance that can be achieved with larger CP-violating couplings. We see that for large enough coupling in the dark matter EFT, there is $\mathcal{O}(10)$ GeV flexibility for the dark matter mass when the phase is approximately $\pi/2$.

We also consider and explore the phenomenology of two different minimal UV realizations of this scenario: singlet-doublet dark matter [913–923] and doublet-triplet dark matter [918, 920, 923, 924]. We study both how these models translate to EFT parameters, and constraints governing these UV realizations, including contributions to the electron electric dipole moment (EDM), the Peskin-Takeuchi parameters, as well as possible collider signatures. We find that while the dark matter mass and CP-violating phase are independent parameters in the EFT, their dependence in the UV completion is quite nonlinear since the Yukawa coupling directly affects the dark matter mass. Specifically, it is difficult to achieve the phase tuning scenario without also tuning the mass in the UV completion, because the large couplings that are required to generate the annihilation cross section when away from resonance also change the dark matter mass. Additionally, we find that the amount of CP-violation in the UV may not be reflective of that observed in the EFT. In the singlet-doublet case, we find two different types of viable parameter space. When the UV couplings are small, both the singlet mass in the UV and the dark matter mass must be very close to $m_h/2$, but the phase is flexible. When the UV couplings are larger, parameters must be chosen such that both the phase of the dark matter-Higgs coupling and the dark matter mass must be somewhat tuned, but there is more flexibility in the dark matter and singlet masses than in the small coupling case. In the doublet-triplet model, we find

that EDM, spin-independent direct detection, and charged fermion collider search constraints are sufficient to rule out any WIMP-scale annihilation signal.

The rest of this chapter is organized as follows. In §5.2, we discuss the effective field theory of Majorana dark matter interacting with the Standard Model through a CP-violating Higgs coupling. The EFT parameters dictate the annihilation and scattering cross sections which are broadly applicable independent of specific UV completions. In §5.3, we UV complete the EFT by introducing a singlet Majorana fermion and a doublet Dirac fermion. In §5.4, we consider another UV completion by introducing a doublet Dirac fermion and a triplet Majorana fermion. We discuss the strong constraints placed on each of these models by a variety of complementary experimental probes such as the electron EDM, precision electroweak parameters, and collider searches. Finally, we offer some remarks in §5.5.

5.2 Model Independent Constraints in the Effective Theory

In this section we take an effective field theory approach and focus on the phenomenology of a single species of Majorana dark matter which couples to the visible sector via a CP-violating Higgs portal. After spontaneous symmetry breaking (SSB), the corresponding terms in the Lagrangian are given by

$$\mathcal{L} \ni \frac{y_{h\chi}}{2\sqrt{2}} h \bar{\chi} P_L \chi + \frac{y_{h\chi}^*}{2\sqrt{2}} h \bar{\chi} P_R \chi + \frac{g_{Z\chi}}{2} Z_\mu \bar{\chi} \gamma^\mu \gamma^5 \chi \tag{5.1}$$

where the CP-violation manifests in the complex nature of dark matter-Higgs coupling $y_{h\chi}$. Furthermore, we have also allowed for a coupling $g_{Z\chi}$ to the Z boson.[1]

As in all WIMP-type solutions to the GCE, the burden of the model is to reconcile the $\mathcal{O}(1)$ pb annihilation cross section necessary to achieve both the observed gamma-ray excess and the dark matter relic density, with the $\mathcal{O}(10^{-10})$ pb bounds on spin-independent scattering with nucleons from direct detection experiments. Traditionally, this is achieved for Higgs-portal dark matter by tuning the dark matter mass to the s-channel resonance $2m_\chi \sim m_h$, but an additional avenue is available in the case of our model.

In the non-relativistic limit, two Majorana fermions form a CP-odd state, so annihilation into the CP-even Higgs through a CP-conserving coupling is p-wave suppressed. It then follows that if the coupling is complex, the annihilation in this limit is dominantly set by the imaginary part of $y_{h\chi}$, which is reflected in the result we obtain in (5.4). Conversely, the dark matter scattering off of the nucleon (or quark) does not require any CP-violation since the initial and final states have the same CP properties, and thus we expect the spin-independent scattering cross section to be proportional to the real part of $y_{h\chi}$. This is reflected in the result we obtain in (5.13). Therefore, the phase of the Higgs coupling can also contribute to a large

[1] χ does not have a vector current coupling because $\bar{\chi}\gamma^\mu\chi$ vanishes identically for Majorana fermions.

hierarchy between the scattering and annihilation cross sections. With this intuition, we describe the details and corresponding phenomenology of this theory in the remainder of this section.

5.2.1 Annihilation

Annihilation is mediated by both the Higgs and the Z boson through an s-channel diagram. The dark sector couplings contributing to dark matter annihilation into SM fermions are given in (5.1), and the visible sector couplings have the form

$$\mathcal{L} \ni \sum_{f} \frac{y_{hf}}{\sqrt{2}} h \bar{f} f + g_{Zf} Z_\mu \bar{f} \gamma^\mu (v_f - a_f \gamma^5) f. \tag{5.2}$$

The couplings are given by their SM values

$$y_{hf} = -\frac{\sqrt{2} m_f}{v}, \quad g_{Zf} = \frac{e}{2 \cos\theta_w \sin\theta_w}, \quad v_f = I_3 - 2Q \sin^2\theta_w, \quad a_f = I_3, \tag{5.3}$$

where v is the Higgs vev, θ_w is the Weinberg angle, and m_f, I_3, and Q are the mass, weak isospin, and electric charge of the fermion respectively. In the non-relativistic limit, the total spin-averaged amplitude squared for annihilation can be written as

$$|\mathcal{M}|^2_{\chi\chi\to f\bar{f}} = 4m_\chi^2 \left[\frac{m_f^2}{m_Z^4} g_{Zf}^2 g_{Z\chi}^2 a_f^2 + y_{hf}^2 \mathrm{Im}[y_{h\chi}]^2 \frac{(m_\chi^2 - m_f^2)}{(m_h^2 - 4m_\chi^2)^2 + m_h^2 \Gamma_h^2} \right], \tag{5.4}$$

where Γ_h denotes the width of the Higgs. The Higgs mediated piece depends only on the imaginary part of the coupling as expected. The cross section is correspondingly given by

$$\langle \sigma v \rangle = \sum_{m_f \le m_\chi} \frac{N_c \sqrt{m_\chi^2 - m_f^2}}{64\pi m_\chi^3} |\mathcal{M}|^2_{\chi\chi\to f\bar{f}}. \tag{5.5}$$

If the dark matter is a thermal relic, then the present-day dark matter abundance, $\Omega_\chi h^2 = 0.11$, sets the annihilation cross section at the time of freeze-out, which is the well-known $\mathcal{O}(1)$ pb weak-scale cross section [517,925–935]. Recent work [936,937] has shown that for models with a hierarchy between annihilation and scattering strengths, early kinetic decoupling before freeze-out alters this number, requiring a larger cross section to achieve the observed abundance. At most extreme, a ~ 20 pb annihilation cross section may be needed for a ~ 57 GeV dark matter with purely imaginary couplings, though this is quite sensitive to the details of the QCD phase transition. However, this effect is significantly weaker for masses $\gtrsim m_h/2$, so we do not take our annihilation cross section to be this large.

At present, dark matter annihilation is expected to produce a distribution of gamma-rays whose flux is given by

$$\frac{d^2\Phi_\gamma}{d\Omega dE_\gamma} = \frac{1}{2}\langle\sigma v\rangle \left(\sum_f \frac{dN_\gamma}{dE\gamma} Br_{\chi\chi\to f\bar{f}}\right) \int_{los} \frac{\rho_\chi^2(r) d\ell}{4\pi m_\chi^2}, \tag{5.6}$$

where $Br_{\chi\chi\to f\bar{f}}$ denotes the branching ratio to the $f\bar{f}$ final state, and dN_γ/dE_γ its corresponding injection spectrum. $\rho_\chi(r)$ denotes the dark matter halo profile and is integrated over the line-of-sight to the Galactic Center. It has been shown that the Fermi GCE data is well-modeled by a Higgs portal dark matter with a cross section $\langle\sigma v\rangle \sim 3$ pb, assuming a modified NFW profile [875]. As the precise best fit depends on many details, including the galactic profile and background modeling [938], in conjunction with the modeling uncertainties of the thermal relic argument, we will consider here a range of cross sections $\langle\sigma v\rangle$ from 1 to 10 pb to be in concordance with both the GCE and the relic abundance.

5.2.2 Direct Detection

In contrast with annihilation, processes relevant for direct detection occur below the weak scale and should be considered in terms of effective interactions with target nuclei. Much of the subsequent discussion follows [488]. At momentum transfers $t \ll m_Z^2$, the interactions in (5.1) and (5.2) are rewritten as the following dimension-6 operators

$$\mathcal{L} \ni \frac{C_S}{m_h^2}\bar{\chi}\chi\bar{f}f + \frac{C_{PS}}{m_h^2}\bar{\chi}\gamma^5\chi\bar{f}f + \frac{C_V}{m_Z^2}\bar{\chi}\gamma^\mu\gamma^5\chi\bar{f}\gamma_\mu f + \frac{C_{PV}}{m_Z^2}\bar{\chi}\gamma^\mu\gamma^5\chi\bar{f}\gamma_\mu\gamma^5 f \tag{5.7}$$

with C_S, C_{PS}, C_V, and C_{PV} denoting the scalar, pseudo-scalar, vector, and pseudo-vector pieces of the quark-gauge couplings respectively. The contributions governed by C_{PS} and C_V are velocity-suppressed and we neglect them in the following. After matching to the UV theory, the coefficients are given by

$$C_S = \frac{1}{2}Re[y_{h\chi}]y_{hf} \qquad C_{PV} = g_{Z\chi}g_{Zf}a_f. \tag{5.8}$$

In the zero momentum transfer limit, the nucleon-level operators are matched to the quark-level ones via form factors

$$\langle N(p)|\bar{f}\gamma^\mu\gamma^5 f|N(p')\rangle = \bar{u}_N(p)\left[\Delta_1^{f,N}(q^2)\gamma^\mu\gamma^5\right]u_N(p') \tag{5.9}$$

$$\langle N(p)|\bar{f}f|N(p')\rangle = \frac{m_N}{m_f}f_f^N\bar{u}_N(p)u_N(p') \tag{5.10}$$

where N represents a nucleon (a proton or neutron), $q = p' - p$ denotes the momentum transfer, and the form factors are listed in Table 5.1. We have neglected higher order terms in q^2. For the scalar term specifically, the heavy quarks also contribute via a gluon loop. After integrating out heavy quarks, the relevant operator

	$\Delta_1^{N,u}$	$\Delta_1^{N,d}$	$\Delta_1^{N,s}$	f_u^N	f_d^N	f_s^N	f_g^N
Protons	0.80	-0.46	-0.12	0.018	0.027	0.037	0.917
Neutrons	-0.46	0.80	-0.12	0.013	0.040	0.037	0.910

Table 5.1: Here we show the light quark and gluon form factors for the proton and neutron. These values come from [939–942] and are summarized in [488].

for each flavor appears as

$$-\frac{C_S}{m_h^2}\frac{\alpha_s}{12\pi m_f}\bar{\chi}\chi G^{\mu\nu}G_{\mu\nu}. \tag{5.11}$$

To match to the nucleon-level picture the following matrix element is taken into account

$$\langle N(p)|G^{\mu\nu}G_{\mu\nu}|N(p')\rangle = -\frac{8\pi}{9\alpha_s}m_N f_g^N \bar{u}_N(p)u_N(p'). \tag{5.12}$$

In terms of the quark-level couplings, the nucleon-level spin-independent cross section is given by

$$\sigma_{SI} = \frac{m_N^2 m_\chi^2}{4\pi(m_\chi + m_N)^2}\left[\frac{Re[y_{h\chi}]}{m_h^2}\left[\sum_{f\in u,d,s} y_{hf}\frac{m_N}{m_f}f_f^N + \sum_{f\in c,b,t} y_{hf}\frac{2}{27}\frac{m_N}{m_f}f_g^N\right]\right]^2. \tag{5.13}$$

As discussed earlier, the cross section only depends on the real part of the Higgs coupling. Furthermore, the dependence on the coupling to the Z boson vanishes in the $q \to 0$ limit. Likewise the spin-dependent cross section is given by

$$\sigma_{SD} = \frac{3m_N^2 m_\chi^2}{\pi(m_\chi + m_N)^2}\left[\frac{g_{Z\chi}}{4m_Z^2}\sum_{f\in u,d,s} g_{Zf}a_f\Delta_1^{N,f}\right]^2. \tag{5.14}$$

5.2.3 EFT Results

In this subsection we examine the phenomenology of the effective theory, and discuss the regions of parameter space where a high annihilation and low scattering cross section can be achieved – specifically we are interested in an annihilation cross section between approximately 1 and 10 pb to fit the GCE and a scattering cross section consistent with direct detection experiments. For spin-independent scattering, the strongest limits come from XENON1T [531, 532], while for spin-dependent scattering, the strongest limits come from both XENON1T [533] and PICO [535, 536]. LZ [943] and XENONnT [944] are projected to improve on current limits within the parameter space of interest. The projected limits are comparable, so we show only one in our figures for clarity. We omit limits from IceCube [945], LUX [946, 947] and PandaX-II [948] because they are slightly weaker than those we've shown for $\mathcal{O}(60)$ GeV dark matter. For the spin-independent constraints, we

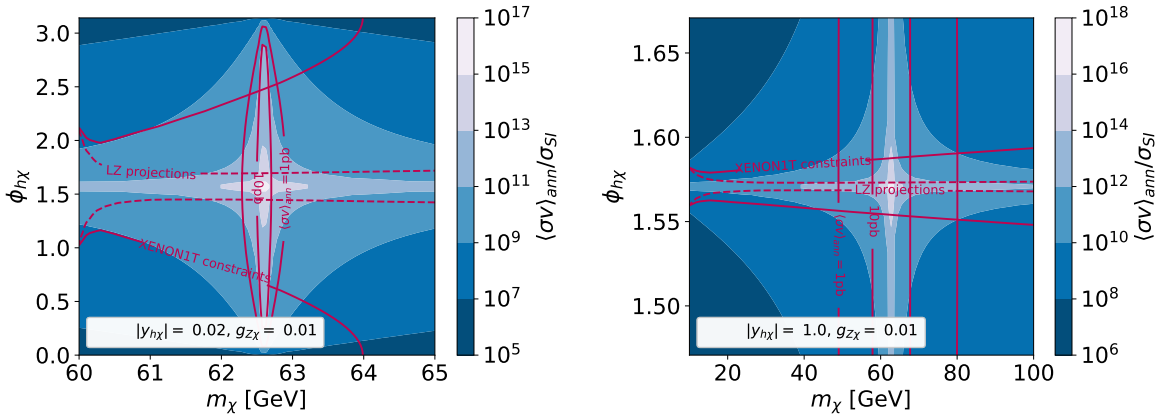


Figure 5.1: The ratio between annihilation and spin-independent direct detection cross sections on the $m_\chi - \varphi_{h\chi}$ plane for different values of $|y_{h\chi}|$. The region allowed by direct detection is inside the solid XENON1T [531, 532] constraint line, while the region allowed by annihilation is between the solid 1 pb and 10 pb lines. We also show projected limits from LZ [943] as dashed lines. Note that the axis scales on the two plots are different. The left plot shows the mass resonance with small $y_{h\chi}$. The right plot shows the phase tuning, with a large coupling. Both plots include a small non-zero Z coupling; the limits are similar for vanishing Z coupling.

consider only dark matter-proton scattering because in this case the difference between proton and neutron cross sections is negligible.

First we review which masses and coupling magnitudes are in general concordance with scattering constraints and annihilation requirements. Typical couplings that can generate an annihilation cross section of ~ 1 pb are shown in (5.15) for two different dark matter masses.

$$\begin{aligned}\left[\frac{\langle\sigma v\rangle}{1\text{ pb}}\right] &= \left[\frac{m_\chi}{80\text{ GeV}}\right]^2 \left[\frac{4m_\chi^2 - m_h^2}{10^4\text{ GeV}^2}\right]^{-2} \left[\frac{y_{h\chi}\sin\varphi_{h\chi}}{1.0}\right]^2 \\ &= \left[\frac{m_\chi}{62.5\text{ GeV}}\right]^2 \left[\frac{4m_\chi^2 - m_h^2}{50\text{ GeV}^2}\right]^{-2} \left[\frac{y_{h\chi}\sin\varphi_{h\chi}}{0.007}\right]^2\end{aligned} \tag{5.15}$$

In (5.16), we show approximate couplings and masses that are consistent with direct detection constraints.

$$\begin{aligned}\left[\frac{\sigma_{SI}}{10^{-10}\text{ pb}}\right] &= \left[\frac{y_{h\chi}\cos\varphi_{h\chi}}{0.02}\right]^2 \\ \left[\frac{\sigma_{SD}}{10^{-5}\text{ pb}}\right] &= \left[\frac{g_{Z\chi}}{0.01}\right]^2\end{aligned} \tag{5.16}$$

We remind the reader that the free parameters of the theory are m_χ, $g_{Z\chi}$, and the complex coupling $y_{h\chi}$ with phase $\varphi_{h\chi}$. While $g_{Z\chi}$ and $\text{Im}[y_{h\chi}]$ set the annihilation cross section, only $\text{Re}[y_{h\chi}]$ sets the magnitude of scattering. In order to generate a large enough annihilation cross section while avoiding direct detection constraints, Higgs portal dark matter models typically tune the dark matter mass close to half the Higgs

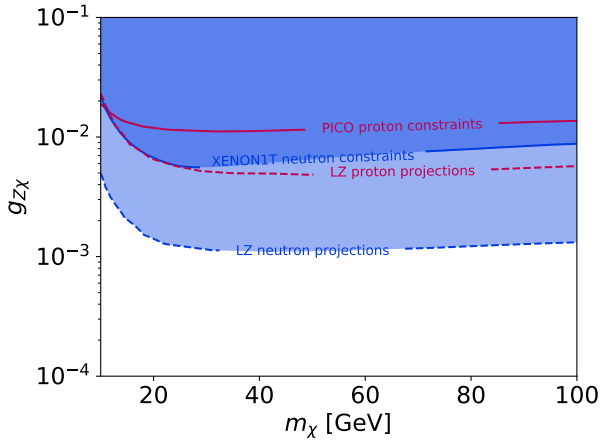


Figure 5.2: Spin-dependent direct detection limits as a function of dark matter mass and dark matter-Z coupling. Constraints are close to horizontal because the spin-dependent cross section depends on the reduced mass. We show the strongest model independent constraints: XENON1T [533] for neutrons and PICO [535, 536] for protons. Additionally, we show projected limits from LZ [943].

mass [904–910, 912, 949]. While tuning the mass is one way to generate the correct ratio in this model, we emphasize that in the EFT, the correct ratio can also be obtained for a wider mass range by increasing the magnitude of the Higgs coupling while tuning the phase, $\varphi_{h\chi}$, of the Higgs coupling close to $\pi/2$ to suppress direct detection constraints. This is illustrated in Fig. 5.1, which plots annihilation and spin independent direct detection constraints in the $m_\chi - \varphi_{h\chi}$ plane for different magnitudes of the Higgs couplings, assuming $m_h = 125.2$ GeV. We can see that near the mass resonance, mass must be tuned to within less than a GeV of the pole. In this regime, a small Higgs coupling (~ 0.02) is sufficient to generate the annihilation cross section and the phase does not need to be near $\pi/2$ to avoid direct detection constraints. However, away from resonance, the larger Higgs coupling required to generate the correct annihilation cross section is allowed with phase tuning, because direct detection only constrains the real part of $y_{h\chi}$. This widens the mass range considerably to $\mathcal{O}(10)$ GeV. Even for the mass resonance, the coupling cannot be purely real, because the leading velocity dependent term is not large enough to generate the required annihilation cross section given the finite Higgs width. See Appendix B for more details. Note that while in principle a large pseudo-vector Z coupling could also generate a sufficient annihilation cross section, this is constrained by spin-dependent direct detection constraints, as shown in Fig. 5.2. Within the range of Z couplings allowed by direct detection, the effect on the allowed annihilation signal is negligible.

5.3 Singlet-Doublet Model

A well-motivated way to UV complete the dark matter EFT provided in §5.2 in a gauge invariant manner is to introduce additional particles charged under G_{SM}. In this section, we discuss a simple potential UV

completion, where the only additional particles we introduce to the Standard Model are a singlet Majorana fermion and a doublet Dirac fermion. This model has previously been discussed in other contexts in [913–923].

5.3.1 Model in the UV

We start by establishing notation and describing the model. The model contains a singlet Majorana fermion ψ_1 and an additional SU(2) doublet Dirac fermion with hypercharge 1/2. We describe the SU(2) doublet with two left handed Weyl fermions ψ_2 (with neutral component ψ_2^0 and charged component ψ_2^1) and $\tilde{\psi}_2$ (with neutral component $\tilde{\psi}_2^0$ and charged component $\tilde{\psi}_2^{-1}$). All new fermions are SU(3) singlets. The Lagrangian for this model is

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \mathcal{L}_{\rm kinetic} - m_2 \psi_2 \cdot \tilde{\psi}_2 - \frac{m_1}{2}\psi_1\psi_1 + Y \psi_1 H^\dagger \psi_2 - \tilde{Y}\psi_1 H \cdot \tilde{\psi}_2 + \text{h.c.} \tag{5.17}$$

As we introduce three new fields and four free parameters, there is one remaining physical phase. We make the choice to fix each of the Yukawa terms to the same phase, which carries the CP-violation,

$$Y \equiv y\, e^{i\delta_{CP}/2}, \qquad \tilde{Y} \equiv \tilde{y}\, e^{i\delta_{CP}/2}. \tag{5.18}$$

After SSB, the mass terms are written as

$$\mathcal{L}_{\rm mass} = -m_2 \left(\tilde{\psi}_2^{-1}\psi_2^1 - \tilde{\psi}_2^0 \psi_2^0 \right) - \frac{m_1}{2}\psi_1\psi_1 + \frac{v}{2} y\, e^{i\delta_{CP}/2}\psi_1\psi_2^0 + \frac{v}{2}\tilde{y}\, e^{i\delta_{CP}/2}\psi_1\tilde{\psi}_2^0 + \text{h.c.} \tag{5.19}$$

Let us define $\psi_2^s \equiv \frac{1}{\sqrt{2}}(\psi_2^0 + \tilde{\psi}_2^0)$ and $\psi_2^d \equiv \frac{1}{\sqrt{2}}(\psi_2^0 - \tilde{\psi}_2^0)$ to be the two Majorana fermions that constitute the neutral Dirac fermion $\{\psi_2^0, \tilde{\psi}_2^0\}$. The mass eigenstates thus result from the mixing of the doublet and singlet Majorana fermions, $\psi_i = (\psi_2^s, \psi_2^d, \psi_1)_i$. We will denote the mass eigenstates $\chi_i = (\chi, \chi_1, \chi_2)_i$, the lightest of which, χ, is the dark matter candidate. Then

$$\mathcal{L}_{\rm mass} = -m_2 \tilde{\psi}_2^{-1}\psi_2^1 - \frac{1}{2}\psi_i M_{ij}\psi_j, \tag{5.20}$$

where M is the mass matrix. This basis change is governed by J, the matrix of eigenvectors that diagonalizes both $M^\dagger M$ and M, phase rotated such that $J^T M J$ has real eigenvalues. After diagonalizing, the Higgs Yukawa couplings are

$$\mathcal{L}_{\rm Higgs} = \frac{1}{2} h \chi_i [J^T U_h J]_{ij} \chi_j + \text{h.c.} \tag{5.21}$$

where

$$U_h = \begin{pmatrix} 0 & 0 & \frac{(Y+\tilde{Y})}{2} \\ 0 & 0 & \frac{(Y-\tilde{Y})}{2} \\ \frac{(Y+\tilde{Y})}{2} & \frac{(Y-\tilde{Y})}{2} & 0 \end{pmatrix}. \tag{5.22}$$

Since one of the new fermions is an SU(2) doublet, the new fermions also couple to the electroweak gauge bosons. The Z couplings are

$$\mathcal{L}_{Z} = \chi_i^\dagger [J^\dagger U_Z J]_{ij} Z \overline{\sigma} \chi_j + g_Z (\cos^2 \theta_W - \sin^2 \theta_W)(\psi_2^{1\dagger} Z \overline{\sigma} \psi_2^1 - \tilde{\psi}_2^{-1\dagger} Z \overline{\sigma} \tilde{\psi}_2^{-1}) \tag{5.23}$$

while the W couplings are

$$\mathcal{L}_{W} = \chi_i^\dagger [bJ^*]_i W^+ \overline{\sigma} \tilde{\psi}_2^{-1} + \psi_2^{1\dagger} W^+ \overline{\sigma} [aJ]_j \psi_j + \text{h.c.} \tag{5.24}$$

where g is the SU(2) gauge coupling, g' is the U(1) hypercharge gauge coupling, and $g_Z \equiv \sqrt{g^2 + g'^2}/2$. Here, $a_i = (g/2, g/2, 0)_i$, $b_i = (g/2, -g/2, 0)_i$ and

$$U_Z = \begin{pmatrix} 0 & -g_Z & 0 \\ -g_Z & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.25}$$

The dark matter candidate χ obtains the couplings seen in the EFT via mixing between the singlet and doublet. The strength of these couplings can be adjusted by altering the makeup of the lightest Majorana fermion. The theory at this level is fully specified by five degrees of freedom: the singlet mass m_1, the doublet mass m_2, the doublet Yukawa coupling magnitudes $\{y, \tilde{y}\}$ and the associated CP-violating phase δ_{CP}.

5.3.2 Translating to the EFT

Now we discuss how the EFT parameters $g_{Z\chi}, m_\chi$, and $y_{h\chi}$ depend on the UV parameters $y, \tilde{y}, m_1, m_2$, and δ_{CP}. We focus mostly on the region where m_2 is large, but also comment on the more general case.[2] Since the theory has a charged fermion with mass m_2, parameter space with small m_2 will generically be ruled out by collider constraints [950, 951]. EDM and electroweak constraints are likewise more stringent in this regime.

Fig. 5.3 shows the EFT mass and phase as a function of m_1 and m_2 for different values of the UV coupling magnitudes and phase. On the left we show multiple values of $y\tilde{y}$ for fixed δ_{CP} while on the right we show multiple values of δ_{CP} for fixed $y\tilde{y}$. In both cases, we can see that only a narrow range in m_1 translates to dark matter with mass near the mass resonance. When m_2 is large, the lightest fermion is mostly m_1. In this limit, mixing is small, so to have the dark matter mass near the mass resonance, m_1 must be fairly

[2] We also omit the case where both m_1 and m_2 are large. In this case, extremely large couplings are required in order to get dark matter with mass near $m_h/2$. This means the δ_{CP} must be small to avoid EDM constraints, which leaves us with $\varphi_{h\chi}$ mostly real and prevents us from simultaneously evading spin-independent constraints.

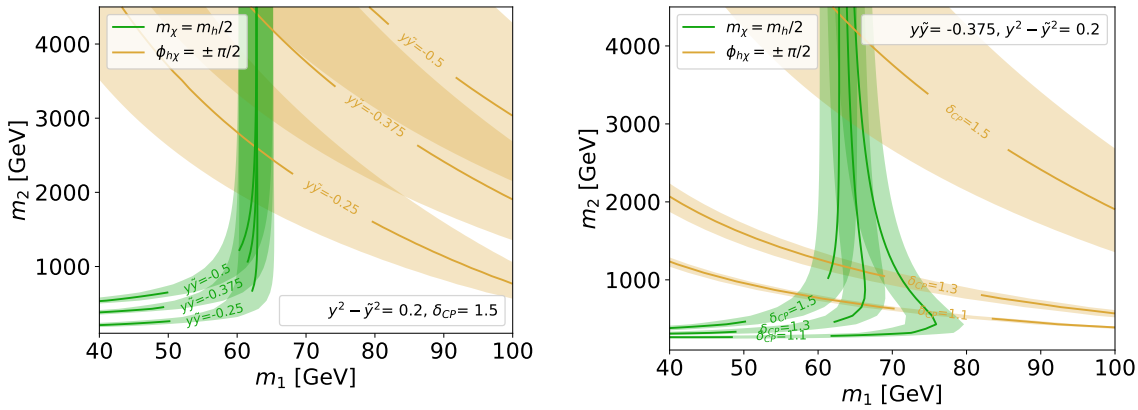


Figure 5.3: EFT coupling phase and dark matter mass as a function of m_1 and m_2 for different values of $y, \tilde{y}$, and δ_{CP}. Left: δ_{CP} is fixed to 1.5 while $y\tilde{y}$ is varied. Right: $y\tilde{y}$ is fixed to -0.375 while δ_{CP} is varied. The shaded regions give a sense of the width of the regions of interest: $60\,\text{GeV} \leq m_\chi \leq 65\,\text{GeV}$ and $1.55 \leq \varphi_{h\chi} \leq 1.60$.

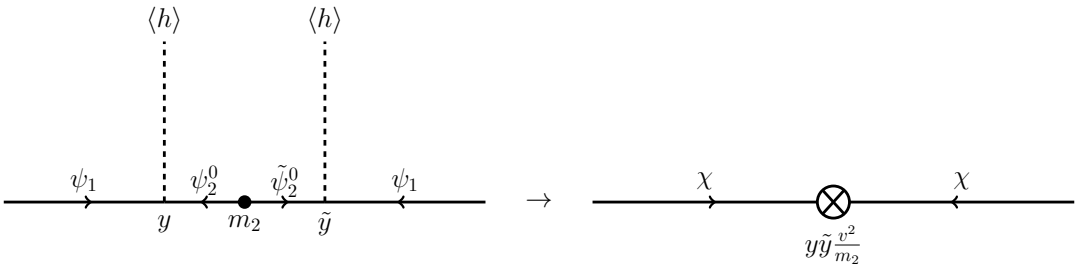


Figure 5.4: Diagram generating dark matter mass in the limit where m_2 is large.

close to half the Higgs mass. We can see that changing $y\tilde{y}$ changes where $\varphi_{h\chi} = \pi/2$ is located but only has a minimal effect on which m_1 value translates to the mass resonance. We can also see that for the same m_1, smaller $y\tilde{y}$ requires a correspondingly smaller m_2 to get dark matter with $m_\chi \approx m_h/2$. Changing δ_{CP} also changes the location of $\varphi_{h\chi} = \pi/2$ contour, but additionally affects the m_1 required to get the mass resonance and the width of the $\varphi_{h\chi} \approx \pi/2$ band.

Fig. 5.4 shows that the corrections to the mass scale as $y\tilde{y}v^2/m_2$.[3] This diagram also tells us that $\varphi_{h\chi} = \delta_{CP}$ in the large m_2 limit, as long as mixing is small and the dark matter mass comes mostly from m_1 rather than the Higgs vev. This can also be seen in Fig. 5.5. When the dark matter mass gets a large contribution from the Higgs vev the story is more complicated: when $y\tilde{y}$ and m_1 have opposite signs, the Higgs contribution can cancel with m_1 at $y\tilde{y} = -m_2 m_1/v^2$ to get a massless state. There is a mass resonance contour for $y\tilde{y}$ both larger and smaller than this value, which can be seen in Fig. 5.6. We might also ask whether a small δ_{CP} in the UV can translate to $\varphi_{h\chi} \approx \pi/2$ in the IR and produce an annihilation signal that evades both direct detection and EDM constraints. However, from the same figure, we can see that

[3] Although we need to phase rotate ψ_2^s, the phase rotations in the couplings and mass insertion cancel out.

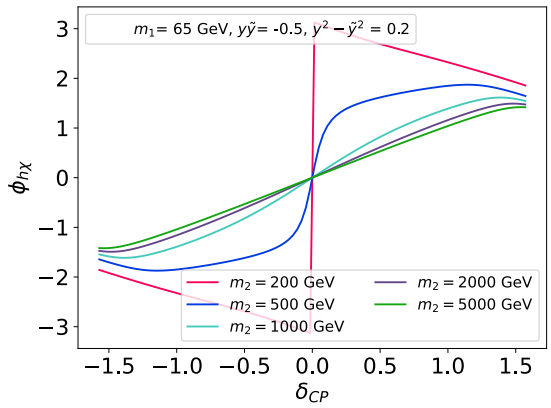


Figure 5.5: $\varphi_{h\chi}$ as a function of δ_{CP} for different values of m_2. As m_2 increases, the IR phase maps directly to the UV phase and $\varphi_{h\chi} \sim \delta_{CP}$.

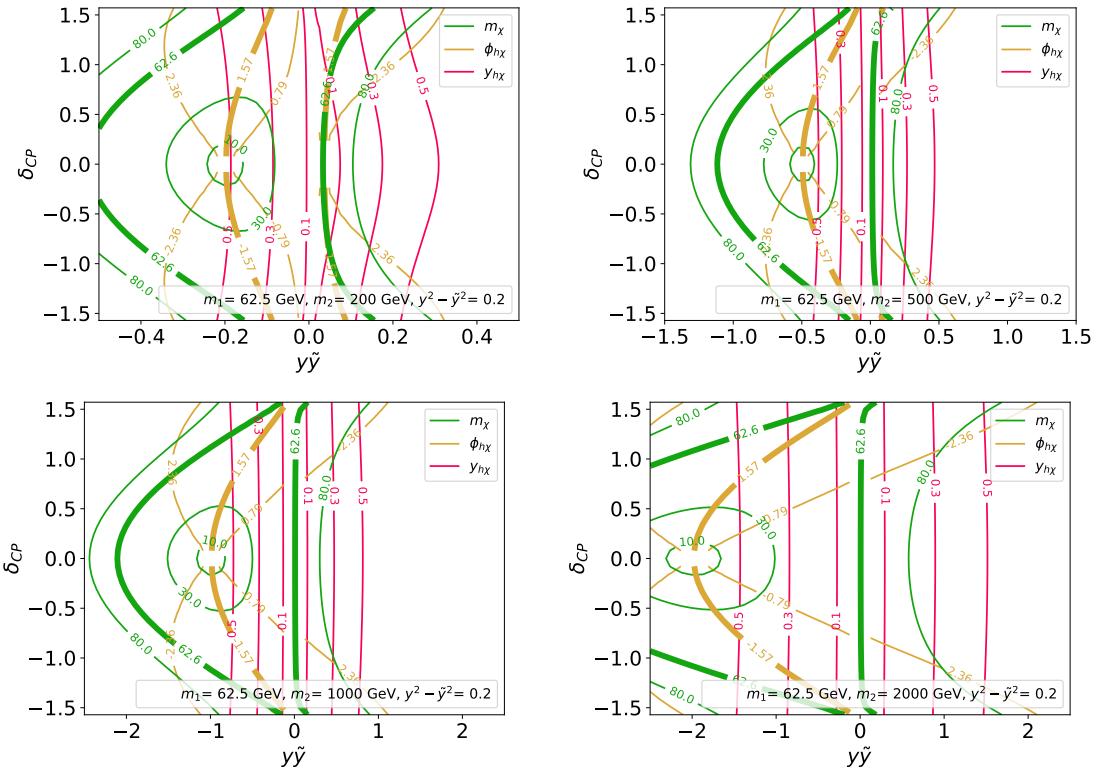


Figure 5.6: Dark matter mass, EFT phase, and dark matter-Higgs Yukawa coupling as a function of the UV parameters $y\tilde{y}$ and δ_{CP} for different values of m_2. In each plot we see a similar mass structure: we see a massless state when $y\tilde{y}$ and m_1 have opposite signs, and have a lightest fermion near 60 GeV for both larger and smaller $y\tilde{y}$ than this value. Note the different values on the $y\tilde{y}$ axis in each of the plots.

although there is a point where small δ_{CP} translates to $\varphi_{h\chi} \approx \pi/2$, it corresponds precisely to the massless state mentioned above and cannot generate our annihilation signal. This occurs because when m_χ is zero, we can freely rotate m_1 to absorb any value of the unphysical phase in y, which is evidenced by all phase contours converging at this massless point.

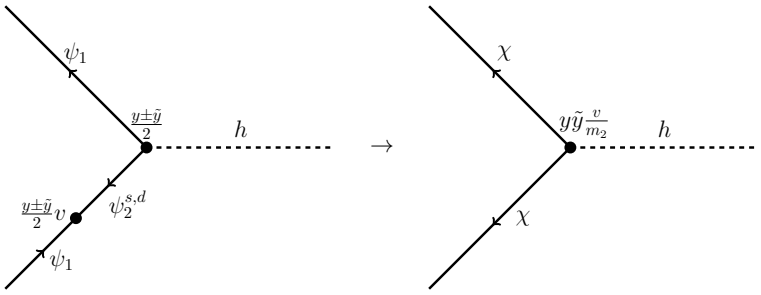


Figure 5.7: Diagram that generates the dark matter-Higgs coupling in the limit where m_2 is large.

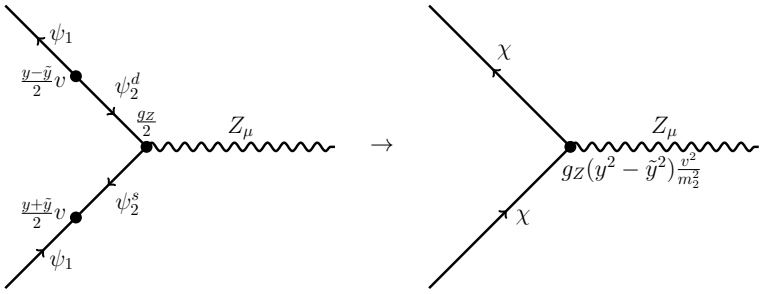


Figure 5.8: Diagram that generates the dark matter-Z coupling in the limit where m_2 is large.

In the small mixing and large m_2 limit, there are two contributions to the Higgs coupling: one where ψ_1 mixes into ψ_2^s and one where it mixes into ψ_2^d, as shown in Fig. 5.7. Each of these contributes $(y \pm \tilde{y})^2 v/m_2$, with a relative minus sign between the two contributions because we need to phase rotate ψ_2^s to have positive mass. This means the Higgs coupling scales as $y\tilde{y}v/m_2$, which determines the scaling of the annihilation signal. This can also be seen from the pink lines in Fig. 5.6. Note that this scaling breaks down once the Yukawa contributions become the dominant contribution to the mass.

In the same limit, the dominant contribution to the Z coupling comes from Fig. 5.8, which scales as $g_Z(y^2 - \tilde{y}^2)v^2/m_2^2$. Even away from this limit, we still get a vanishing Z coupling for $y = \tilde{y}$, because only one of the doublet states mixes with the singlet when $y = \tilde{y}$. For small m_2, spin-dependent direct detection constraints require $y \approx \tilde{y}$, but for $m_2 \gtrsim 500$ GeV this constraint becomes irrelevant, since the Higgs coupling (which determines the annihilation signal) scales as m_2^{-1} while the Z coupling scales as m_2^{-2}. This can be seen in Fig. 5.9.

5.3.3 Constraints

In this section, we discuss the experimental constraints that apply to the singlet-doublet model. We focus on constraints that apply directly to the parameters in the UV theory, including discussing their scaling in the large m_2 limit.

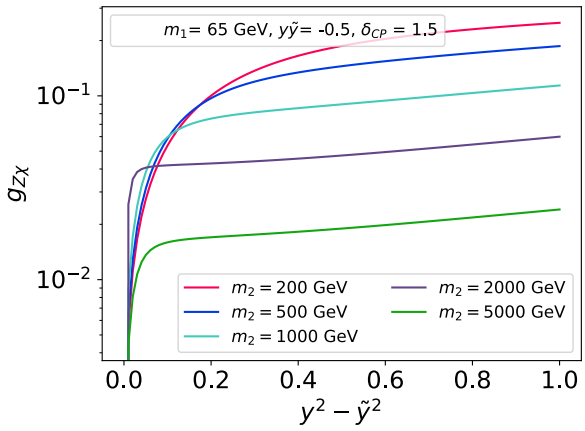


Figure 5.9: Plot of $g_{Z\chi}$ as a function of $y^2 - \tilde{y}^2$ for different values of m_2. $g_{Z\chi}$ increases with increasing $y^2 - \tilde{y}^2$ and decreases with increasing m_2, corroborating the scaling derived from the diagrams in Fig. 5.8.

Electric Dipole Moment

Any new source of CP-violation in a given model can lead to additional contributions to electric dipole moments. Since our model contains new CP-violating couplings to the Higgs, we expect electron EDM constraints to be relevant for our model. For small m_2, the EDM limit will be one of the strongest on our model, since the EDM is precisely constrained to be below 1.1×10^{-29} e cm [952].

For the singlet-doublet model above, the only relevant diagram is the Barr-Zee diagram with W bosons in the outer loop [953], displayed in Fig. 5.10. There are no other Barr-Zee diagrams with Higgs or Z legs; since CP-violation is only in the neutral sector of this model and a charged particle is necessary to radiate a photon, the inner loop must contain both a neutral and charged particle. Additionally, there are no other non-Barr-Zee diagrams that contribute to the EDM at 2 or fewer loops. For any non-Barr-Zee diagrams to contribute, there would have to be a CP-odd correction to a gauge boson or Higgs propagator. With only a single external momentum, it is impossible to contract with an epsilon tensor and make a non-vanishing CP-odd Lorentz invariant.

To compute the value of the relevant Barr-Zee diagram, we use a simplified version of equation 21 in [954], where we have neglected the neutrino mass, approximated lepton couplings as flavor diagonal, and used the fact that one of the fermions in the loop is neutral:

$$\frac{d_e}{e} = -\frac{g^2}{(4\pi)^4} \sum_i \mathrm{Im}([aJ]_i^*[bJ^*]_i) \left(\frac{m_c m_{n,i} m_e}{M_W^4}\right) G(x_c, x_i, 0). \tag{5.26}$$

Here, $x_\alpha = m_\alpha^2/M_W^2$ and $G(a, b, c)$ is defined as

$$G(a, b, c) = \frac{1}{1-c} \int_0^1 \frac{dx}{1-x} \left(\frac{c}{z-c} \log\left(\frac{c}{z}\right) + \frac{1}{1-z} \log\left(\frac{1}{z}\right)\right) \tag{5.27}$$

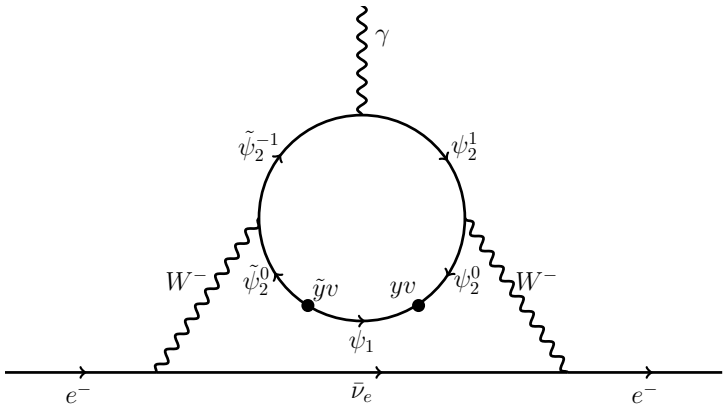


Figure 5.10: Leading non-zero Barr-Zee diagram contribution in the large m_2 limit. In this limit, we can work perturbatively in the gauge basis. The relevant W couplings are the coefficients of $\chi_{21}^\dagger \frac{\sqrt{g}}{2} W^- \chi_{20}$ and $\tilde{\chi}_{20}^\dagger \frac{\sqrt{g}}{2} W^+ \tilde{\chi}_{2-1}$.

with

$$z(x, a, b) = \frac{b}{x} + \frac{a}{1-x}. \tag{5.28}$$

Recall from §5.3.1 that couplings a_i and b_i parameterize the W boson couplings to the inner loop fermions in the gauge basis, which are given in (5.24), and J is the change of basis matrix.

When m_2 is large enough that we can integrate out the doublet and mixing is small, the dominant contribution to the EDM comes from Fig. 5.10, since each helicity of charged fermion couples to a different neutral doublet component and mixing with the singlet is necessary to generate CP-violation. This contribution scales as $y\tilde{y}v^2/m_2^2$. The m_2 scaling follows from dimensional analysis: three factors of m_2 from the integral measure cancel with three of the five factors of m_2 from the propagators.[4]

Electroweak parameters

Here we consider constraints from electroweak precision measurements, where deviations from the SM are parametrized by oblique parameters S, T, U, W, and Y [955, 956], defined as

$$S \equiv \frac{4c_W^2 s_W^2}{\alpha_e}\left[\Pi'_{ZZ}(0) - \frac{c_W^2 - s_W^2}{cs}\Pi'_{Z\gamma}(0) - \Pi'_{\gamma\gamma}(0)\right], \tag{5.29}$$

$$T \equiv \frac{1}{\alpha}\left[\frac{\Pi_{WW}(0)}{m_W^2} - \frac{\Pi_{ZZ}(0)}{m_Z^2}\right], \tag{5.30}$$

$$U \equiv \frac{4s_W^2}{\alpha}\left[\Pi'_{WW}(0) - \frac{c_W}{s_W}\Pi'_{Z\gamma}(0) - \Pi'_{\gamma\gamma}(0)\right] - S, \tag{5.31}$$

$$W \equiv \frac{m_W^2 s_W^2 c_W^2}{8\pi\alpha}\left[\Pi''_{ZZ}(0) + \frac{2s_W}{c_W}\Pi''_{Z\gamma}(0) + \frac{s_W^2}{c_W^2}\Pi''_{\gamma\gamma}(0)\right], \tag{5.32}$$

$$Y = \frac{m_W^2 s_W^2}{8\pi\alpha}\left[c_W^2\Pi''_{\gamma\gamma}(0) + s_W^2\Pi''_{ZZ}(0) - 2s_W c_W \Pi''_{\gamma Z}(0)\right]. \tag{5.33}$$

[4] The ψ_1 propagator also scales as m_2^{-1} since $p = m_2 \gg m_1$.

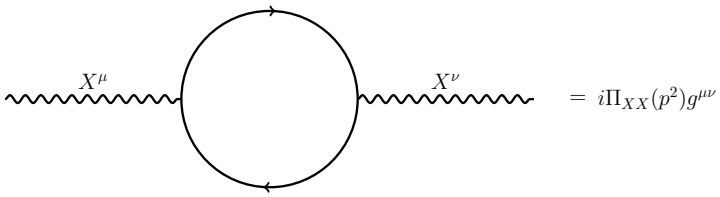


Figure 5.11: New particles that couple to the Standard Model gauge bosons contribute to the vacuum polarization at 1-loop through this diagram. The X^μ represents an electroweak gauge boson. We ignore the $p^\mu p^\nu$ terms since they aren't relevant for (5.29) - (5.33).

The masses and couplings are evaluated at m_Z^2 and c_W and s_W are $\cos\theta_W$ and $\sin\theta_W$ respectively. Π_{XX} represents the new particles' contribution to the vacuum polarization of the gauge boson X at 1-loop, computed in $\overline{\text{MS}}$ scheme under the convention shown in Fig. 5.11.

To lend intuition, we note that T parametrizes custodial SU(2) breaking inherent in the asymmetry within the doublet terms; in our theory this manifests in the difference in Yukawa couplings y and $\tilde{y}$. U is the derivative of T, and thus is typically smaller. All these parameters fall off with increasing m_2.

The most recent constraints, at 95% CL, from the LHC yield

$$S = -0.01 \pm 0.10 \qquad T = 0.03 \pm 0.12 \qquad U = -0.01 \pm 0.10 \tag{5.34}$$

with correlations +0.92 between S and T, -0.80 between S and U, and -0.93 between T and U [957]. The singlet-doublet model predominately contributes to the S and T parameters; see Appendix C for more details. W and Y are measured to be

$$W = (-2.7 \pm 2.0) \times 10^{-3} \qquad Y = (4.2 \pm 4.9) \times 10^{-3} \tag{5.35}$$

with correlation -0.96 [958], though we find these to be subdominantly constraining for this theory.

Collider Experiments

Constraints from many collider searches (in particular SUSY searches) can be applied to this model. Specifically, we consider those searches included in the database of the publicly available SModelS version 1.2.4 software [959–967]. To generate the necessary input, we use SARAH 4.14.3 [968–970] to create modified versions of SPheno [971, 972] and Madgraph [973, 974] which include the singlet and doublet. Then we use this version of SPheno at tree level to compute the spectrum and branching ratios for SModelS and the run card for Madgraph, which was used to obtain the production cross sections that SModelS also needs as input. These constraints are combined into a single exclusion limit labeled LHC when included in our plots. In addition to this constraint, we also show the constraint from invisible Higgs decay. We do not include the

constraint from invisible Z decay, since it is not kinematically allowed in the parameter space of interest.

5.3.4 Full Exclusion Limits and Discussion

Finally, combining all of these constraints, we examine the remaining parameter space for singlet-doublet dark matter that has the desired amount of annihilation. Our results are shown in Figs. 5.12 and 5.13. We find that in all cases, some tuning of the parameters is required, but that there is flexibility in which UV parameters we need to tune.

As in the EFT, in order to achieve a pure mass resonance (and not have to tune the EFT phase) we need small couplings. This can be seen in Figs. 5.12a and 5.12b. The spin-independent constraints are weak for small couplings, regardless of δ_{CP} or the EFT phase. Other constraints are even less restrictive, except for the EDM at very large δ_{CP}. Since the couplings are small, m_1 must be tuned near $m_h/2$ in order to achieve a sufficient annihilation signal, but there is flexibility in the value of m_2, as can be seen in Fig. 5.13a. This is the region of parameter space that is relevant for the best fit in [912].

If instead we choose our parameters so that we allow the EFT phase to be tuned near $\pi/2$, there is other viable parameter space with larger couplings. Here, we have slightly more flexibility in m_1 (which still needs to be roughly $60-70$ GeV), but m_2 must be large ($m_2 \gtrsim \mathcal{O}(1)$ TeV) to avoid EDM, electroweak, and collider constraints. This can be seen in Fig. 5.13b. Additionally, to achieve an EFT phase near $\pi/2$ and avoid spin-independent constraints, generally $\delta_{CP} \gtrsim 1$. Note that limits from spin-dependent scattering can be avoided, since they vanish when $y = \pm\tilde{y}$. This part of parameter space generally requires proximity to both the mass resonance and the phase $\pi/2$ line. However, there is still some flexibility in both values; masses $m_\chi > 65\,\text{GeV}$ and phases $\varphi_{h\chi} < 1.5$ are allowed in these intersections, albeit not simultaneously. Unlike in the case of the EFT, it is very difficult to tune only the phase because we cannot make couplings arbitrarily large without affecting the mass spectrum, as we saw in §5.3.2.

Figures 5.12b - 5.12d shows several examples of this. In Fig. 5.12b, we can see the case where we still choose m_1 to be near $m_h/2$ but allow larger couplings. If instead we choose m_1 further away from $m_h/2$, the only viable parameter space also requires large couplings in order to get the dark matter mass sufficiently close to resonance. This is shown in Figs. 5.12c and 5.12d. Comparing these two plots, we can see that there is more flexibility in δ_{CP} and larger required coupling values for higher m_2, because higher m_2 changes the shape of the EFT phase $\pi/2$ contour. Specifically, there is more overlap between $\varphi_{h\chi}$ near $\pi/2$ and the annihilation signal in the large m_2 case because the condition $\varphi_{h\chi} = \pi/2$ becomes less dependent on $y\tilde{y}$ at larger m_2.[5]

[5]This is because the $\varphi_{h\chi} = \pi/2$ contour always goes through the massless state that exists for negative $y\tilde{y}$, which occurs at larger couplings for larger m_2. All phase contours go through this point since the phase becomes unphysical.

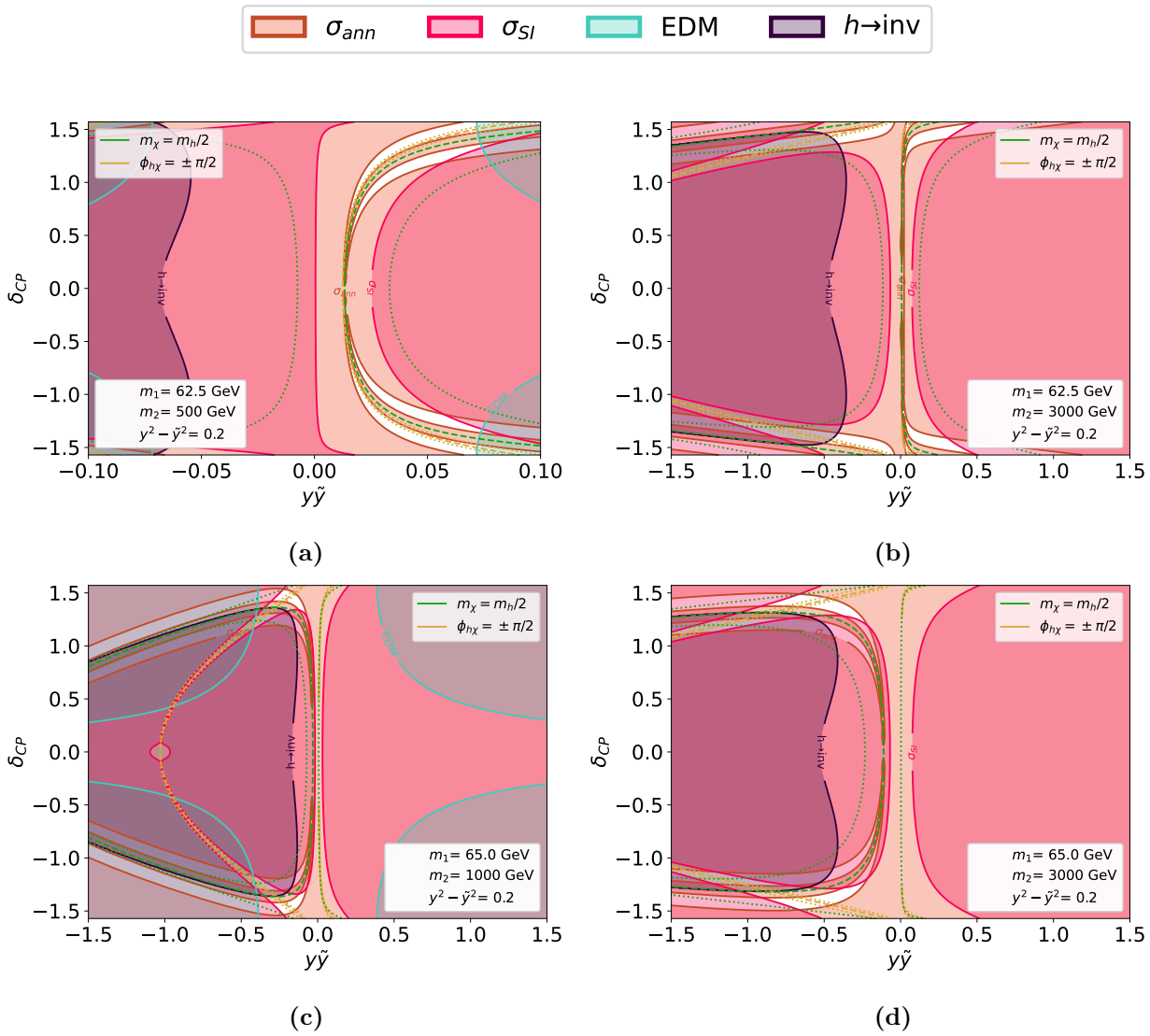


Figure 5.12: Full constraints on the singlet-doublet parameter space in the $y\tilde{y} - \delta_{CP}$ plane, for different values of m_1 and m_2. In this and subsequent plots the shaded regions denote parameter space ruled out by experimental bounds [531, 532, 952, 975]. For annihilation, we include both an upper and lower bound. Constraints that are not relevant for these slices of parameter space are omitted. Dotted lines indicate proximity to mass resonance and pure imaginary EFT coupling: the green dotted lines bound a region with dark matter mass $60\,\text{GeV} \leq m_\chi \leq 65\,\text{GeV}$, the yellow with EFT phase $1.55 \leq \varphi_{h\chi} \leq 1.6$. In figures (a) and (b) we show that viable parameter space can be found at small couplings, corresponding to a pure mass resonance with flexibility in $\varphi_{h\chi}$. In this case, smaller values of m_2 are allowed but m_1 must be close to $m_h/2$. Figures (b) - (d) also show allowed parameter space for larger couplings: (b) shows $m_1 \approx m_h/2$; (c) and (d) show m_1 further away from $m_h/2$ for two different values of m_2. In all of (b) - (d), viable parameter space requires $m_1 \approx 60 - 70$ GeV, $\delta_{CP} \gtrsim 1$, and $\varphi_{h\chi} \approx \pi/2$.

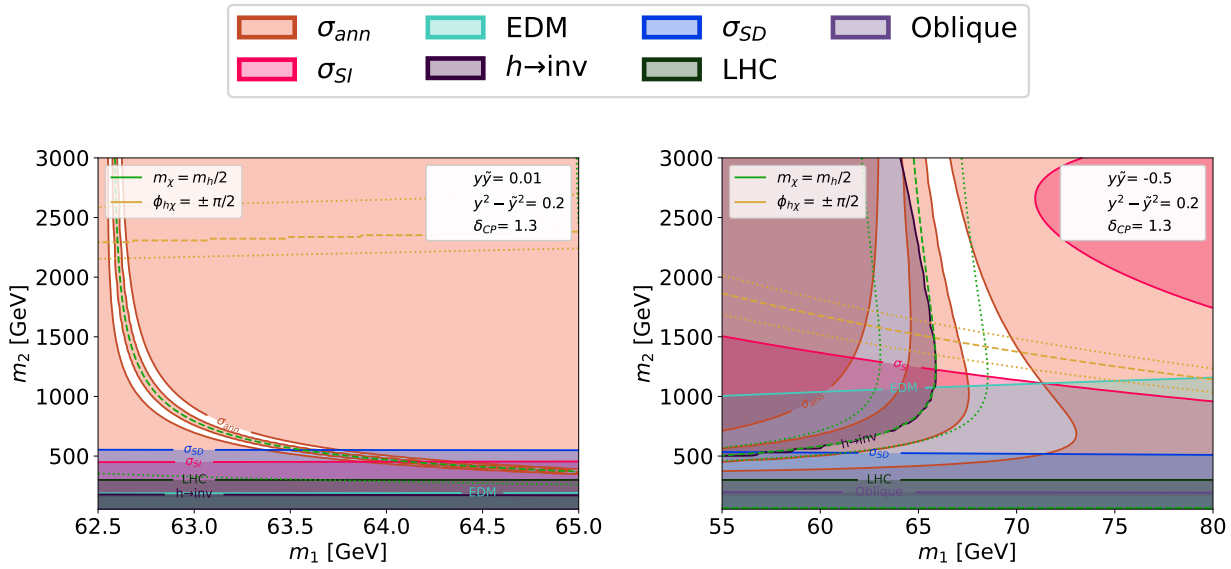


Figure 5.13: Similar to Fig. 5.12, in slices of the $m_1 - m_2$ plane and for different values of $y\tilde{y}$ and δ_{CP}. We show the same constraints as Fig. 5.12 and constraints from [533,535,536,957–967]. Dotted lines around the critical mass and phase values give a guide towards the proximity of any viable space to mass resonance and pure imaginary EFT coupling: the green dotted lines bound a region with dark matter mass $60\,\text{GeV} \leq m_\chi \leq 65\,\text{GeV}$, the yellow with EFT phase $1.55 \leq \varphi_{h\chi} \leq 1.6$. The left shows a mass resonance with small couplings, where m_2 down to ~ 500 GeV is allowed. The right shows larger couplings, where we need $m_2 \gtrsim \mathcal{O}(1)$ TeV. We omit light charged fermion constraints since small m_2 is already ruled out.

5.4 Doublet-Triplet Model

In this section, we describe another potential UV completion: doublet-triplet dark matter. This model includes the addition of a doublet Dirac fermion and a triplet Majorana fermion to the Standard Model. It has been previously discussed in other contexts in [918,920,923,924].

5.4.1 Model in the UV

We begin by describing our model and establishing the notation. This model contains a Dirac doublet of two left handed Weyl fermions with hypercharge 1/2 (denoted by ψ_2 and $\tilde{\psi}_2$ as in the singlet-doublet case) and a triplet of Majorana fermions (with components $\psi_3^{-1}, \psi_3^0, \psi_3^1$), all of which are SU(3) singlets. The Lagrangian is given by

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{kinetic}} - \frac{1}{2} m_3 \psi_3 \psi_3 - m_2 \psi_2 \cdot \tilde{\psi}_2 - Y H^\dagger \psi_3 \psi_2 - \tilde{Y} (\varepsilon H^*)^\dagger \psi_3 \tilde{\psi}_2 + \text{h.c.} \tag{5.36}$$

As in the singlet-doublet case, this theory also has a single physical phase, and we can choose the same convention as the previous section to localize CP-violation to the Yukawa couplings, where

$$\mathrm{Y} \equiv \mathrm{y}\, \mathrm{e}^{\mathrm{i}\delta_{\mathrm{CP}}/2}, \qquad \tilde{\mathrm{Y}} \equiv \tilde{\mathrm{y}}\, \mathrm{e}^{\mathrm{i}\delta_{\mathrm{CP}}/2}. \tag{5.37}$$

Next we describe our notation after SSB. We denote the gauge basis neutral particles by $\psi_{\mathrm{n}} = \{\psi_3^0, \psi_2^{\mathrm{s}}, \tilde{\psi}_2^{\mathrm{d}}\}$ and the gauge basis charged particles by $\psi_{\mathrm{c}}^+ = \{\psi_3^{+1}, \psi_2^{+1}\}$ and $\psi_{\mathrm{c}}^- = \{\psi_3^{-1}, \tilde{\psi}_2^{-1}\}$. We label the neutral mass eigenstates $\chi_{\mathrm{n}} = \{\chi, \chi_1, \chi_2\}$ and the charged mass eigenstates $\chi_{\mathrm{c}}^+ = \{\chi_1^{+1}, \chi_2^{+1}\}$, and $\chi_{\mathrm{c}}^- = \{\chi_1^{-1}, \chi_2^{-1}\}$. Each is ordered from least to most massive, and χ again denotes the dark matter. We call the basis change matrices $\mathrm{J_n}$ and $\mathrm{J_c^{\pm}}$, which are defined by $\psi_{\mathrm{n}} = \mathrm{J_n}\chi_{\mathrm{n}}$, $\psi_{\mathrm{c}}^{\pm} = \mathrm{J_c^{\pm}}\chi_{\mathrm{c}}^{\pm}$. The phases of the eigenvectors are chosen such that the mass eigenvalues are real. Then the mass terms are given by

$$\mathcal{L}_{\mathrm{mass}} = -\frac{1}{2}\chi_{\mathrm{n}}[\mathrm{J_n^T M_n J_n}]\chi_{\mathrm{n}} - \chi_{\mathrm{c}}^-[\mathrm{J_-^T M_c J_+}]\chi_{\mathrm{c}}^+ + \mathrm{h.c.} \tag{5.38}$$

with

$$\mathrm{M_n} \equiv \begin{pmatrix} \mathrm{m_3} & (\mathrm{y}-\tilde{\mathrm{y}})\mathrm{v}/2\sqrt{2} & (\mathrm{y}+\tilde{\mathrm{y}})\mathrm{v}/2\sqrt{2} \\ (\mathrm{y}-\tilde{\mathrm{y}})\mathrm{v}/2\sqrt{2} & -\mathrm{m_2} & 0 \\ (\mathrm{y}+\tilde{\mathrm{y}})\mathrm{v}/2\sqrt{2} & 0 & \mathrm{m_2} \end{pmatrix}, \qquad \mathrm{M_c} \equiv \begin{pmatrix} \mathrm{m_3} & -\mathrm{yv}/\sqrt{2} \\ -\tilde{\mathrm{y}}\mathrm{v}/\sqrt{2} & \mathrm{m_2} \end{pmatrix}. \tag{5.39}$$

The Higgs Yukawa couplings are

$$\mathcal{L}_{\mathrm{Higgs}} = \frac{1}{2}\mathrm{h}\chi_{\mathrm{n}}[\mathrm{J_n^T Y_n J_n}]\chi_{\mathrm{n}} + \mathrm{h}\chi_{\mathrm{c}}^-[\mathrm{J_-^T Y_c J_+}]\chi_{\mathrm{c}}^+ + \mathrm{h.c.} \tag{5.40}$$

with

$$\mathrm{Y_n} \equiv \begin{pmatrix} 0 & -(\mathrm{y}-\tilde{\mathrm{y}})/2\sqrt{2} & -(\mathrm{y}+\tilde{\mathrm{y}})/2\sqrt{2} \\ -(\mathrm{y}-\tilde{\mathrm{y}})/2\sqrt{2} & 0 & 0 \\ -(\mathrm{y}+\tilde{\mathrm{y}})/2\sqrt{2} & 0 & 0 \end{pmatrix}, \qquad \mathrm{Y_c} \equiv \begin{pmatrix} 0 & \mathrm{y}/\sqrt{2} \\ \tilde{\mathrm{y}}/\sqrt{2} & 0 \end{pmatrix}. \tag{5.41}$$

The Z couplings are

$$\mathcal{L}_{\mathrm{Z}} = \frac{1}{2}\mathrm{Z}^{\mu}\bar{\sigma}_{\mu}\chi_{\mathrm{n}}[\mathrm{J_n^\dagger U_n J_n}]\chi_{\mathrm{n}} + \mathrm{Z}^{\mu}\chi_{\mathrm{c}}^+\bar{\sigma}_{\mu}[\mathrm{J_+^\dagger U_+ J_+}]\chi_{\mathrm{c}}^+ + \mathrm{Z}^{\mu}\chi_{\mathrm{c}}^-\bar{\sigma}_{\mu}[\mathrm{J_-^\dagger U_- J_-}]\chi_{\mathrm{c}}^-, \tag{5.42}$$

with

$$\begin{gathered} \mathrm{U_n} \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\sqrt{\mathrm{g}^2+\mathrm{g}'^2} \\ 0 & -\sqrt{\mathrm{g}^2+\mathrm{g}'^2} & 0 \end{pmatrix}, \\ \mathrm{U_+} \equiv \begin{pmatrix} \frac{\mathrm{g}^2}{\sqrt{\mathrm{g}^2+\mathrm{g}'^2}} & \\ 0 & \frac{(\mathrm{g}^2-\mathrm{g}'^2)}{2\sqrt{\mathrm{g}^2+\mathrm{g}'^2}} \end{pmatrix}, \quad \mathrm{U_-} \equiv \begin{pmatrix} -\frac{\mathrm{g}^2}{\sqrt{\mathrm{g}^2+\mathrm{g}'^2}} & 0 \\ 0 & -\frac{(\mathrm{g}^2-\mathrm{g}'^2)}{2\sqrt{\mathrm{g}^2+\mathrm{g}'^2}} \end{pmatrix}, \end{gathered} \tag{5.43}$$

while the W couplings are

$$\mathcal{L}_W = W^{\mu+}\chi_n\bar{\sigma}_\mu[J_n^\dagger D_- J_-]\chi_c^- + W^{\mu-}\chi_n\bar{\sigma}_\mu[J_n^\dagger D_+ J_+]\chi_c^+ + \text{h.c.} \tag{5.44}$$

with

$$D_- \equiv \begin{pmatrix} g & 0 \\ 0 & g/2 \\ 0 & -g/2 \end{pmatrix}, \qquad D_+ \equiv \begin{pmatrix} -g & 0 \\ 0 & g/2 \\ 0 & g/2 \end{pmatrix}. \tag{5.45}$$

The charged fermions also couple to the photon with charge ± 1.

5.4.2 Constraints

We treat most of the constraints in the doublet-triplet model similarly to those in the singlet-doublet model. There are two exceptions that we discuss in more detail: the EDM and collider constraints.

The EDM calculation differs from the singlet-doublet case because there are additional diagrams. Like in the singlet-doublet case, the relevant contributing diagrams are all Barr-Zee diagrams [953]. The diagram with charged W legs, shown in Fig. 5.10, that contributed in the singlet-doublet case is still relevant, but for the doublet-triplet model there are two additional relevant Barr-Zee diagrams: Zh and γh, shown in Fig. 5.14. There is still no γZ contribution because in that case the same charged fermion runs through the entire loop, leaving no place for CP-violation to enter since the diagonal Z coupling is real. We also neglect the hh diagram since it is suppressed by two factors of the electron Yukawa. We use the general forms of the Zh and γh contributions from [976],

$$d_e^{hV} = \frac{1}{16\pi^2 m_h^2}\int_0^1 dx \frac{1}{x(1-x)} j\left(\frac{m_V^2}{m_h^2}, \frac{\tilde{\Delta}^V}{m_h^2}\right) g_e^V c_O^V \frac{m_e}{v}, \tag{5.46}$$

where g_e^V is the electron coupling to Z or γ, and we define

$$j(r,s) = \frac{1}{r-s}\left(\frac{r\log r}{r-1} - \frac{s\log s}{s-1}\right). \tag{5.47}$$

c_O^V and $\tilde{\Delta}^V$ are determined by the inner fermion loop which only contains charged fermions for both γh and γZ. They are given by

$$\begin{aligned} c_O^Z &= -\frac{e}{2\pi^2}\text{Re}\Big(m_c^i x^2(1-x)(g_{ij}^S g_{ji}^{V*} + i g_{ij}^P g_{ji}^{A*}) + (1-x)^3 m_c^j (g_{ij}^S g_{ji}^{V*} - i g_{ij}^P g_{ji}^{A*})\Big), \\ \tilde{\Delta}^Z &= \frac{x m_c^i + (1-x)m_c^j}{x(1-x)}, \qquad c_O^\gamma = -\frac{e^2 g_{jj}^P}{2\pi^2}(1-x)m_c^j, \qquad \tilde{\Delta}^\gamma = \frac{(m_c^j)^2}{x(1-x)}, \end{aligned} \tag{5.48}$$

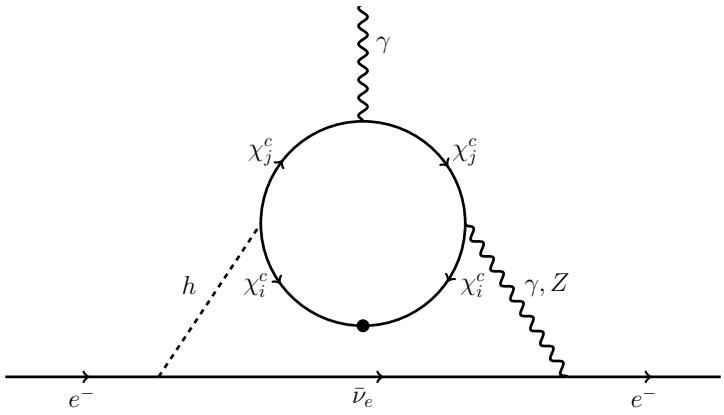


Figure 5.14: Additional class of Barr-Zee diagrams contributing to the electron EDM. χ_c is the tuple of charged fermions in the mass basis. For the γh diagram, i = j, whereas for the Zh diagram, we also have contributions where $i \neq j$.

where

$$\begin{aligned} g^S &= \frac{1}{2}(J_-^T Y_c J_+ + J_+^\dagger Y_c^\dagger J_-^*), & g^P &= \frac{i}{2}(J_-^T Y_c J_+ - J_+^\dagger Y_c^\dagger J_-^*), \\ g^V &= J_-^T U_+ J_-^* + J_+^\dagger U_+ J_+, & g^A &= J_-^T U_+ J_-^* - J_+^\dagger U_+ J_+ \end{aligned} \tag{5.49}$$

are given in terms of the matrices defined in §5.4.1. By definition, χ_j is the fermion which radiates the on-shell external photon, and $g_{ij}^* = (g_{ji})^*$.

A key difference between the singlet-doublet and doublet-triplet cases is that in the latter the mass of the lightest charged fermion is set by similar scales as those that set the mass of the dark matter, and thus generically the lighest charged fermion mass is $\mathcal{O}(100)$ GeV for the doublet-triplet model. This allows us to treat collider constraints differently here than in the singlet-doublet case; we apply generic LEP constraints on charged fermions rather than running the full collider pipeline we considered previously. Specifically, charged fermions lighter than 92.4 GeV are ruled out as long as the mass splitting between the lightest neutral and lightest charged particle is ≥ 100 MeV [950, 951].[6]

5.4.3 Full Exclusion Limits and Discussion

Unlike in the singlet-doublet case, there is no viable parameter space in this model. In order to show this, we consider three different cases. First, we discuss the case where the magnitude of the couplings is small, for any phase. Then we discuss the case of large coupling and large phase. Finally we discuss the case of large coupling but very small phase.

[6] If the lightest charged state is more than 3 GeV heavier than the lightest neutral state, then there is a stronger bound ruling out charged fermions up to mass 103.5 GeV [951]. We use the smaller of the two values for simplicity since it is sufficient for our purposes.

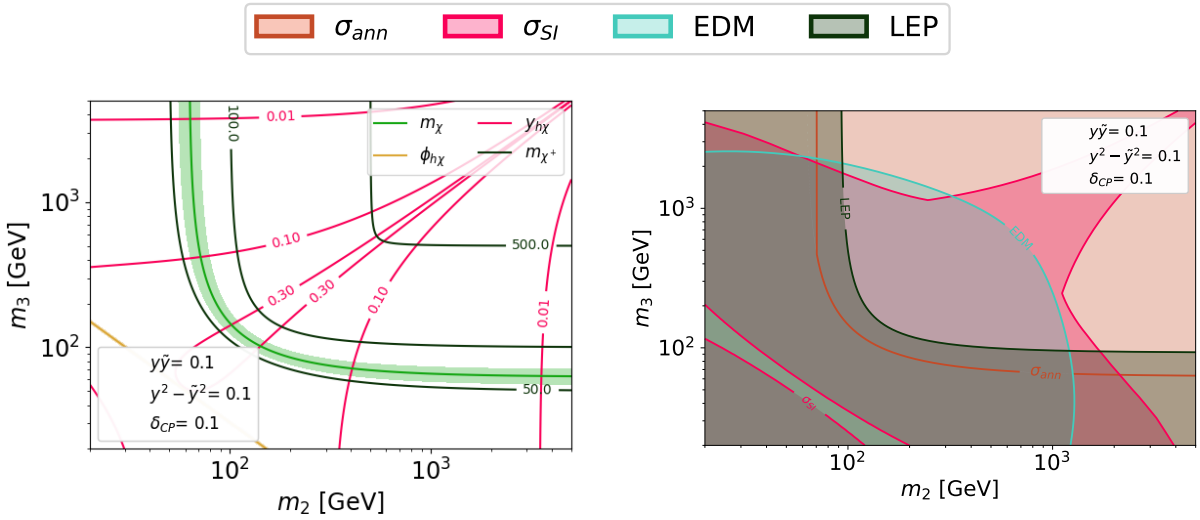


Figure 5.15: Example where the magnitude of couplings y and $\tilde{y}$ are small, for different values of m_2 and m_3. The left plot shows the values of the EFT parameters: dark matter mass, dark matter-Higgs coupling phase, dark matter-Higgs coupling magnitude, and lightest charged fermion mass. Regions around the mass and phase points of interest are shaded: 55 GeV $\leq m_\chi \leq$ 70 GeV and $1.3 \leq \varphi_{h\chi} \leq 1.85$. The right shows the annihilation signal and a subset of relevant constraints including EDM [952], spin-independent direct detection [531, 532], and charged fermion constraints from LEP [950, 951]. The annihilation signal appears as a single brown line because a viable annihilation signal is only achievable in a tuned region of parameter space.

In the first case, parameter space is entirely ruled out by charged fermion constraints, as we can see from Fig. 5.15. On the left, this figure shows the values of several EFT parameters for fixed y, $\tilde{y}$, and δ_{CP} and various values of m_2 and m_3. On the right, we show the annihilation signal and a subset of constraints that are sufficient to rule out this region of parameter space.[7] From these plots, we can see that since the couplings are small, in order to get a sufficient annihilation signal one of m_2 or m_3 must be $\gtrsim m_h/2$, with the other UV mass larger. Since the magnitude of the couplings is small while the UV masses are large, in this region there will only be a very small splitting between charged and neutral fermions. Therefore, the parameter space here will be entirely ruled out by charged fermion constraints from LEP. This occurs regardless of phase, though EDM constraints are also strong enough to rule this out for larger phases.

In the second case of large coupling and large phase, EDM constraints are typically very strong. The only exceptions are if both m_2 and m_3 are very large (which can't generate the necessary annihilation signal) or if one of m_2 or m_3 is very small. This is because in the limit where one of m_2 or m_3 is exactly zero, the phase becomes unphysical since we can rotate it away. In the limit where m_2 is small, the lightest state will have mass even less than m_2 and the DM mass won't be in the right mass range to generate the necessary annihilation signal. But in the limit where m_3 is small, if the couplings are large enough we can potentially

[7] The other constraints from the singlet-doublet case still apply here, but we omit them from these plots for clarity.

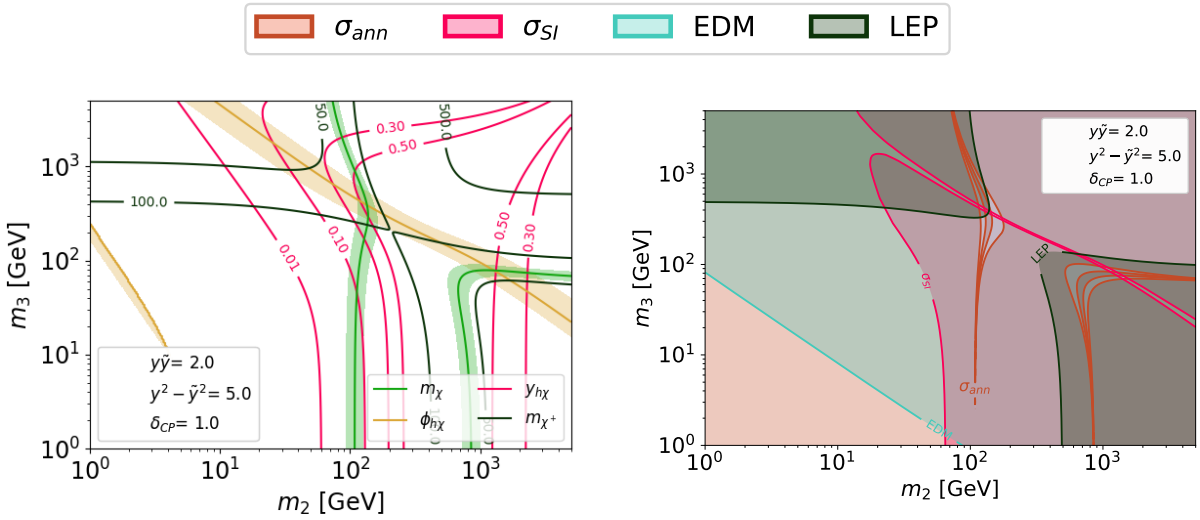


Figure 5.16: A case where the magnitudes of couplings y and $\tilde{y}$ are large and δ_{CP} is also large, for different values of m_2 and m_3. The left plot shows the values of EFT parameters: dark matter mass, dark matter-Higgs coupling phase, dark matter-Higgs coupling magnitude, and lightest charged fermion mass. Regions around the mass and phase points of interest are shaded: 55 GeV $\leq m_\chi \leq$ 70 GeV and $1.3 \leq \varphi_{h\chi} \leq 1.85$. The right shows the annihilation signal and a subset of relevant constraints, and from here we can see that the combination of EDM constraints and spin-independent constraints entirely rule out the parameter space generating a viable annihilation signal.

generate the right annihilation signal. However, since the physical phase is small, the EFT phase will also be small, and spin-independent direct detection constraints will always rule out any part of the annihilation signal that isn't constrained by the EDM. This can be seen in Fig. 5.16, which again shows various values of EFT parameters for fixed y, $\tilde{y}$, and δ_{CP} and different m_2 and m_3 values on the left, and the annihilation signal and a subset of constraints on the right.

The third case of large magnitude coupling but very small phase is shown in Fig. 5.17. The top plots show the case where y and $\tilde{y}$ are similar in magnitude, while the bottom plots show a large splitting between y and $\tilde{y}$. In both, the EFT coupling is mostly real since the phase is small. There are two different trends depending on the magnitude of the coupling. In both plots, we see regions where the magnitude of the EFT coupling is large, and the annihilation signal is ruled out by spin-independent constraints. In the case of small splitting, we also see a region where the EFT coupling is small (because the lightest state doesn't mix), which is unable to generate the necessary annihilation signal.

5.5 Discussion

Given that the GCE is one of the most persistent signals of potential new physics, it is worth cataloging and understanding what could generate it. While there is still substantial debate over the source of the

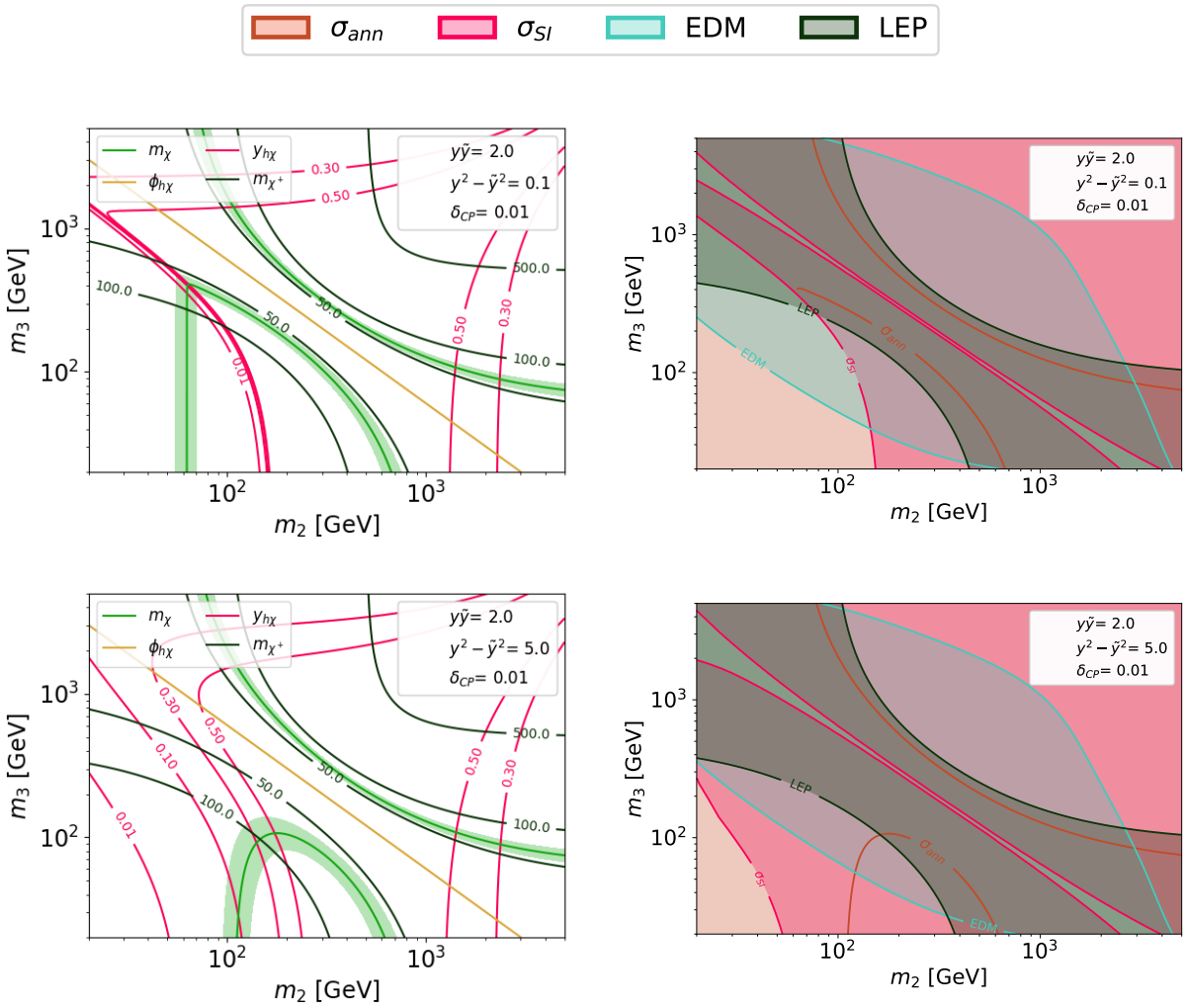


Figure 5.17: Two examples where the magnitude of couplings y and $\tilde{y}$ are large and δ_{CP} is small, for different values of m_2 and m_3. The top plots show the case where y and $\tilde{y}$ have similar magnitudes; the bottom plots show the case where their magnitudes are very different. As in the other doublet-triplet plots, the left plots show the values of various EFT parameters with shaded regions of interest and the right plots show the annihilation signal and a subset of relevant constraints. The annihilation signal appears as two brown lines on each plot, since the region of allowed masses is so narrow. In both cases, spin-independent constraints rule out the signal. In the case where the couplings are nearly equal, there is also a region where the lightest neutral state decouples, and the dark matter-Higgs coupling is insufficient to generate the annihilation signal despite the dark matter mass being close to $m_h/2$.

GCE, one promising and well explored possibility is dark matter annihilating to $b\bar{b}$. In this work, we revisit the particular case where dark matter is a Majorana fermion with a CP-violating Higgs coupling, which allows annihilation and spin-independent scattering to be governed by different parameters. Specifically, the leading contribution to annihilation is determined by the imaginary part of the coupling to the Higgs, while spin-independent scattering constraints depend primarily on the real part of the coupling to the Higgs. We study the EFT of this dark matter model for the GCE in detail, and find that while tuning the dark matter mass very close to half the Higgs mass is one potential way to obtain a large enough signal, tuning the phase of the Higgs coupling to be near pure imaginary loosens this restriction in the context of the EFT.

We also explore two potential UV completions: singlet-doublet dark matter and doublet-triplet dark matter. In both, the story is more complicated than the EFT because the UV phase and mass are not independent parameters. Although more elaborate supersymmetric realizations of a CP-violating Higgs portal have been discussed in [912], our goal throughout this chapter has been to gain a more detailed qualitative understanding of the mechanism through simpler models. In particular, we have discussed the scaling of the signal and various constraints with the different parameters in the simplified models, as well as quantified how much tuning is necessary to explain the signal without running into constraints. The singlet-doublet dark matter case is particularly interesting because it is a minimal working example of how Majorana dark matter could explain the GCE through the Higgs portal.

We find that in the minimal singlet-doublet case, there is still viable parameter space when the doublet mass is much larger than the singlet mass. There are two viable regions of parameter space for the singlet-doublet model. In the case where the UV couplings are small, the tuning of the dark matter mass manifests as a tuning of the singlet mass, but the restriction on both UV and EFT phase is loose. When the couplings are larger, the doublet mass is required to be $\gtrsim \mathcal{O}(1)$ TeV. The EFT phase, and often the UV phase as well, must be close to pure imaginary to avoid spin-independent constraints, and the dark matter and singlet masses also must still be relatively close to $m_h/2$ to generate a sufficient annihilation signal (though the allowed region is comparatively much wider).

Upcoming direct detection and EDM experiments, such as LZ, XENONnT and ACME, will search through significant portions of the remaining parameter space. These two types of probes combine to explore both the limits of minimal and maximal CP-violation, and we expect to definitively rule out doublet masses below the TeV scale in the small coupling case. In the more optimistic case of larger coupling, new experiments will be able to probe doublet masses up to $\mathcal{O}(15)$ TeV or larger for some phases. In either case, this type of model offers a range of complementary detection avenues that may combine to elucidate the nature of annihilating dark matter.

In the doublet-triplet case, we do not find any viable parameter space. Spin-independent and EDM

constraints restrict the size of the real and imaginary parts of the Higgs coupling, respectively. When the coupling is small in overall magnitude, the annihilation signal requires a dark matter mass near the $m_h/2$ resonance, and the small splitting between the lightest charged and neutral states results in a prohibitively light charged fermion. Hence, the remaining parameter space is ruled out by LEP.

While this chapter is framed in the context of the GCE, models with a CP-violating Higgs portal coupling the dark and visible sectors are also compelling for other reasons. These interactions could be the key to other mysteries in particle physics, such as the particle nature of dark matter and problems that CP-violation is necessary to solve, including the matter/antimatter asymmetry of the universe and the strong CP problem. For example, for some models the addition of a CP phase around the weak scale could increase the viability of electroweak baryogenesis. While new Higgs boson couplings have the potential to make the hierarchy problem worse, the minimal models we studied can also be realized within the larger framework of SUSY [912] which can ameliorate this issue. These connections could be potential avenues for further exploration, if it turns out that dark matter communicates with the Standard Model through a CP-violating Higgs portal.

6

Oblique Lessons from the W Mass Measurement at CDF II

6.1 Introduction

The Standard Model of particle physics (SM) has been remarkably successful in explaining various experimental results. The discovery of the Higgs boson [977, 978] at the Large Hadron Collider (LHC) was imperative to confirming the pattern of spontaneous symmetry breaking in the electroweak sector of the SM. However, as we continue to collect data and improve analysis techniques, we have seen a proliferation of precision measurements that deviate from SM predictions, such as the muon magnetic moment [683, 684, 979] and the R_K/R_K^* anomalies [980–982]. The most recent anomalous measurement reported is the mass of the W boson M_W [11]. A discrepant measurement of M_W could be an indication of supersymmetry (SUSY), composite Higgs, or other phenomena beyond the Standard Model (BSM) at potentially very high energy scales. It is therefore essential that we explore the phenomenological implications of this new M_W measurement.

In order to quantify the compatibility of the W mass measurement with the SM prediction with high precision, we perform a global fit of the SM, known as the electroweak fit. This method involves fitting over a set of well-measured SM observables, and minimizing the χ^2 value over both the fitted ('free') observables as well as derived observables, see Refs. [983–985]. The electroweak fit leverages the small uncertainties of the fitted observables to produce precise predictions of the derived observables. Additionally, since this fit is an exceptional probe of precision measurements, it is also highly sensitive to BSM effects.

For scenarios where new physics contributions dominantly appear as corrections to the SM gauge boson propagators, we can parameterize the effects of new physics phenomena on the electroweak sector using oblique parameters S, T, and U [955,986], which were previous defined in (5.29) - (5.31) (see also Refs. [987–989]). These parameters capture the effects of higher-dimension operators [990, 991] that can arise in a variety of UV completions. In many models, S and T are the dominant corrections since they arise from dimension-6 operators, whereas U is dimension-8 and therefore suppressed by a factor of v^2/Λ_{UV}^2.

The power of the electroweak fit is dependent on precision of experimental measurements of SM observables, and improves along with collider technology and luminosity. The leading measurements are made at the Large Electron-Positron Collider (LEP), Stanford Linear Collider (SLC), Tevatron, and LHC. The discovery of the Higgs greatly improved the electroweak fit as it provided the final measured value to span the free parameters of the SM [992–994].

The most recent update to the SM values used in the fit comes from the CDF collaboration at the Tevatron [11]. Their analysis was completed with a four-fold increase of data, reduced uncertainty in PDFs and track reconstruction, and updated measurements compared to their previous result [995]. They report

$$M_{W,\text{CDF II}} = 80.4335 \pm 0.0094 \text{ GeV}, \tag{6.1}$$

which, without averaging with other experimental results, shows a 7σ deviation from the SM prediction. This value is notably higher than the previous measurement averaged from the Tevatron and LEP experiments ($M_W = 80.385 \pm 0.015$ GeV) [996], as well as ATLAS ($M_W = 80.370 \pm 0.019$ MeV) [997] and LHCb ($M_W = 80.354 \pm 0.032$ GeV) [998].

In this chapter we explore how new physics contributions, parameterized by the values of the oblique parameters, can adjust the electroweak fit such that M_W is consistent with the updated CDF measurement. We first perform our fits scanning over values of S and T with U fixed to zero (since U is suppressed) and identify the range of these variables that can resolve the observed anomaly in M_W. We then study how the fit changes if we allow U to float. Large values of U can easily accommodate the observed increase in M_W; however, it is difficult to construct models with the primary new physics contributions affecting only U while leaving S and T unchanged.

Next we consider several well-motivated simple extensions of SM that can produce nonzero S and T values. The models discussed in this chapter include a scalar singlet, a two Higgs doublet model (2HDM), a neutral scalar $SU(2)_L$ triplet (referred to here as a *swino*), and various singlet-doublet fermion scenarios. For each model we check if there is available parameter space that corresponds to the fitted values of T and S. We find that extending the SM with a scalar singlet or doublet cannot explain the observed anomaly in M_W measurements, while a singlet-doublet fermion extension is strongly constrained by various experimental

bounds. A $\mathcal{O}$(TeV) swino, on the other hand, can explain the observed anomaly while evading current bounds and provides a well-motivated target for future high energy colliders.

The remainder of the chapter is organized as follows. In §6.2, we define the parameters and methodology of our electroweak fit. §6.3 discusses the results and the implications of the oblique parameters on fitting the measured observables. In §6.4 we map the values of the fitted oblique parameters to the parameters of various models, and comment on the viability of this space. We discuss in §6.5.

6.2 Electroweak Fit

To assess the impact of the new measurements of M_W, and the implications for potential new physics, we perform an electroweak fit to a representative set of observables, following the strategy of the GFitter group [983–985, 999][1] with a modified version of the code used in Refs. [1001, 1002]. A set of five core observables are free to vary in the fit: the Z boson mass M_Z, the top mass M_t, the Higgs mass M_h, the Z-pole value of the strong coupling constant $\alpha_s(M_Z)$, and the hadronic contribution to the running of α, denoted $\Delta\alpha^{(5)}_{had}(M_Z^2)$.

These five values float between their experimental uncertainties. The other observables in the fit have theoretical predictions formulated with the floating observables and are compared to their measured values (see Table 6.1). In addition to measurements of these five parameters, the observables considered include the W mass and a host of other electroweak precision measurements performed at SLC, LEP, the Tevatron, and the LHC, which are listed with their measured values below the horizontal line in Table 6.1. These other observables can be determined in the SM as functions of the five core observables, the Fermi constant G_F, and the fine structure constant $\alpha(q^2 = 0)$. In the electroweak fit, $G_F = 1.1663787 \times 10^{-5}\,\text{GeV}^{-2}$ and $\alpha = 1/137.03599084$ are treated as fixed values since they are determined with much higher precision than the rest of the observables [992].

For the W mass, we will consider several different values to assess the impact of the recent CDF measurement on the overall state of the global EW fit. These are,

$$\begin{aligned} M_W &= 80.4335 \pm 0.0094 \text{ GeV} \qquad \text{(CDF II)}, \\ M_W &= 80.4112 \pm 0.0076 \text{ GeV} \qquad \text{(LHC + LEP + Tevatron)}, \\ M_W &= 80.379 \pm 0.012 \text{ GeV} \qquad \text{(PDG 2020)}, \end{aligned} \tag{6.2}$$

where the uncertainties quoted above include the statistical, systematic and modeling uncertainties used in each experiment. The second scenario is our estimate for the global average of different M_W measurements,

[1] With respect to the GFitter results in Ref. [999], we consider an updated value of the Higgs and Top-quark masses and the revised values of Γ_Z and σ^0_{had} from Ref. [1000].

Observable	Measured Value
M_Z [GeV]	91.1876 ± 0.0021
M_h [GeV]	125.25 ± 0.17
M_t [GeV]	172.69 ± 0.58
$\alpha_s(M_Z^2)$	0.1181 ± 0.0011
$\Delta\alpha_{had}^{(5)}(M_Z^2)$	0.02766 ± 0.00007
Γ_Z [GeV]	2.4955 ± 0.0023
Γ_W [GeV]	2.085 ± 0.042
σ_{had}^0 [nb]	41.481 ± 0.0325
R_ℓ^0	20.767 ± 0.0247
$A_{FB}^{0,\ell}$	0.0171 ± 0.0010
A_ℓ	0.1499 ± 0.0018
$\sin^2\theta_{eff}^\ell(Q_{FB})$	0.2324 ± 0.0012
$\sin^2\theta_{eff}^\ell$(Tevt.)	0.23148 ± 0.00033
A_b	0.923 ± 0.020
A_c	0.670 ± 0.027
$A_{FB}^{0,b}$	0.0992 ± 0.0016
$A_{FB}^{0,c}$	0.0707 ± 0.0035
$R^{0,b}$	0.21629 ± 0.00066
$R^{0,c}$	0.1721 ± 0.0030

Table 6.1: Summary of the observables included in the fit, and their experimental values. The five observables above the horizontal line are allowed to float in the fit, while the SM values of the remaining observables are determined from these five values, as discussed in the main text. The values of M_Z, M_t, M_h, $\alpha_s(M_Z^2)$, $\Delta\alpha_{had}^{(5)}(M_Z^2)$, and Γ_W are taken from the most recent PDG average [992]. Following [999], for M_t we also include an additional theory error of 0.5 GeV in addition to the experimental error from [992]. For Γ_Z and σ_{had}^0 we use the updated values computed in Ref. [1000]. The remaining Z-pole observables are taken from the LEP and SLC measurements [1003]. For A_ℓ we use the average of the LEP and SLC values, following Ref. [999].

assuming zero correlations between experimental result to first approximation.[2] In addition, to assess the particular impact of the new, high precision measurement from CDF II, we will also perform the fit with M_W taken to be the CDF II value with the systematic uncertainty artificially inflated by a factor of two, $M_W = 80.4335 \pm 0.0157$, to better understand the compatibility of the CDF measurement with the SM prediction. This scenario is referred to as the CDF II (2× Syst.) throughout the chapter.

The SM values of the other observables are determined from the free parameters using the full two-loop electroweak results available in the literature. The running of α is computed using the floating value of $\Delta\alpha_{had}^{(5)}$ as well as the leptonic piece, $\Delta\alpha_{lep} = 0.031497686$ [1004], which is kept fixed in the fit. The W mass is determined using the parameterization in Ref. [1005], which also includes corrections up to $\mathcal{O}(\alpha\alpha_s^3)$

[2]While there are sources of uncertainty such as parton distribution functions that might introduce some correlation between these results, when we repeated the world average M_W scenario (Tevatron + LEP + LHC) with a few different values for the correlations, we arrived at similar qualitative results. A comprehensive global averaging of these experimental results considering all correlations is left for future work.

for the radiative correction (referred to as Δr in the literature). The expression for the width of the W is taken from the parameterization in Ref. [1006]. For the Z width Γ_Z, hadronic peak cross section σ^0_{had}, and width ratios R^0_ℓ, R^0_b, R^0_c, we use the parameterizations in Ref. [1007]. For the effective weak mixing angle, $\sin^2\theta^\ell_{eff}$, we use the results in Ref. [1008]. The value of $\sin^2\theta^\ell_{eff}$ is used as a proxy for the weak mixing angle to determine the left- and right-handed couplings of the Z, allowing us to compute the asymmetries:

$$\mathcal{A}_f = \frac{g^2_{Lf} - g^2_{Rf}}{g^2_{Lf} + g^2_{Rf}} \tag{6.3}$$

for $f = \ell, c, b$. The value of $\sin^2\theta^\ell_{eff}$ is also used to compute the forward-backward asymmetry $A^{0,\ell}_{FB}$. Finally, for the other forward-backward asymmetries, we compute the effective weak mixing angles $\sin^2\theta^b_{eff}$ and $\sin^2\theta^c_{eff}$ using the parameterizations in Refs. [1008, 1009], respectively. These are then translated to $A^{0,b,c}_{FB}$ using the standard relations summarized e.g., in Ref. [1009]. See also Ref. [1010] for a recent review of the status of relevant theoretical calculations.

We parameterize potential effects of BSM physics in the electroweak fit in terms of the oblique parameters, S, T and U [955,986], as defined in (5.29) through (5.31). Note that S, T and U do not completely characterize potential BSM effects in the electroweak precision data—a larger set of oblique parameters was developed in Refs. [956, 958]. We will not consider their effects in this chapter, as they are typically smaller in universal perturbative theories [1001, 1011].[3] The new physics contributions to the electroweak observables can be expressed as linear functions of S, T and U [955, 986, 1012–1014], which are summarized in Appendix A of Ref. [993].

For a class of universal effective theories, both S and T are related to the Wilson coefficients [990,991,1015] of dimension-6 operators[4]:

$$\begin{aligned}\mathcal{L}_{ob} \supset & \frac{c_W s_W}{v^2}\left(\frac{i}{2s_W}E_W(H^\dagger\sigma^a\overleftrightarrow{D}^\mu H)D^\nu W^a_{\mu\nu} + \frac{i}{2c_W}E_B(H^\dagger\overleftrightarrow{D}^\mu H)\partial^\nu B_{\mu\nu} + E_{WB}H^\dagger\sigma^a H W^a_{\mu\nu}B^{\mu\nu}\right) \\ & - E_T\left(\frac{2}{v^2}\right)|H^\dagger\overleftrightarrow{D}_\mu H|^2\end{aligned} \tag{6.4}$$

where

$$\begin{aligned} S &= \frac{4s^2_W}{\alpha}g^2\left(E_{WB} + \frac{1}{4}E_W + \frac{1}{4}E_B\right) \\ T &= \frac{1}{\alpha}E_T. \end{aligned} \tag{6.5}$$

[3] There are some models which contribute dominantly to these other observables; in order to study these models, an electroweak fit including these additional parameters would have to be performed. One such model is the dark photon, which contributes only to Y (defined in (5.33)) at tree-level when the expansion in p^2 is done correctly. We leave the study of these types of models for future work.

[4] This basis choice may look unfamiliar; see Ref. [1016] for a detailed discussion of the relationship between the oblique parameters and effective theories.

		CDF-II	CDF-II (2× syst.)	World Average	PDG
SM	$\chi^2/(n_{d.o.f.} = 15)$	4.03	2.29	2.97	0.97
Best Fit (U = 0)	S	0.15 ± 0.08	0.13 ± 0.08	0.10 ± 0.08	0.03 ± 0.08
	T	0.25 ± 0.06	0.22 ± 0.07	0.18 ± 0.06	0.07 ± 0.06
	$\chi^2/(n_{d.o.f} = 13)$	1.23	1.18	1.03	0.87
Best Fit (U floating)	S	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10
	T	0.03 ± 0.12	0.03 ± 0.12	0.03 ± 0.12	0.03 ± 0.12
	U	0.20 ± 0.09	0.20 ± 0.10	0.14 ± 0.09	0.04 ± 0.09
	$\chi^2/(n_{d.o.f} = 12)$	0.93	0.93	0.93	0.93

Table 6.2: Fit results including the oblique parameters and χ^2 per degree of freedom. Different columns correspond to different input M_W measurement scenarios around Eq. (6.2). The first row shows the χ^2 per degree of freedom for the SM in each M_W scenario. Results of the fit including (S,T) and excluding (including) U in the list of floating parameters are included in the middle (bottom) row. See Appendix D for correlations.

The U parameter is often fixed to zero in electroweak fits, as it corresponds to a dimension-8 operator from an effective field theory point of view, and its effects are therefore subleading compared to S and T. We will frequently set U = 0 in our fits, but consider its effect in more details in §6.3.2. We will discuss new physics interpretations of S and T following the results of the fit with U = 0 in §6.4.

With all of these inputs, we perform the electroweak fit by minimizing a χ^2 function,

$$\chi^2 = \sum_{i,j} (M_i - O_i)(V^{-1}_{cov})_{ij}(M_j - O_j) \tag{6.6}$$

where the sum runs over all the observables in Table 6.1, in addition to the W mass. Here, M_j is the experimentally measured value of the observable, O_j is the predicted value in terms of the five free parameters and S, T, U, and V^{-1}_{cov} is the inverse-covariance matrix for the observables. For the Z lineshape and heavy flavor observables measured at LEP, we use the experimental correlations from refs. [1000, 1003] to compute the covariance matrix. For other observables, we neglect any correlations so that the covariance matrix is diagonal with $(V^{-1}_{cov})_{jj} = 1/\sigma^2_j$. We repeat this calculation for all the four scenarios for M_W measurements defined around Eq. (6.2).

6.3 Results of the Fit

6.3.1 Fitting S and T

We first consider the fit results where U is fixed to zero. The results of our electroweak fit with different values of M_W are summarized in Table 6.2. Correlations are shown in Appendix D.

The first row of Table 6.2 indicates the χ^2 per degree of freedom (d.o.f.) for the SM for the fit with

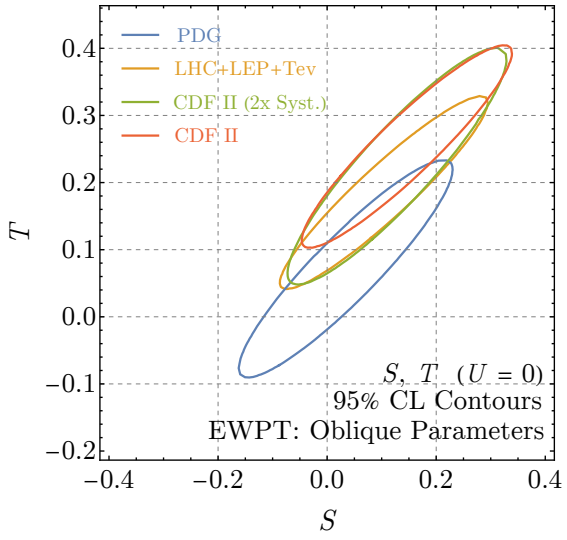


Figure 6.1: The 95% CL preferred regions in the S and T plane with U = 0 from the electroweak fit, marginalizing over the five input parameters and for various experimental values of M_W (see the discussion around Eq. (6.2)). We do not include U in these fits. The blue curve is in good agreement with results of GFitter group [983–985, 999]. Including the recent CDF II measurement of M_W [11] (in the green, yellow, and red curves) moves the best-fit region to larger positive values of S and T. The SM (with (S, T) = (0, 0)) is strongly disfavored when the new CDF II M_W measurement is included in the fit.

each value of M_W. We observe that, prior to the CDF measurement, the Standard Model provides a good fit to the data using the PDG 2020 value of M_W, with $\chi^2/(n_{\text{d.o.f.}} = 15) = 0.97$ (p = 0.48). Taking instead the recent CDF II measurement of M_W, however, the p-value for the SM drops to 2.11×10^{-7}, exemplifying the tension discussed in Ref. [11]. This is somewhat ameliorated when considering the smaller world average value of M_W ($p = 9.01 \times 10^{-5}$), but notable tension remains.

In the middle row of Table 6.2 we summarize the results of the fit when we allow S and T to float in addition to the five free observables. We report the best fit values and confidence intervals of S and T, and then the χ^2 per degree of freedom. We find a good fit to the data with the PDG average value of M_W, prior to the CDF measurement (p = 0.60), where the fit prefers small values of S and T at 0.03 and 0.07, respectively. This is consistent with the electroweak fit presented in Ref. [992]. For all of the fits accounting for the new measurement of M_W from CDF II, the fit instead prefers much larger values of S and T. Despite this, we still find a good fit to the data, with p values ranging from 0.24 when using the CDF measurement alone to 0.42 using the combination of measurements at LHC, Tevatron, and LEP.

The last row of Table 6.2 shows the results of the fit done when U is allowed to float as well. We find that when including the CDF-II measurement, the fit favors a large value of U and small S and T. Since this result is unnatural from a model building perspective, we proceed with the results of the fit with U = 0.

The results of the fit for the oblique parameters S and T are illustrated in Fig. 6.1. Here we show ellipses indicating the 95% CL contours around the best-fit values of S and T. These are computed by

computing the χ^2 at each point in the S − T plane, marginalizing over the free observables, and requiring $\Delta\chi^2 \equiv \chi^2(S,T) - \chi^2_{\rm min} < 6.18$, where $\chi^2_{\rm min}$ is the minimum value of the χ^2 as a function of all the free parameters as well as S and T.

The 95% CL contours of the fit with the PDG average value of M_W (excluding the recent CDF II measurement) are shown in blue and agree with the results of Ref. [999]. This fit slightly prefers $T > 0$, though the correlation between S and T leaves some parameter space with $S, T < 0$ as well. Once the new measurement of M_W from CDF II is included, however, the preferred region in the S − T plane shifts dramatically. The correlation between S and T remains, but values of $T < 0$ are no longer allowed, even when the systematic error on the CDF measurement is artificially inflated. In all, we find a strong preference for BSM contributions in the electroweak fit, particularly for positive, nonzero values of T.

For each fit, we also find the best fit value of each individual observable both for the SM (with S and T fixed to zero) and for the best-fit value of S and T. The results are shown in Table 6.3. Each entry indicates the best-fit value of the observable, along with the pull (calculated as the fit value minus the measured value, divided by the experimental uncertainty) shown in parentheses. For all three values of M_W including the new CDF measurement, we see a significant pull (ranging from -4.6 to -7.0) on the fit value of M_W in the Standard Model. This is entirely ameliorated at the best fit values of S and T, at the cost of a small tension in the value of Γ_Z, which has a fit value larger than the experimental value when S and T are allowed to float. All of the other observables have quite similar values at their best-fit point and at the SM, regardless of the experimental value of M_W used in the fit. Note also that the previously existing tension in the forward-backward asymmetry, $A_{FB}^{0,b}$, measured at LEP is unaffected by the floated values of S and T and is roughly the same for any value of M_W.

6.3.2 The U Parameter

In the fits described above, we have fixed $U = 0$. As discussed in §6.2, this is motivated by the fact that the U-parameter is dimension 8, and is typically suppressed relative to S and T in concrete models.

Nevertheless, in light of the large value of M_W measured at CDF II, it is worth examining the effects of the U-parameter on the electroweak fits in more detail. This is because, of all the electroweak precision observables we consider, the U parameter affects only two: the W mass and width [993, 1012, 1013]:[5]

$$\begin{aligned} M_W &= M_{W,SM}\left(1 - \frac{\alpha(M_Z^2)}{4(c_W^2 - s_W^2)}\left(S - 2c_W^2 T\right) + \frac{\alpha(M_Z^2)}{8s_W^2}U\right), \\ \Gamma_W &= \Gamma_{W,SM}\left(1 - \frac{3\alpha(M_Z^2)}{4(c_W^2 - s_W^2)}\left(S - 2c_W^2 T\right) + \frac{3\alpha(M_Z^2)}{8s_W^2}U\right). \end{aligned} \tag{6.7}$$

[5] We thank Ayres Freitas for emphasizing this point to us.

	Fit Value (Pull) at SM / Best Fit (S,T)							
	CDF-II		CDF-II (2× syst.)		World Average		PDG	
(S, T)	(0,0)	(0.15, 0.25)	(0,0)	(0.13, 0.22)	(0,0)	(0.10, 0.19)	(0,0)	(0.03, 0.07)
M_Z [GeV]	91.1913 (+1.8)	91.1880 (+0.2)	91.1895 (+0.9)	91.1880 (+0.2)	91.1911 (+1.7)	91.1880 (+0.2)	91.1886 (+0.5)	91.1879 (+0.1)
M_h [GeV]	125.24 (-0.1)	125.26 (+0.1)	125.25 (0.0)	125.26 (+0.1)	125.24 (-0.1)	125.25 (0.0)	125.25 (0.0)	125.25 (0.0)
M_t [GeV]	173.94 (+2.2)	172.75 (+0.1)	173.29 (+1.0)	172.71 (0.0)	173.87 (+2.0)	172.70 (0.0)	172.97 (+0.5)	172.68 (0.0)
$\alpha_s(M_Z^2)$	0.1179 (-0.2)	0.1180 (-0.1)	0.1182 (+0.1)	0.1181 (0.0)	0.1180 (-0.1)	0.1181 (0.0)	0.1183 (+0.2)	0.1183 (+0.2)
$\Delta\alpha^{(5)}_{had}(M_Z^2)$	0.02761 (-0.7)	0.02766 (0.0)	0.02763 (-0.4)	0.02766 (0.0)	0.02761 (-0.7)	0.02766 (0.0)	0.02765 (-0.1)	0.02766 (0.0)
M_W [GeV]	80.3681 (-7.0)	80.4261 (-0.8)	80.3613 (-4.6)	80.4182 (-1.0)	80.3674 (-5.8)	80.4075 (-0.5)	80.3579 (-1.8)	80.3769 (-0.2)
Γ_Z [GeV]	2.4950 (-0.2)	2.4994 (+1.7)	2.4947 (-0.3)	2.4988 (+1.4)	2.4949 (-0.3)	2.4981 (+1.1)	2.4946 (-0.4)	2.4960 (+0.2)
Γ_W [GeV]	2.091 (+0.1)	2.096 (+0.3)	2.091 (+0.1)	2.095 (+0.2)	2.091 (+0.1)	2.094 (+0.2)	2.090 (+0.1)	2.092 (+0.2)
$\sigma^0_{had.}$ [nb]	41.489 (+0.2)	41.491 (+0.3)	41.489 (+0.2)	41.491 (+0.3)	41.489 (+0.2)	41.490 (+0.3)	41.489 (+0.2)	41.490 (+0.3)
R^0_ℓ	20.748 (-0.8)	20.751 (-0.6)	20.750 (-0.7)	20.751 (-0.6)	20.748 (-0.8)	20.751 (-0.6)	20.751 (-0.6)	20.751 (-0.6)
$A^{0,\ell}_{FB}$	0.0163 (-0.8)	0.0164 (-0.7)	0.0162 (-0.9)	0.0163 (-0.8)	0.0163 (-0.8)	0.0163 (-0.8)	0.0162 (-0.9)	0.0162 (-0.9)
A_ℓ	0.1474 (-1.4)	0.1477 (-1.2)	0.1470 (-1.6)	0.1476 (-1.3)	0.1473 (-1.4)	0.1475 (-1.3)	0.1469 (-1.7)	0.1472 (-1.5)
$\sin^2\theta^\ell_{eff}(Q_{FB})$	0.2315 (-0.8)	0.2314 (-0.8)	0.2315 (-0.8)	0.2314 (-0.8)	0.2315 (-0.8)	0.2315 (-0.8)	0.2315 (-0.8)	0.2315 (-0.8)
$\sin^2\theta^\ell_{eff}$(Tevt.)	0.23148 (0.0)	0.23144 (-0.1)	0.23152 (+0.1)	0.23145 (-0.1)	0.23148 (0.0)	0.23146 (-0.1)	0.23154 (+0.2)	0.23150 (+0.1)
A_b	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)	0.936 (+0.7)
A_c	0.668 (-0.1)	0.668 (-0.1)	0.668 (-0.1)	0.668 (-0.1)	0.668 (-0.1)	0.668 (-0.1)	0.667 (-0.1)	0.668 (-0.1)
$A^{0,b}_{FB}$	0.1033 (+2.6)	0.1035 (+2.7)	0.1031 (+2.4)	0.1035 (+2.7)	0.1033 (+2.6)	0.1034 (+2.6)	0.1030 (+2.4)	0.1032 (+2.5)
$A^{0,c}_{FB}$	0.0739 (+0.9)	0.0739 (+0.9)	0.0737 (+0.9)	0.0738 (+0.9)	0.0738 (+0.9)	0.0738 (+0.9)	0.0736 (+0.8)	0.0736 (+0.9)
$R^{0,b}$	0.21583 (-0.7)	0.21587 (-0.6)	0.21585 (-0.7)	0.21587 (-0.6)	0.21583 (-0.7)	0.21587 (-0.6)	0.21586 (-0.7)	0.21587 (-0.6)
$R^{0,c}$	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)	0.1722 (0.0)

Table 6.3: The best fit values of the observables, and their pulls (calculated as the fit value minus the measured value, divided by the experimental uncertainty). Note that to generate this table, we have explicitly set U = 0.

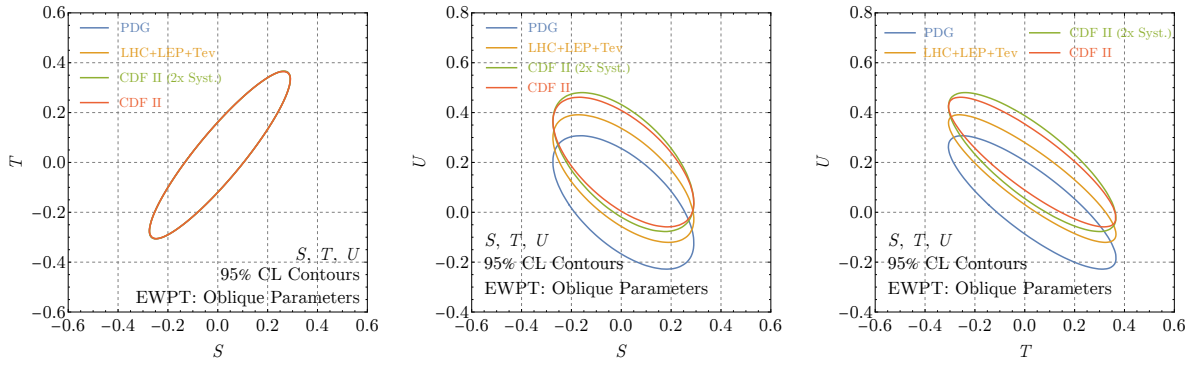


Figure 6.2: Similar to Fig. 6.1, but now also including U in the global fit. We show the 95% CL preferred region of all oblique parameters in the S − T plane (left), S − U plane (center), and T − U plane (right). In each plot, we marginalize over the third parameter. We find that when we include U in the fit, S and T remain nearly centered about 0, whereas U has a notable positive shift. Getting such large values of U are quite challenging in perturbative models. Note that only one ellipse is visible in the S − T plane as the other contours overlap completely.

The W decay width is not measured to nearly as high precision as M_W, so the observed discrepancy in the W mass at CDF II [11] can be accommodated in the SM electroweak fit by setting $U \approx 0.11$, without affecting any of the other observables as compared to the SM fit. These other unaffected observables include S and T which take their SM values as well as the existing tension in the EW fit from the forward-backward asymmetry.

To illustrate this in more detail, we perform the fit as described above but also allow the U parameter to float, in addition to the S and T parameters and the free observables. We then plot 95% confidence intervals for pairs of the electroweak precision parameters while marginalizing over the third parameter and the other free parameters. The results are shown in Fig. 6.2.

We see that, when marginalizing over U, the 95% CL preferred range of S and T with the new CDF measurement of M_W is quite similar to the allowed region using the smaller value of M_W. Instead, the U parameter is inflated to account for the shift in mass.

The difficulty in this interpretation is that a large value of U is challenging to generate in perturbative models, because, as mentioned in §6.2, U corresponds to a dimension-8 operator [1017], and a value of $\mathcal{O}(0.1)$ indicates scales of order few 100 GeV for tree-level models, and $\ll 100$ GeV for particles contributing in loops. As the U parameter violates custodial symmetry, it is difficult to imagine a model that generates a large, nonzero value of U without also generating large values of T. We therefore do not attempt to construct models generating large values of U. In the concrete BSM models we consider in the next section, we will ignore the (subleading) U-dependence altogether.

6.4 Implications for BSM Models

From the results of our electroweak fit shown in §6.3, we see that the value of M_W can dramatically change the preferred values of the oblique parameters. While the 95% CL region fitting with PDG measurements is nearly centered around the predicted SM values of $(S, T, U) = (0, 0, 0)$, the updated value of M_W shifts this region to positive $\mathcal{O}(0.1)$ values of oblique parameters (see Figs. 6.1, 6.2).

In this section we explore various tree-level and loop-level contributions to the oblique parameters from simple models, and assess their viability. For clarity, we focus on the scenario of M_W equal to the updated world average from Tevatron, LEP, and LHC measurements (the second scenario in Eq. (6.2)).

It is first worthwhile to estimate the scale of new physics implied by $\mathcal{O}(0.1)$ values of S and T. Comparing to the dimension-6 operators defined in (6.4), we see that for tree-level matching with perturbative couplings, these operators can be generated by new physics at a scale $\Lambda \sim$ TeV. If the new physics arises in loops, on the other hand, the loop factor suppression implies a scale closer to $\mathcal{O}(100\,\text{GeV})$. We will examine this matching in both scenarios, first considering minimal extensions to the SM that can be integrated out at tree-level, such as an additional scalar, then consider a one-loop example with new singlet-doublet fermion pairs. Note that, as indicated in Fig. 6.1, it is important for these models to shift T to positive values to be consistent with our electroweak fit.

6.4.1 Tree-Level Models

Here we consider models that lead to corrections to the oblique parameters at tree level. Given the results of the fits shown in Fig. 6.1, we are particularly interested in models that can accommodate large positive values of S and T.

The simplest examples of models leading to oblique parameter corrections are new scalars. An $SU(2)_L$ singlet scalar leads only to an overall rescaling of the Higgs couplings that do not affect S and T or shifts in the Higgs self-coupling. Models with extra $SU(2)_L$ doublet scalars, such as a 2HDM [1018], can affect the Higgs couplings to the gauge bosons, but these deviations are proportional to $\cos^2(\beta - \alpha)$, the square of the alignment parameter, which from an effective field theory perspective is dimension-8, and therefore cannot affect the oblique parameters S and T, which are dimension-6.

An $SU(2)_L$ triplet scalar ϕ^a, however, leads to more interesting possibilities [1019].[6] Such a triplet can have interactions with the SM Higgs $\sim \phi^a H^\dagger \sigma^a H$. After electroweak symmetry breaking, this interaction leads to a small vacuum expectation value for the scalar triplet, which shifts the mass of the W bosons without changing the mass of the Z, therefore offering a possibility of resolving the tension between the CDF

[6] We thank Matthew Strassler for bringing this model to our attention.

measurement of M_W and the SM expectation.

For concreteness, we will consider a real scalar $SU(2)_L$ triplet ϕ^a with $Y = 0$ which we will refer to as a swino; see Refs. [1020, 1021] for possible UV-completions and Ref. [1022] for a recent study of swino phenomenology. The Lagrangian takes the form

$$\mathcal{L} \supset \frac{1}{2} D_\mu \phi^a D^\mu \phi^a - \frac{1}{2} M_T^2 \phi^a \phi^a + \kappa \phi^a H^\dagger \sigma^a H - \eta H^\dagger H \phi^a \phi^a. \tag{6.8}$$

The oblique parameters have been worked out in Ref. [1023, 1024], where they include the matching up to one-loop order. At tree-level, the contribution to S from scalar triplets vanishes. The $Y = 0$ swino does, on the other hand, lead to a contribution to the T parameter given by

$$T = \frac{v^2}{\alpha} \frac{\kappa^2}{M_T^4} \tag{6.9}$$

This contribution is *positive* for any value of κ and can naturally explain the observed discrepancy in M_W measurement.

One can also consider scalar triplets with $Y = 1$, but these lead to the wrong sign for T at tree level. At one loop, both $Y = 0$ and $Y = 1$ triplets lead to additional corrections to both S and T, which can be potentially large and positive, depending on the quartic couplings to the Higgs. We leave a more detailed study of these possibilities to future works.

In Fig. 6.3, we show the band of values of κ and M_T that are compatible with the electroweak fit with the combined value of M_W at 95% CL. As is clear from the scaling in (6.9), the necessary large value of T can be achieved even for large triplet masses. Requiring $\kappa/M_T \lesssim 1$, the triplet mass can be up to $\mathcal{O}$(few TeV), evading any potential collider bounds.

6.4.2 Singlet-Doublet Model

We now shift our attention to another simple extension of the SM, the $SU(2)_L$ singlet-doublet fermion model. Unlike the previous discussion, the contribution of this model to electroweak precision measurements first occurs at loop level. The model includes N_f families of a singlet Majorana and doublet Dirac fermion charged under the electroweak sector [4, 913–923].[7] This is a minimal, UV complete, anomaly-free construction which can generate a Higgs portal coupling, with the added benefit that such a setup can be readily embedded inside supersymmetric extensions of the SM. The $SU(2)_L$ doublet has hypercharge 1/2 and is composed of two left-handed Weyl fermions ψ_2 and $\tilde{\psi}_2$. The Lagrangian is

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_{N_f} \mathcal{L}_{\text{kinetic}} - m_2 \psi_2 \cdot \tilde{\psi}_2 - \frac{m_1}{2} \psi_1 \psi_1 + y\, e^{i\delta_{CP}/2} \psi_1 H^\dagger \psi_2 - \tilde{y}\, e^{i\delta_{CP}/2} \psi_1 H \cdot \tilde{\psi}_2 + \text{h.c.} \tag{6.10}$$

[7] For simplicity, we consider the scenario where these fermions do not mix with each other, but in principle mixing could lead to richer phenomenology.

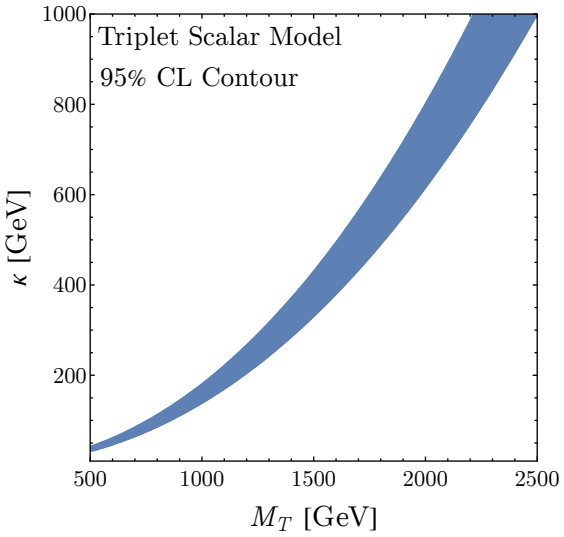


Figure 6.3: The 95% CL band using the results from an electroweak fit with the updated world average M_W measurement in the $M_T - \kappa$ plane of the triplet scalar model. We find viable parameter space for $\mathcal{O}$(TeV) swino masses that can potentially be probed with future high-energy colliders.

This Lagrangian has a physical CP-violating phase, as we have four new parameters and three new fields. Since S and T are CP-even observables, we set $\delta_{CP} = 0$ in this analysis for simplicity. However, this model is also interesting with nonzero values of δ_{CP} as it can potentially explain the Galactic Center Excess (see §5 or [4] for details). Additionally, because of the Yukawa terms, there is mass mixing between the fermions and the ψ_i fields are not the propagating degrees of freedom. We call attention to this point because the mass of the lightest propagating fermion is relevant for Higgs (and Z) decay constraints, which require $M_\chi > M_h/2$. The singlet-doublet model contributes to the S and T parameters at loop-level with the new fermions running in the loop. We provide a quick summary of this calculation in Appendix C.

The size of the contributions to S and T in this model scales linearly with the number of new fermion generations, N_f. We can only get a nonzero T value when the custodial symmetry is broken, i.e. $y \neq \tilde{y}$. Because of this, the value of T depends on the difference $y - \tilde{y}$, so a relatively large difference between y and $\tilde{y}$ is required to generate a sufficiently large T. Furthermore, S and T both decrease as m_2 or m_1 increase, making it difficult to reach values consistent with both the updated electroweak fit and existing experimental constraints without including multiple generations of new fermions.

In Fig. 6.4 we plot 95% CL region from our electroweak fit using the updated world average as a function of the new fermion mass parameters m_1 and m_2 to get a benchmark value of the couplings. Lower values of m_1 and m_2 are strongly constrained by a host of different measurements (including LEP bounds on charged fermions, Higgs and invisible Z decays, and direct searches for light fermions carrying electroweak charge). In the left panel of the figure we consider the model with only one generation of new fermions. We find that the contribution to S and T is only sufficiently large to fit M_W with the updated CDF II measurement in a

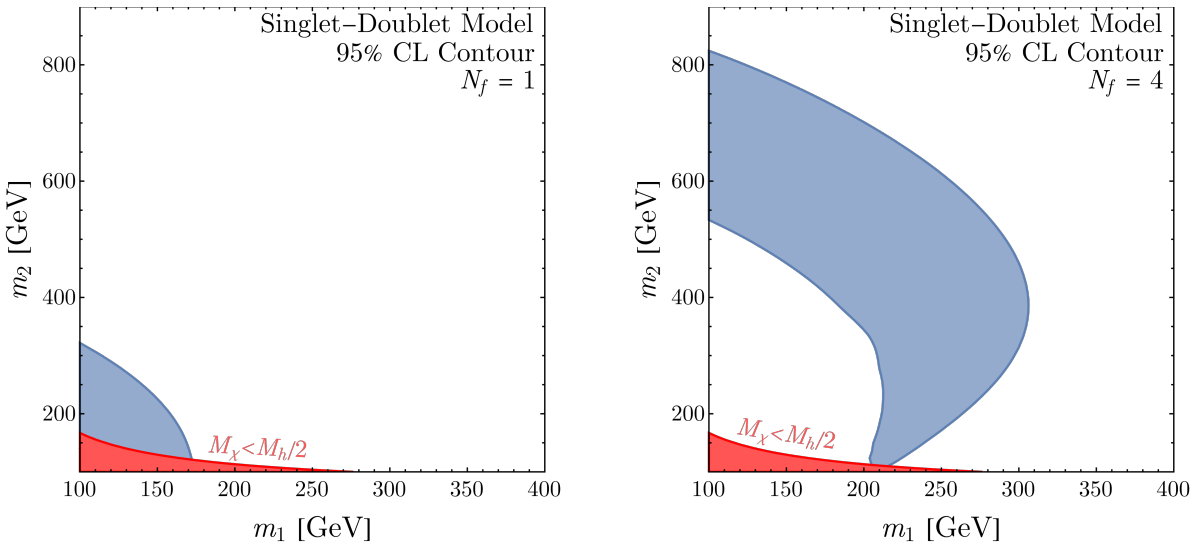


Figure 6.4: The regions of singlet-doublet parameter space that are inside the 95% CL region from the electroweak fit including the S and T parameters for the updated world average M_W value. The couplings are set to benchmark values ($y = 0.1$, $\tilde{y} = 1$, $\delta_{CP} = 0$), and we consider N_f generations of new fermions, where $N_f = 1$ ($N_f = 4$) on the left (right). We consider $y \neq \tilde{y}$ since nonzero T depends on the custodial symmetry breaking $y - \tilde{y}$. Relevant constraints on the model are briefly discussed in the text; in particular, direct LHC searches can potentially rule out most of the blue band for $N_f = 1$ and probe much of the $N_f = 4$ allowed region.

small corner of the parameter space; direct searches at LHC strongly constrain this range of masses.

In the right panel of Fig. 6.4 we show the contribution of the model to the oblique parameters with $N_f = 4$. We now find a larger range of masses that give rise to M_W values within 95% confidence of the global average measurement. Direct LHC searches can again rule out some of this parameter space, but there is still viable parameter space in the range of masses shown in the figure, specifically in the limit of degenerate masses or at high values of m_2. A more thorough exploration of the viable parameter space (including with other values of y and $\tilde{y}$) is left for future work.

6.5 Discussion

In this chapter we studied the effect of the recent M_W measurement at CDF II on global fits of electroweak precision observables and the implications for physics beyond the SM. By performing a standard χ^2 fit over SM parameters as well as the oblique parameters S, T, and U, we explored the efficacy of a variety of models for generating an upward shift in the M_W mass. After combining all M_W measurements at the Tevatron, LEP, and the LHC, there exists a significant discrepancy with SM predictions.

The results of our fit suggest that new physics models that contribute to S and, more substantially, a positive T are potential candidates to explain the anomaly. While we considered a global fit also including U, the results did not have a natural model-building interpretation. Of the models we consider, we find

that a singlet scalar extension of SM and a 2HDM model fail to yield S and T contributions consistent with our fit. However, the swino model was markedly successful since it generated positive $\mathcal{O}(0.1)$ values of T in unconstrained regions of parameter space. Viable triplet mass values were found to be near or above the TeV scale, which can evade current experimental bounds while giving rise to interesting signatures in future high energy colliders such as FCC-hh or muon colliders. We leave a detailed study of such signals for future work. Additionally, we found some success with a singlet-doublet fermion model when considering multiple generations.

As previously mentioned, there are other anomalies in the SM that could arise from discrepant electroweak precision measurements, such as the anomalous magnetic moment of the muon $g-2$. It was pointed out in Ref. [1025–1027] that the existing discrepancy between the theoretical and measured values of $(g-2)_\mu$ can be absorbed in a shift to the hadronic vacuum polarization contribution by changing $\Delta\alpha^{(5)}_{\rm had}$, at the cost of increasing the tension in the SM electroweak fit, particularly by *decreasing* the preferred value of M_W. It is of high importance to explore if the necessary change in the fit to ameliorate the $(g-2)_\mu$ discrepancy can be accommodated by the BSM effects of interest for the W mass measurement as studied in Refs. [1028, 1029], or if something much more exotic is required.

Finally, we would like to call attention to the fact that tension arising from the global SM electroweak fit is not unique to the W boson mass. For example, significant deviations from the SM have been evident in the forward-backward asymmetry observable at LEP for many years [1003], and there are numerous attempts at explaining this with BSM physics (e.g. Refs. [1030], among others). This further motivates future study of how potential new physics affects electroweak precision observables.

These results can be interpreted as new *oblique* signs of BSM appearing around the TeV scale. In light of this new measurement, further experimental results, including improvement to measurement of M_W at the LHC or future colliders, are strongly motivated.

Note added: As the paper this chapter is based on was being finalized, Refs. [1028, 1029, 1031–1035] appeared, which also consider the implications of the recent M_W measurement. In particular, Refs. [1032, 1034] similarly consider an electroweak fit to evaluate possibility of new physics contributions to the W mass.

7

Wrinkles in the Froggatt–Nielsen Mechanism

7.1 Introduction

Flavor physics has been a harbinger of physics beyond the Standard Model (BSM) at various points in time, from predicting the existence of the charm quark [1036, 1037] to estimating the mass of the top quark [1038–1041] long before its discovery at the Tevatron [1042, 1043]. Precision experiments, in particular, help establish or find violations of the Standard Model (SM) symmetry structures, and prove to be noteworthy indirect probes of new physics whose mass scale lies beyond the reach of direct collider searches; see [1044, 1045] for reviews of many such experiments.

A primary goal of flavor physics is to understand the appearance of large hierarchies in the masses and mixing angles of the SM fermions. The two most popular solutions to this puzzle are (i) the Froggatt–Nielsen (FN) mechanism and its variations [1046–1050], and (ii) extra dimensional models where an $\mathcal{O}(1)$ difference in the bulk masses of fermions gives rise to an exponential hierarchy between the observed masses in the IR [1051–1056]. Other notable possibilities include generating the mass hierarchy via running to the IR in extensions of the SM with scale invariant sectors in the UV [1057], or radiatively generating the Yukawas with the hierarchy governed by powers of the loop expansion parameter [1058–1061]. A review of these and other dynamical solutions to the flavor puzzle can be found in [1062–1064]. In what follows, we focus our attention on the FN mechanism.

In the FN mechanism, the hierarchies in the SM fermion sector arise as different powers of a small expansion parameter. This expansion parameter is given by the ratio of the vacuum expectation value (vev) of a scalar field, known as the flavon, over a heavy mass scale. The SM Yukawa couplings are generated by non-renormalizable operators involving the chiral SM fermions, the Higgs, and the flavon. The dimensionality of these operators—and the resulting power of the expansion parameter that appears—is dictated by the charges of the SM fermions under a new Abelian horizontal symmetry, $\mathrm{U}(1)_{\mathrm{H}}$, which is broken by the flavon. As we will discuss, there is additional freedom in the assignment of these charges that was overlooked in Ref. [1046]. In the original FN paper, it was supposed that these irrelevant operators are generated by "chains" including heavy vector-like matter, also charged under $\mathrm{U}(1)_{\mathrm{H}}$. A number of variations to this model have been proposed, including "inverted" models [1065], where the flavon vev is larger than the heavy mass scale.

One of the drawbacks of invoking the FN mechanism is that the new dynamics responsible for the SM hierarchies can exist at scales far above the weak scale, beyond the reach of direct experimental probes. Nevertheless, given the other shortcomings of the SM—the electroweak hierarchy problem in particular—there is ample reason to expect new physics at or near the TeV scale. If the new physics is *flavorful* (i.e., it involves non-universal couplings to SM matter fields), its flavor structure may also be dictated by the FN dynamics. This argument can also be run in reverse: given the stringent constraints from precision measurements of the SM, for new physics to exist at the TeV scale it must either be flavor-blind or incorporate some symmetry arguments to suppress flavor-violation [1066, 1067]. This reasoning is familiar in the supersymmetric context, where it is understood that squarks must either be degenerate or flavor-aligned [1068].

In this light, it is clearly worthwhile to study the application of the FN mechanism to the couplings of new BSM fields. This is particularly true when flavorful new physics is invoked to explain potential discrepancies between experimental results and the SM expectations: should one of these discrepancies become an unambiguous signal of new physics, we might glean information about the dynamics associated with flavor in the UV. This approach was advocated in [1069–1071], and we will review it extensively in this work. An immediate consequence of this framework is that many different experimental observables become correlated. These correlations challenge some of the simplest solutions to various flavor anomalies, as the couplings and masses required to explain the discrepancy violate bounds set by other observables such as lepton flavor violating (LFV) processes or flavor changing neutral currents.

The goal of this work is to explore how these considerations can change if the FN setup is amended with additional symmetries or structure in the UV. We do this by working in an effective field theory (EFT) framework, including the SM and new BSM fields, with their couplings to fermions treated as spurions

under the $U(3)^5$ flavor symmetry of the SM. In this framework, we can introduce controlled deviations from the size of these spurions dictated by the horizontal charges. We refer to these deviations as *wrinkles*, since they appear in the UV as changes in the length of the chain diagrams responsible for the Yukawas in the IR. Wrinkles can exist in SM or BSM spurions, and allow us to relax the correlations between different observables, permitting sizable new physics contributions to some observables while satisfying other experimental bounds.

Importantly, while wrinkles allow for much greater flexibility in the couplings of BSM fields to SM fermions, this flexibility is not without bound. If the effective theory is to be faithfully embedded in the FN mechanism, radiative corrections must not spoil the relationship between the couplings in the IR and the non-renormalizable operators in the UV. This requirement has been previously formulated as a consistency condition in the context of minimal flavor violation EFTs [1069] (see also Ref. [1070]). While these conditions are trivially satisfied in ordinary FN models, we show that they put meaningful bounds on wrinkled FN setups.

Since this wrinkled FN setup can be applied to any new physics, we will illustrate its application in an example, where the SM is extended by a single leptoquark, denoted S_1 in the nomenclature of Ref. [1072]. See [1071, 1073–1075] for previous discussions of leptoquark models with horizontal symmetries. We will use this leptoquark to enhance the branching ratio of $B^+ \to K^+ \bar{\nu}\nu$, which currently shows a small discrepancy with SM predictions [698] and will be precisely measured at the Belle II experiment in the coming years. Without wrinkles, the charges and masses required to generate a large $B^+ \to K^+ \bar{\nu}\nu$ signal also imply the existence of large signals in other correlated observables, such as LFV decays or leptonic meson decays. We will show a simple example where a wrinkled FN setup evades these bounds while satisfying the consistency conditions alluded to above. As we will see, the bound on the wrinkles implies other correlated signals are generated near detection thresholds in this example, and could potentially be seen in the near future.

In the coming years, troves of new data from colliders and small-scale experiments searching for signs of flavorful new physics will begin stress-testing the delicate flavor structure of the SM. Given the substantial motivation for BSM physics, this structure could break and potentially start showing signs of deviations from the SM expectation. In preparation for such deviations, it is timely to develop new model-building tools which enable embedding their solutions in UV complete frameworks. Wrinkles in an FN ansatz are a flexible, bottom-up tool that allow for a broader exploration of the complementarity of different flavor probes, while reliably parameterizing more sophisticated UV models of flavor. As such, they present a natural setup to search for a consistent IR picture of new physics with flavor, should any deviations from the SM come to light.

This chapter is organised as follows: in §7.2, we review the FN mechanism, its solution to the flavor

hierarchy problem in SM and how it furnishes suitable ansätze for couplings arising from new BSM physics. Next, in §7.3, we introduce the concept of wrinkles for the FN mechanism, discuss constraints on them, and provide examples for how they can arise from UV complete models. In §7.4, we provide a concrete example of applying wrinkles to the S_1 scalar leptoquark embedded in a FN model. We demonstrate that wrinkles allow one to simultaneously explain bounds on BSM physics from current precision flavor observables, while also retaining predictive power for potential future measurements. We conclude in §7.5. Appendix E.1 provides details about bounds on wrinkles arising from consistency conditions. Appendix E.2 provides details on flavor observable computations in the S_1 leptoquark model.

7.2 Froggatt–Nielsen and BSM Physics

The lepton and quark Yukawas and mixing angles present a clear generational hierarchy, with the charged particle masses ranging over five orders of magnitude. This hierarchy implores an explanation in the UV. Searches for flavorful new physics are carried out in pursuit of such an explanation. Hence, if any anomaly emerges in these experiments, it is well-motivated to embed its BSM solutions within UV models that explain the flavor hierarchy as well.

The FN mechanism [1046] provides a four-dimensional, field-theoretic explanation for this hierarchy, replacing the small dimensionless parameters with a power counting in powers of an inverse mass scale, fixed by a symmetry. In this section, we review how this mechanism can explain the parameters in the SM matter sector, with an emphasis on the EFT point of view. We will then discuss how this perspective can naturally be extended to BSM physics.

7.2.1 Review of the Froggatt–Nielsen Mechanism

The basic idea of the FN mechanism is to introduce a horizontal symmetry, $U(1)_H$, under which different generations of the SM fermions have different charges. The horizontal symmetry is assumed to be spontaneously broken by the vacuum expectation value of a SM singlet scalar field, ϕ—the *flavon*. Assuming our EFT is valid up to some cutoff scale M, we are led to a natural expansion parameter $\lambda = \langle\phi\rangle/M$, which appears in non-renormalizable operators involving the SM fermions. Later on, we will associate this scale M with the mass of new heavy fermions. Without loss of generality, we take the SM Higgs to be neutral under $U(1)_H$ and take the flavon charge to be -1.

At scales just below the cutoff, the lowest dimension operators involving the SM fermions and the Higgs take the form

$$\mathcal{L} \supset r^u_{ij}\frac{\phi^{(\dagger)m_{ij}}}{M^{m_{ij}}}Q_i H\bar{u}_j + r^d_{ij}\frac{\phi^{(\dagger)n_{ij}}}{M^{n_{ij}}}Q_i H^c\bar{d}_j + r^e_{ij}\frac{\phi^{(\dagger)l_{ij}}}{M^{l_{ij}}}L_i H^c\bar{e}_j + \text{h.c.} \tag{7.1}$$

where $(Q_i,\ \bar{u}_i,\ \bar{d}_i,\ L_i,\ \bar{e}_i)$ are different SM fermions, subscripts on fermion fields refer to different generations, r_{ij} are $\mathcal{O}(1)$ couplings,

$$m_{ij} = \big|[Q_i] + [\bar{u}_j]\big|, \qquad n_{ij} = \big|[Q_i] + [\bar{d}_j]\big|, \qquad l_{ij} = \big|[L_i] + [\bar{e}_j]\big|, \tag{7.2}$$

and the square brackets indicate the $U(1)_H$ charge. The hermitian conjugate on ϕ appears if the sum of charges inside the absolute value is negative. At energies below $\langle\phi\rangle$, these operators appear as the Yukawa couplings of the SM Higgs, with the coupling matrices given by

$$Y^{ij}_{Q\bar{u}} = r^u_{ij}\frac{\langle\phi^{(\dagger)}\rangle^{m_{ij}}}{M^{m_{ij}}} \sim \lambda^{m_{ij}}, \qquad Y^{ij}_{Q\bar{d}} = r^d_{ij}\frac{\langle\phi^{(\dagger)}\rangle^{n_{ij}}}{M^{n_{ij}}} \sim \lambda^{n_{ij}}, \qquad Y^{ij}_{L\bar{e}} = r^e_{ij}\frac{\langle\phi^{(\dagger)}\rangle^{l_{ij}}}{M^{l_{ij}}} \sim \lambda^{l_{ij}}. \tag{7.3}$$

This scaling implies that even modest differences in horizontal charges give rise to exponential hierarchies in Yukawa couplings. To connect with the observed flavor structure of the SM, we identify λ with the Cabbibo angle, ~ 0.2, so that the CKM matrix hierarchies follow naturally from the Wolfenstein parameterization [1076]. We refer to this setup as vanilla FN.

At the $\mathcal{O}(1)$ level, the masses and mixing angles are

$$V_{ij} \sim \lambda^{\big|[Q_i] - [Q_j]\big|}, \quad U_{ij} \sim \lambda^{\big|[L_i] - [L_j]\big|}, \tag{7.4}$$

$$m^u_i \sim \lambda^{\big|[Q_i] + [\bar{u}_i]\big|}, \quad m^d_i \sim \lambda^{\big|[Q_i] + [\bar{d}_i]\big|}, \quad m^l_i \sim \lambda^{\big|[L_i] + [\bar{e}_i]\big|}, \tag{7.5}$$

where V (U) is the CKM [1077, 1078] (PMNS [1079, 1080]) matrix.

The most general horizontal charge-assignment that gives rise to the observed structure of the CKM and PMNS matrices and SM fermion masses is given in Table 7.1.[1] We have the overall freedom to shift the charges of all quarks (leptons) by the same amounts q_0 (l_0), respectively. Once these shifts are chosen, the CKM and PMNS structure constrain the other LH quarks' and leptons' charges. As indicated in Eq. (7.4), these mixing matrices only fix the absolute value of the difference between charges, hence the freedom in choosing $X, Y = \pm 1$ in the table. The appearance of X, Y in multiple entries captures the correlation between those charges. To find the RH fermion charges we use the measured values of masses in the SM. As in the case of mixing, Eq. (7.5) only fixes the absolute value of the charge difference between LH and RH fermions, leaving the sign undetermined. We choose the signs so that the eigenvector associated with the heaviest (lightest) mass eigenstate has the biggest overlap with the third (first) generation for each type of fermion. To check this, we generated 10000 mass matrices for each charge assignment, drawing new random numbers $r^{u,d,e}_{ij} \in (0.2, 1)$ for each test. For every charge assignment, we confirmed that a substantial

[1] In general, shifts of ± 1 in most of these charges can be tolerated when random $\mathcal{O}(1)$ Yukawa couplings in the UV model are taken into account and the fact that the expansion parameter λ is not particularly small is considered. The anarchic structure of the PMNS matrix, in particular, leaves room for such small changes in the charges; see [1081, 1082] for further exploration of these shifts.

	Gen. 1	Gen. 2	Gen. 3
Q	$-q_0 - 3X$	$-q_0 - 2X$	$-q_0$
$\bar{u}$	$q_0 + 3X \pm 7$	$q_0 - X$	q_0
$\bar{d}$	$q_0 + 3X \pm 6$	$q_0 - 3X$	$q_0 - 2X$
L	$l_0 + Y$	l_0	l_0
$\bar{e}$	$-l_0 - Y \pm 8$	$-l_0 + 5Y$	$-l_0 + 3Y$

Table 7.1: The most general horizontal charge assignment that explains the SM masses and mixings in FN with $\lambda \sim 0.2$. q_0 and l_0 denote general shifts in quark and lepton charges, respectively, that leave the IR masses and mixings unchanged. $X, Y = \pm 1$ denote the correlations between different charges that are required by the CKM and PMNS matrices. For every value of (q_0, l_0), we have 2^5 choices for the charge assignments. In supersymmetric theories, holomorphy sets $X = -Y = -1$ and picks the positive sign for first generation RH fermions.

fraction of trials yield the correct mixing patterns and mass eigenvalues that are within a factor of two of the experimentally-measured values.

In the original FN proposal, it was assumed that the charges of all five types of fermions (Q, $\bar{u}$, $\bar{d}$, L, and $\bar{e}$) are ordered monotonically between different generations. Table 7.1 indicates that, while some correlations between LH and RH fermions of the second and third generation (captured by X, Y) are needed to generate the correct mass eigenstates, the monotonicity condition can be removed for first generation RH fermions without distorting the model's prediction for SM masses. This manifests itself as a binary choice in the charge of each first generation RH fermion ($\bar{u}$, $\bar{d}$, $\bar{e}$).

It is also popular to consider supersymmetric variations of FN models. In the supersymmetric case, holomorphy of the superpotential forbids terms with $\phi^\dagger$ instead of ϕ [1047, 1048]. This eliminates a great deal of the freedom in charge assignments tabulated in Table 7.1. Specifically, it fixes $X = -Y = -1$ and picks the positive sign for first generation RH fermions, leaving only the separate overall shifts in the quark and lepton charges, q_0 and l_0. It also enforces the monotonicity of the horizontal charges across different generations. However, since we do not explore the supersymmetric case in detail in the rest of this chapter, we do not need to enforce these constraints.

The simplest UV completion of this effective theory (and the one imagined by Froggatt and Nielsen [1046]) is to introduce a set of vector-like fermions F with mass M that live in an SM representation permitting Yukawa couplings between the Higgs and SM fermions. We assume the existence of heavy fermions with all horizontal charges necessary to complete the SM Yukawas with Yukawa couplings to the flavon $\sim \phi F \bar{F}'$. The flavon Yukawa couplings are assumed to be $\mathcal{O}(1)$, leading to the effective theory in Eq. (7.1) with $\mathcal{O}(1)$ Wilson coefficients denoted by r_{ij}.

As an example, the up-type Yukawa couplings can be generated by "chain" diagrams such as those shown in Fig. 7.1. The top Yukawa arises at the renormalizable level, but the suppressed couplings arise

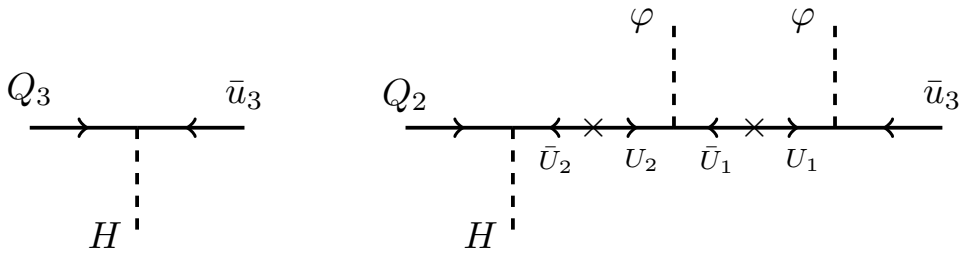


Figure 7.1: Example diagrams leading to the effective operators for the up-type Yukawa couplings with vector-like heavy fermions U and $\bar{\text{U}}$, where $\bar{\text{U}}$ has the same SM quantum numbers as $\bar{\text{u}}$. The subscripts on the U fields refer to the horizontal charge of U, and we have taken charges from Table 7.1 with $\text{X} = -1$ and $\text{q}_0 = 0$.

by introducing the vector-like pair U and $\bar{\text{U}}$, where $\bar{\text{U}}$ has the same quantum numbers under the SM gauge groups as $\bar{\text{u}}$. The subscripts indicate the $\text{U}(1)_\text{H}$ charge of U. For instance, the chain shown on the right side of Fig. 7.1 gives rise to a λ^2 suppression in the coupling of $\text{Q}_2\bar{\text{u}}_3$. If there exist heavy fermions with SM charges similar to Q and the correct horizontal charges, chain diagrams with the Higgs and flavon insertions interchanged will contribute as well. Similar chains give rise to the Yukawa couplings for other SM fermions.

Models of FN constructions with additional symmetries, multiple expansion parameters, or expansion parameters that are allowed to freely vary have also been developed in the literature, e.g. see [1047, 1048, 1081, 1083–1085]. For simplicity, however, in this work we focus on FN setups with only one expansion parameter, which we identify with the Cabbibo angle, and develop a systematic way for small deviations from them. We can straightforwardly generalize our discussions below to more baroque FN setups.

As a final note, in a UV complete model, quantum gravity considerations require that the horizontal symmetry be embedded in a gauge symmetry [602], which in turn demands the cancellation of all its anomalies.[2] We have checked that the general charge assignment of Table 7.1 can not cancel all gauge anomalies in the typical FN UV completion, see also Ref. [1091] for a similar conclusion. This conclusion is also corroborated by [1092, 1093], which deduce that the general charge assignments that can explain the SM Yukawa hierarchy can not be anomaly-free by studying general extensions of the SM with a new anomaly-free U(1) gauge group. As a result, in such a construction one should resort to either introducing new heavy chiral fermions (and subsequently extending the scalar sector so as to generate a mass for these fermions) or the Green-Schwarz mechanism to cancel anomalies [1094]. We leave further investigations of anomaly cancellation for future work.

[2]The lack of evidence for the (pseudo-)Nambu–Goldstone boson associated with the spontaneous breaking of the horizontal symmetry is also often used as motivation for gauging it. However, models with a potentially viable Goldstone exist. See [1086–1091] for examples where the Goldstone is identified with the QCD axion.

7.2.2 Froggatt–Nielsen and Flavorful New Physics

When introducing new physics, some assumptions must be made about the couplings of SM fields to new particles. These couplings are generically non-universal unless governed by additional structure such as new gauge symmetries. Given the hierarchies that exist in the SM fermion couplings, it is a priori unclear what a "natural" size for such non-universal couplings should be. However, if one assumes a UV explanation of the flavor hierarchy such as the FN mechanism, there is a natural ansatz for the new physics couplings as well.

The phenomenological significance of such an ansatz lies in the fact that it correlates predictions of a BSM model for various flavorful observables in the IR. Thus, depending on the ansatz, a model built for explaining a discrepancy in the data will give rise to correlated signals in other constraining observables. For instance, any solutions of the $(g-2)_\mu$ anomaly with non-minimal flavor ansatz gives rise to unacceptably large contributions to various LFV decays, especially $\tau \to \mu\gamma$.

To better understand such ansätze, it is useful to organize our thinking in terms of the global flavor symmetry of the SM:

$$G_{\text{flavor}} = SU(3)_Q \times SU(3)_u \times SU(3)_d \times SU(3)_L \times SU(3)_e \times U(1)^5, \tag{7.6}$$

where three of the U(1) factors can be identified with hypercharge, baryon number and lepton number. This symmetry acts on the generation indices of the chiral matter in the SM, with the unbarred (barred) fields transforming as triplets (anti-triplets), respectively. The symmetry is broken explicitly by the Yukawa matrices, but formal invariance under G_{flavor} can be restored if we promote the Yukawas to transform as spurions:

$$Y_{Q\bar{u}} \sim (\bar{3}_Q, 3_u), \qquad Y_{Q\bar{d}} \sim (\bar{3}_Q, 3_d), \qquad Y_{L\bar{e}} \sim (\bar{3}_L, 3_e). \tag{7.7}$$

This formalism can be extended in a straightforward way to new physics with any new spurions of G_{flavor} [1069–1071]. New fields are taken to be singlets of the $SU(3)^5$ part of the SM flavor group, and their couplings to SM fermions then have definite transformation properties under G_{flavor}.

As an example, consider the scalar leptoquark S_1, a color anti-fundamental with hypercharge $Y = 1/3$. This allows for the renormalizable couplings to SM fields,[3]

$$\mathcal{L} \supset -\Delta^{ij}_{QL}\varepsilon^{ab}S_1 Q_{bi} L_{aj} - \Delta^{ij}_{\bar{u}\bar{e}} S_1^\dagger \bar{u}_i \bar{e}_j + \text{h.c.}, \tag{7.8}$$

where the spinor indices are implicit, a, b are $SU(2)_L$ fundamental indices, $\varepsilon^{12} = +1$, and (i, j) are flavor

[3]The SM gauge symmetries also permit the couplings $S_1\bar{u}\bar{d}$ and $S_1 Q^\dagger Q^\dagger$, which lead to proton decay. We can forbid these couplings by enforcing conservation of baryon number and endowing the leptoquark with a baryon number of $-1/3$, or by potentially gauging some discrete subgroup. Therefore, in the rest of this work, we ignore these couplings. We note that the "wrinkles" introduced in §7.3 cannot entirely alleviate the proton decay constraint, necessitating a symmetry-based explanation.

indices. The $\Delta_{\rm QL}$ coupling also appears in R-parity violating supersymmetric models, where S_1 is identified with a down squark; these models have $\Delta_{\bar{u}\bar{e}} = 0$ [1095]. The new Yukawa couplings $\Delta_{\rm QL}$ and $\Delta_{\bar{u}\bar{e}}$ transform as

$$\Delta_{\rm QL} \sim (\bar{3}_{\rm Q}, \bar{3}_{\rm L}), \qquad \Delta_{\bar{u}\bar{e}} \sim (3_{\rm u}, 3_{\rm e}). \tag{7.9}$$

In the absence of any flavor ansatz, the matrices $\Delta_{\rm QL}$ and $\Delta_{\bar{u}\bar{e}}$ are arbitrary 3×3 complex matrices. However, when embedded in a vanilla FN setup, and assuming the S_1 leptoquark is neutral under $U(1)_{\rm H}$, we find an ansatz for the hierarchies present in the spurions $\Delta_{\rm QL}$ and $\Delta_{\bar{u}\bar{e}}$. In analogy with Eq. (7.3), we find:

$$\Delta^{ij}_{\rm QL} \sim \lambda^{\left|[{\rm Q}_i] + [{\rm L}_j]\right|}, \qquad \Delta^{ij}_{\bar{u}\bar{e}} \sim \lambda^{\left|[\bar{e}_j] + [\bar{u}_i]\right|}. \tag{7.10}$$

Put differently, the SM charges and flavor symmetries are enough to determine how the new S_1 field should be embedded in the effective theory below M. The power counting of the effective theory then dictates that the expected FN scaling above holds, up to the $\mathcal{O}(1)$ Wilson coefficients of the effective theory (analogous to the r_{ij} in Eq. (7.1)). This ansatz generalizes to arbitrary new spurions of $G_{\rm flavor}$ that can arise in other leptoquark models. A complete list of these spurions is given in Ref. [1070].

Once the effective theory is known, we can make predictions for the contributions of new physics to various observables. Because the same spurion contributes to multiple observables, these predictions are correlated by a FN ansatz. These correlations can lead to inconsistencies with experimental results. Consequently, it is useful to have a systematic way of deviating from this scaling while still maintaining the predictivity of FN models. We discuss a systematic way of doing this in the next section. Specifically, we show how modifications of the UV spectrum of a FN construction can allow a controlled deviation from correlations between various observables in the IR, alleviating violations of experimental bounds.

7.3 Wrinkles in Froggatt–Nielsen

As described in the previous section, the FN mechanism provides a natural ansatz for the hierarchies of new flavor spurions coupled to the SM quarks and leptons. However, given our lack of knowledge about the dynamics underlying the flavor structure of the SM, it is worth exploring how this ansatz could change within the general framework of horizontal symmetry explanations for the SM flavor pattern.

In this spirit, we introduce the notion of "wrinkles", as a way of parametrically changing the FN ansatz for the flavor spurions that is described above without introducing additional scales. In §7.3.1, we will define them precisely, and argue that they allow for more flexibility in correlations between different flavor observables. While this flexibility inherently makes our ansatz less predictive, the freedom to introduce wrinkles is not absolute: there is a bound on the number of wrinkles imparted by radiative corrections,

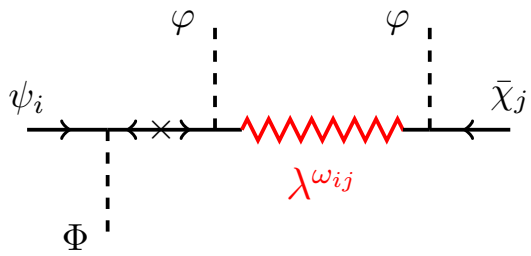


Figure 7.2: A cartoon illustrating a "wrinkle" in the Yukawa coupling $\Phi\psi_i\bar{\chi}_j$, which leads to a change in the predicted scaling from the FN ansatz.

which we will discuss in §7.3.2. In §7.3.3, we give several explicit examples of realizations of wrinkles in UV models.

7.3.1 Wrinkled Froggatt–Nielsen Chains

In §7.2.2, we described how the FN ansatz leads to a natural power counting for new flavor spurions in powers of $\lambda \equiv \langle\phi\rangle/M$, which we identify with the Cabbibo angle. Here, we generalize this power counting by considering modifications to the power of λ that appears in the spurion.

Consider a flavor spurion $Y_{\psi\bar{\chi}}$, where ψ, $\bar{\chi}$ are given SM matter fields. We introduce what we call "wrinkles" to modify the scaling of a given element of $Y_{\psi\bar{\chi}}$:

$$Y^{ij}_{\psi\bar{\chi}} \sim W^{ij}_{\psi\bar{\chi}}\lambda^{|[\psi_i]+[\bar{\chi}_j]|} \equiv \lambda^{\omega_{ij}+|[\psi_i]+[\bar{\chi}_j]|}. \tag{7.11}$$

Here we denote the power of λ that appears in $W^{ij}_{\psi\bar{\chi}}$ by $\omega^{ij}_{\psi\bar{\chi}}$ which, for simplicity, is assumed to be an integer. This additional scaling is motivated by allowing for additional structure in the UV, such as symmetries inducing obstructions in the heavy fermion chains which generate the non-renormalizable operators, and is illustrated schematically in Fig. 7.2. In general, any modification of the UV theory that gives rise to deviations from predictions of the vanilla FN setup without changing the number of power counting parameters can be considered a wrinkle. Different UV completions can lead to different correlated patterns of matrix entries $\omega^{ij}_{\psi\bar{\chi}}$ as we will discuss in §7.3.3, but from the IR perspective, these correlations are not apparent.

To be concrete, consider the example of the spurions Δ_{QL} and $\Delta_{\bar{u}\bar{e}}$ for the S_1 leptoquark, as in Eqs. (7.8) and (7.9). With additional wrinkles, the couplings in Eq. (7.10) are modified to

$$\Delta^{ij}_{QL} \sim \lambda^{\omega^{ij}_{QL}+|[Q_i]+[L_j]|}, \qquad \Delta^{ij}_{\bar{u}\bar{e}} \sim \lambda^{\omega^{ij}_{\bar{u}\bar{e}}+|[\bar{u}_i]|+[\bar{e}_j]} \tag{7.12}$$

where ω_{QL} and $\omega_{\bar{u}\bar{e}}$ are matrices of integers, whose elements ω^{ij}_{QL} and $\omega^{ij}_{\bar{u}\bar{e}}$ can vary across generations independently for both fermions. The idea of wrinkles can also be extended to models with additional scales by allowing wrinkles for each power counting parameter. Here we will focus on the case with a single expansion

parameter and not discuss the case of multiple parameters further.

Note that there are two distinct possibilities allowed by introducing wrinkles. The most straightforward one is that the number of factors of λ in some couplings of a *new* flavor spurion are modified, suppressing or enhancing their contributions to some flavor observables. For instance, wrinkles could suppress BSM contributions to observables such as electric and magnetic dipole moments (EDMs and MDMs) or light meson decays, which are generally strongly constrained, and allow for spurions with smaller mass scales to contribute to other observables. We will discuss this possibility thoroughly, again in the case of the S_1 leptoquark, in §7.4.

The second possibility is that wrinkles could exist in SM chains—i.e., $Y_{Q\bar{u}}$, $Y_{Q\bar{d}}$, or $Y_{L\bar{e}}$ could have fewer or additional factors of λ. In the IR, the SM Yukawa matrices must still match the measured masses and mixing angles of the quarks and leptons. Wrinkles in SM chains therefore necessitate different horizontal charges than the ones shown in Table 7.1. This changes the expected scaling for BSM spurions, leading to different couplings than expected in a naïve FN ansatz between the SM fermions and new particles. We will not comment in detail on particular phenomenological applications of this scenario, but highlight that this is an interesting direction for further exploration.

7.3.2 Bounds on Wrinkles from Radiative Corrections

Allowing for wrinkles would appear to entirely eliminate the predictivity of the FN ansatz. However, there is a natural bound on the size of the wrinkles that arises from demanding that the observed flavor structure in the IR arises predominantly from *tree-level* contributions to the effective operators below the scale M. Requiring that the tree-level contribution (including wrinkles) to the Yukawa coupling is larger than any subleading corrections from loops leads to a number of *consistency conditions* on the Yukawas, which in turn set a bound on the wrinkles. Provided these conditions are satisfied, the flavor structure in the IR is still determined by the FN mechanism in a predictive way, with departures from the minimal implementation parameterized by the wrinkles.

To illustrate these constraints, consider the Yukawa coupling matrix between the right-handed up-type quarks and the right-handed charged leptons for the S_1 leptoquark model in Eq. (7.8), $\Delta^{ij}_{\bar{u}\bar{e}}$. In a FN setup, this coupling arises from a non-renormalizable operator with a minimal number of flavons. It can be UV completed with a tree-level chain of heavy fermions and flavons with a single leptoquark vertex, as illustrated on the left in Fig. 7.3. However, the same operator can also be generated at higher order by including SM fermions in the FN chain $\bar{u}^i \to Q^k \to L^l \to \bar{e}^j$, as shown on the right in Fig. 7.3. The first and last connections include additional Higgs insertions that are tied together to form a loop, and the $Q^k \to L^l$ connection involves a leptoquark interaction. Thus, the higher-order contribution to $\Delta^{ij}_{\bar{u}\bar{e}}$ is:

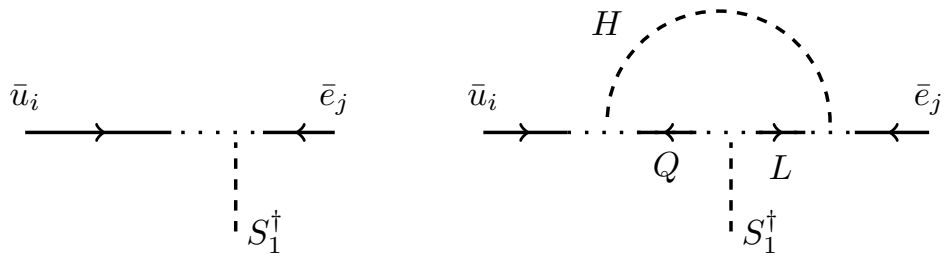


Figure 7.3: Left: the tree level $S_1^\dagger \bar{u}_i \bar{e}_j$ coupling. Right: A loop contribution to the same spurion, leading to the spurion contribution described by Eq. (7.13). In both diagrams the dots indicate a chain of flavon and heavy fermion vertices, whose length is determined by the horizontal charges of the particles, which we suppress for clarity.

$$\Delta^{ij}_{\bar{u}\bar{e}}\Big|_{\text{loop}} \sim \frac{1}{16\pi^2}\left(Y^T_{Q\bar{u}} \cdot \Delta^*_{QL} \cdot Y_{L\bar{e}}\right)^{ij}. \tag{7.13}$$

Demanding this contribution to be smaller than the tree-level contribution, and assuming the absence of any artificial cancellations, leads to a lower bound on the Yukawa coupling $\Delta^{ij}_{\bar{u}\bar{e}}$ and an upper bound on the entries of Δ^*_{QL}. This bound begets a set of consistency conditions on the wrinkles:

$$\begin{aligned} \left|\Delta^{ij}_{\bar{u}\bar{e}}\right| &\gtrsim \frac{1}{16\pi^2}\left|\left(Y^T_{Q\bar{u}} \cdot \Delta^*_{QL} \cdot Y_{L\bar{e}}\right)^{ij}\right|, \\ \implies \lambda^{\omega^{ij}_{\bar{u}\bar{e}} + \left|[\bar{u}_i] + [\bar{e}_j]\right|} &\gtrsim \frac{1}{16\pi^2}\left|Y^T_{Q\bar{u}}\right|^{ik} \lambda^{\omega^{kl}_{QL} + \left|[Q_k] + [L_l]\right|}\left|Y_{L\bar{e}}\right|^{lj}, \end{aligned} \tag{7.14}$$

where there is an implicit summation over the indices k and l above. While the SM Yukawas on the right hand side of this relation may also contain wrinkles, it is the IR value of the coupling that appears, which is fit to the SM masses and mixing angles.

Similar *consistency conditions* were proposed in [1069–1071], neglecting the loop factor. Other similar constraints (including the loop factor) have been considered as naturalness constraints on models of flavorful new physics [1096]. We settle for the weaker constraint, including the loop factor, as a concrete, irreducible bound.[4] Note that there are also other higher order contributions to the spurions, such as those from higher-dimensional operators with the Higgs replaced by its vacuum expectation value, but they will be smaller than the one in Eq. (7.13), since $v^2/M^2 < 1/16\pi^2$.

More generally, a complete set of consistency conditions can be derived by again considering the Yukawas as spurions under G_{flavor}. In the absence of any additional symmetries, contributions similar to Eq. (7.13) arise from any combination of Yukawa couplings that transform in the same representation of G_{flavor}. The complete list of leading consistency conditions for all of the Yukawa couplings in the SM extended with the

[4] RG evolution of the leptoquark couplings also does not change the above set of bounds, as long as one imposes the consistency conditions at the matching scale of order M. The structure of the one loop Yukawa RGEs involves the same higher order operators as appearing in our consistency condition. Thus, imposing the consistency condition at the matching scale ensures that running is a small effect and can be neglected. Consequently, RG evolution to scales below the matching scale ensures that the consistency condition (inequality) holds at all such scales.

S_1 leptoquark are listed in Appendix E.1.

These inequalities must be satisfied for any wrinkled FN setup involving additional flavor spurions, and they impose non-trivial constraints on the size of the wrinkles introduced in Eq. (7.11).[5] The details of these constraints depend on the particular charge assignment of the SM fermions, but once these are fixed, a degree of predictiveness is returned to the FN ansatz, even in the presence of wrinkles. As pointed out in [1069–1071], these consistency conditions are trivially satisfied in a vanilla FN setup without wrinkles, as a result of the triangle inequality.

As an example of how this bound works with nonzero wrinkles, consider the charge assignment in Table 7.1 with $q_0 = 0$, $l_0 = -1$, $X = +1$, $Y = -1$ and all other sign choices being positive. Assuming no wrinkles in the SM Yukawas, the bound on $\omega_{\bar{u}\bar{e}}^{33}$ from Eq. (7.14) becomes

$$\begin{aligned}\omega_{\bar{u}\bar{e}}^{33} &\lesssim \sum_{k,l}\Big(\big|[Q_k]+[\bar{u}_3]\big|+\big|[Q_k]+[L_l]\big|+\omega_{QL}^{kl}+\big|[L_l]+[\bar{e}_3]\big|\Big) \\ &\qquad +\log_\lambda\frac{1}{16\pi^2}-\big|[\bar{e}_3]+[\bar{u}_3]\big| \\ &\lesssim 2+\omega_{QL}^{33}+\log_\lambda\frac{1}{16\pi^2},\end{aligned} \tag{7.15}$$

where in the last line we have assumed that $k = l = 3$ is the largest entry in ω_{QL}^{kl}, which is typically the case. We see that, at least for this consistency condition, up to five wrinkles on $\Delta_{\bar{u}\bar{e}}^{33}$ are allowed, even without extra wrinkles on Δ_{QL}^{33}.

A similar argument for general couplings, again using the triangle inequality, makes it clear that if all $\omega_{\psi\bar{\chi}}^{ij} \geq 0$, a *sufficient* condition on the wrinkles is that they are all greater than a loop factor:

$$(W_{\psi\bar{\chi}})^{ij} \gtrsim \frac{1}{16\pi^2}. \tag{7.16}$$

Note that in this equation, we have assumed a mild separation of scales so that the logarithms in the loop contribution can be neglected along with other $\mathcal{O}(1)$ factors in the loop calculation. In this work, we focus on the bound in Eq. (7.16) and leave further studies of more accurate lower bounds on wrinkles for future work. As shown in Eq. (7.15), this bound may be overly restrictive, but it provides a useful shortcut for employing wrinkles in an EFT without having to manually check all the consistency conditions.

7.3.3 UV Completions

We now turn to UV completions of the wrinkles introduced in Eq. (7.11). Our goal is not to provide an exhaustive or detailed list of examples, but demonstrate a proof of principle of potential ways these wrinkles can arise from more complicated UV completions.

[5] The consistency conditions, as written, hold neglecting $\mathcal{O}(1)$ couplings; there may be small deviations from including them.

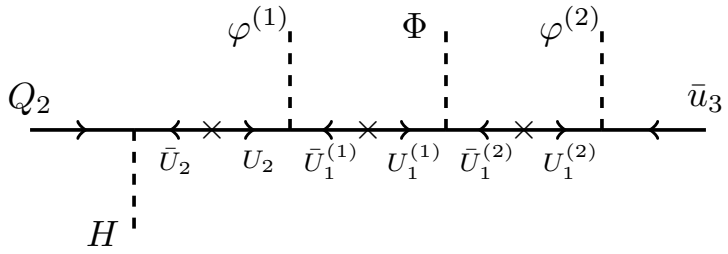


Figure 7.4: An explicit realization of a "wrinkled" FN chain, where the heavy quark with horizontal charge +1 is replaced by two heavy quarks, along with additional flavons, transforming under additional symmetries.

Missing Heavy Fermions

As a first concrete realization of the idea sketched in Fig. 7.2, we consider a situation where one of the heavy fermions with a particular horizontal charge does not exist in the spectrum. Instead, the chain leading to the effective operator can only be completed by including additional fermions and scalars, causing additional suppression.

To illustrate this mechanism, we consider the example in Fig. 7.1 and replace a single heavy vector-like pair of fermions U_1, $\bar{\mathrm{U}}_1$ with two sets of vector-like pairs, which we will denote by $\mathrm{U}_1^{(1)}$, $\bar{\mathrm{U}}_1^{(1)}$ and $\mathrm{U}_1^{(2)}$, $\bar{\mathrm{U}}_1^{(2)}$. These are assumed to have the same SM and horizontal charges as U_1, $\bar{\mathrm{U}}_1$, but also transform as conjugate pairs under new symmetry groups, G_1 and G_2, respectively. To be explicit, we will take $\mathrm{G}_1 = \mathrm{SU}(\mathrm{N}_1)$ and $\mathrm{G}_2 = \mathrm{SU}(\mathrm{N}_2)$ to be two different continuous, non-Abelian groups, but the following construction works for arbitrary (continuous or discrete) groups as well, with straightforward modifications. To complete the chain diagram, we must also introduce new flavons, which we take to be in the representations,

$$\phi^{(1)} : (\mathbf{N}_1, 1)_{-1}, \qquad \phi^{(2)} : (1, \overline{\mathbf{N}}_2)_{-1}, \qquad \Phi^{(1,2)} : (\overline{\mathbf{N}}_1, \mathbf{N}_2)_0, \tag{7.17}$$

where the parentheses indicate the $\mathrm{SU}(\mathrm{N}_1)\times\mathrm{SU}(\mathrm{N}_2)$ representation, and the subscript is the horizontal charge. These allow us to construct the diagram shown in Fig. 7.4, where both of the extra heavy fermion pairs are traversed between Q_2 and $\bar{\mathrm{u}}_3$. The charge assignments forbid the couplings $\phi^{(1)}\mathrm{U}_2\bar{\mathrm{U}}_1^{(2)}$ and $\phi^{(2)}\mathrm{U}_1^{(1)}\bar{\mathrm{u}}_3$, so that this diagram is the leading effective operator containing $\mathrm{HQ}_2\bar{\mathrm{u}}_3$.

Assuming all the scalars acquire vevs $\sim \langle\phi\rangle$ and that the new fermions have vector-like masses $\sim \mathrm{M}$, this replaces the λ^2 suppression inferred from the horizontal charges with a λ^3 suppression. In other words, this leads to a "wrinkle", $\mathrm{W}_{\mathrm{Q}\bar{\mathrm{u}}}^{23} \sim \lambda$.

This construction can be extended to include arbitrarily many wrinkles in place of a single heavy fermion. For example, $\mathrm{W}_{\mathrm{Q}\bar{\mathrm{u}}}^{23} \sim \lambda^2$ is obtained by introducing additional mirror quarks, $\mathrm{U}_1^{(3)}$, $\bar{\mathrm{U}}_1^{(3)}$, replacing $\phi^{(2)}$ with a bi-fundamental $\Phi^{(2,3)}$ transforming as $(1, \overline{\mathbf{N}}_2, \mathbf{N}_3)_0$, closing the chain with $\phi^{(3)}$, which transforms as a

$(1, 1, \overline{\mathbf{N}}_3)_{-1}$; further wrinkles are obtained for additional mirror quarks. In these types of examples, the Higgs and chiral fermions of the SM are neutral under the new symmetries, so as to be compatible with the general arguments in Ref. [1047].

Note that with this mechanism, we see an example of the correlation between wrinkles in different chains. We constructed this wrinkle in the context of the Q_2 and $\bar{u}_3$ chain, but since we have removed U_1, $\bar{U}_1$ from the spectrum, the wrinkle necessarily appears in any chain involving them. For instance, assuming heavy up-like quarks are responsible for all of the up-type Yukawa couplings, it would also appear in the $Q_1 H \bar{u}_3$ operator.

Extra Abelian Symmetries

Another concrete example in which wrinkles can appear in an effective theory with the FN ansatz is realized by considering additional Abelian symmetries in the UV, under which the SM fermions are charged. In particular, we can consider gauging the non-anomalous combinations of baryon number, B, and the individual lepton numbers, L_e, L_μ, and L_τ, as is frequently done in model-building for various flavor anomalies [1097]. These symmetries are preserved by the SM Yukawa couplings, but generically violated by neutrino masses and additional Yukawa couplings between SM fermions and new BSM fields, such as leptoquarks. For concreteness, we again work with the S_1 leptoquark and assume it is neutral under the new symmetry; therefore the flavor spurion must absorb the remaining U(1) charge. This means that additional flavons charged under the extra symmetries also must be included in order to complete the leptoquark Yukawa couplings. The usual flavon, with $U(1)_H$ charge -1, is still present, since it is required to complete the SM Yukawa couplings.[6]

In contrast to the UV completions discussed in §7.3.3, where the wrinkles are always additional suppression factors, the wrinkles that result from these extra symmetries can naturally either suppress or enhance the size of the flavor spurions. Another distinction is that we have not removed any fermions of particular charges from the UV spectrum in this case: we allow fermions with all required quantum numbers to exist.

Just like other UV models, additional symmetries and the flavons charged under them can generate a correlated pattern of wrinkles for the different chains. The details of those correlations depend on whether the new symmetries are flavor universal or flavor specific; we will discuss examples of both cases. In order to maintain the predictivity of our example, we also assume additional symmetries are spontaneously broken at similar scales to the $U(1)_H$ symmetry.

The flavor universal case is simpler, but also less flexible because of interdependence between different chains. For example, assuming $U(1)_{B-L}$ is a symmetry of the theory, we can construct the leptoquark Yukawa

[6] In the presence of neutrino masses, the extra flavons may also be required to generate the PMNS matrix structure, depending on the additional symmetries we impose.

spurions by introducing an additional flavon, $\tilde{\phi}$, which we take to have $B-L$ charge $1/3$. The new flavon will not change any of the SM chains, since they respect the $B-L$ symmetry, but the leptoquark chains can be different from the usual FN scenario. For instance, if $\tilde{\phi}$ has no $U(1)_H$ charge, all the leptoquark Yukawas will become smaller by λ^2, since the external fermions all have $B-L$ charge difference $\pm 2/3$ without the new flavon. If $\tilde{\phi}$ also has $U(1)_H$ charge ≥ 1, then the pattern of leptoquark chains becomes more intricate. Since each leptoquark chain must contain exactly two copies of the new $B-L$ flavon and the remaining difference in horizontal charge requires insertions of the original flavon, whether a given chain becomes shorter or longer depends on the details of the assigned horizontal charges.

We have somewhat more freedom in the flavor specific case. As an example, consider introducing a new $U(1)_{B-3L_e}$ symmetry. Now both the PMNS matrix and chains for the leptoquark Yukawa couplings require new flavons charged under both $U(1)_H$ and $U(1)_{B-3L_e}$. We introduce two additional flavons: ϕ_2 is necessary to generate PMNS matrix entries of the correct size, and ϕ_3 is necessary to complete the leptoquark chains while respecting the additional symmetries. These flavons have $B-3L_e$ charges

$$[\phi] = 0 \qquad [\phi_2] = 3 \qquad [\phi_3] = -1/3. \tag{7.18}$$

Each also carries $U(1)_H$ charge -1. Including these extra symmetries and flavons charged under them creates wrinkles by changing the required number of vev insertions for the leptoquark couplings compared to the spurion size we would naively expect with only these $U(1)_H$ charges. For example, if we consider only the couplings to the third generation leptons, we make the right-handed μ and τ couplings smaller while leaving the right-handed e coupling and the left-handed couplings unaffected. This is shown in Fig. 7.5. Nonetheless, despite the additional freedom in the flavor specific case, it is still challenging to obtain certain patterns of wrinkles, such as those constrained by the triangle inequality.

Finally, we comment on a few modifications to the examples above. First, we note that it is possible to modify this approach by charging the leptoquark under $U(1)_H$ instead of/in addition to additional flavon(s). Similar to the $B-L$ charged flavons, this is another mechanism to add wrinkles to the leptoquark couplings without affecting the SM couplings. In principle, we can also charge the leptoquark under the additional symmetries we discussed in this section, but note that we are not always guaranteed a charge assignment which makes all of the couplings invariant. Second, we note that like the previous case, other modifications such as using discrete Abelian symmetries also behave similarly. However, we can not replace these Abelian symmetries with continuous non-Abelian ones [1047], because we are charging the SM fermions under the new symmetry. This is in contrast to the previous case, where only internal fermions are charged under new non-Abelian symmetries.

While we have provided two different ways in which wrinkles could be generated, we have not exhausted

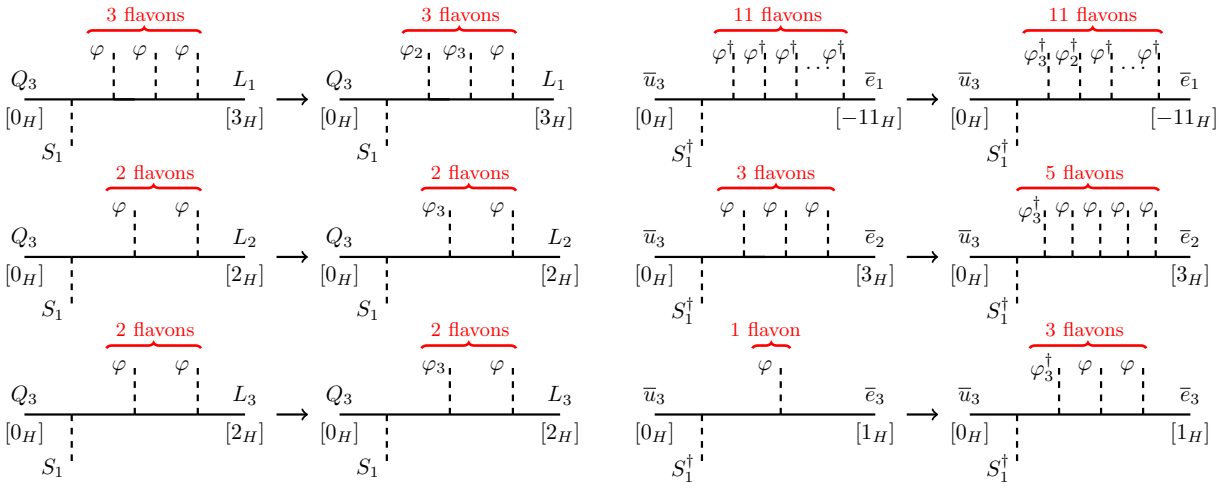


Figure 7.5: Chains before and after adding flavons charged under a $U(1)_{B-3L_e}$ symmetry to generate wrinkles. The horizontal charges correspond to Table 7.1 with $q_0 = 0$, $l_0 = 2$, and $Y = 1$. We observe that the new $U(1)_{B-3L_e}$ symmetry and its flavons modify the prediction of the model for some of the leptoquark couplings in the IR.

the possibilities. These are only examples, and there are undoubtedly many more options for generating wrinkles, which would be interesting for future work. Since the details of a particular model are not the central point of this chapter, we now move to discussing a full example in the IR.

7.4 $B \to K\bar{\nu}\nu$ in a Wrinkled Setup

To demonstrate the ideas of the previous sections with a specific example, in this section we study the phenomenology of the S_1 leptoquark introduced in Eq. (7.8) with particular flavor ansätze in detail. Such ansätze correlate the contribution of S_1 to different observables. As mentioned in the previous section, the inclusion of wrinkles in a FN ansatz can change the relative sizes of predictions for different flavor observables. This could allow a model to accommodate a significant excess over the SM in one observable, while suppressing other observables that would otherwise be too constraining.[7]

As an illustration, we will focus on constructing a model that can give rise to a large signal in the semi-leptonic decay $B^+ \to K^+\bar{\nu}\nu$. $\mathrm{BR}(B^+ \to K^+\bar{\nu}\nu)$ is an interesting test case for several reasons. Assuming the vanilla FN ansatz, the mass range preferred for new physics near the current experimental sensitivity is in the few TeV range, and small hints of flavorful new physics may have already been detected [698]. Like all flavor-changing neutral currents, the $b \to s\bar{\nu}\nu$ transition is greatly suppressed in the SM. It is also relatively clean theoretically, with the uncertainties in the hadronic form factors and from perturbative effects well

[7] Signals of leptoquarks in all flavor experiments can also be suppressed by choosing q_0 and l_0 (defined in Table 7.1) such that the quarks' and leptons' charges are very far apart, but this requires an unnaturally large separation of charges. This choice also does not permit the explanation of any discrepancies in flavor experiments because it suppresses leptoquark contribution to all observables.

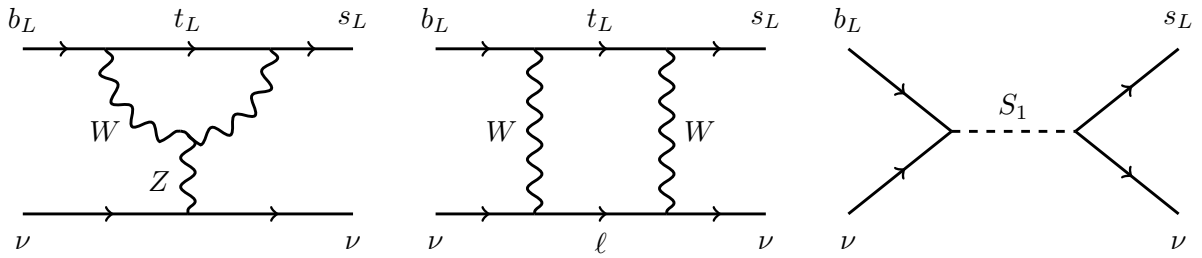


Figure 7.6: Example Feynman diagrams leading to $b \to s\bar{\nu}\nu$ transitions in the SM extended with an S_1 leptoquark. The left (center) diagram show the leading one loop SM contributions with the penguin (box) topology, while the right diagram illustrates the tree-level leptoquark contribution. The Z in the left diagram could also connect to the top line instead.

under control [1098–1106]. This situation, along with the prospect of observing the decay at the Belle II experiment in the near future, make it an intriguing probe of BSM physics [1107–1109]. We use this specific observable as a testbed of various ideas introduced in the previous section; similar studies can be carried out for any other flavorful anomalies that may emerge in experimental data.

7.4.1 $B \to K\bar{\nu}\nu$ in the SM and Beyond

In order to understand how various FN ansätze contribute to $\text{BR}\left(B^+ \to K^+\bar{\nu}\nu\right)$, we first need to discuss the SM and leptoquark contributions, as well as experimental bounds. Typically, $\text{BR}\left(B^+ \to K^+\bar{\nu}\nu\right)$ is parameterized in terms of the Wilson coefficients C_R^{ij} and C_L^{ij}, which are defined implicitly in the effective Hamiltonian governing $b \to s\bar{\nu}\nu$ transitions

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left(C_L^{ij} \mathcal{O}_L^{ij} + C_R^{ij} \mathcal{O}_R^{ij}\right) + \text{h.c.} \tag{7.19}$$

where

$$\mathcal{O}_L^{ij} = \frac{\alpha_{em}}{2\pi} \left(s_L^\dagger \bar{\sigma}^\mu b_L\right) \left(\nu_j^\dagger \bar{\sigma}_\mu \nu_i\right), \qquad \mathcal{O}_R^{ij} = \frac{\alpha_{em}}{2\pi} \left(s_R^\dagger \sigma^\mu b_R\right) \left(\nu_j^\dagger \bar{\sigma}_\mu \nu_i\right), \tag{7.20}$$

and $i, j = e, \mu, \tau$ are neutrino flavor indices.

In the SM (and in the S_1 leptoquark model we consider below), only C_L is non-zero. The leading contribution to the SM value of the Wilson coefficient arises from diagrams such as those in Fig. 7.6. Also including NLO QCD corrections [1098–1100] and two-loop electroweak contributions [1104], the SM Wilson coefficient is

$$C_L^{ij,\,\text{SM}} = (-6.353 \pm 0.074)\, \delta_{ij}, \tag{7.21}$$

where δ_{ij} captures the fact that the SM contributions are lepton flavor conserving. This leads to a prediction for the branching ratio [1108, 1109],

$$\text{BR}\left(B^+ \to K^+\bar{\nu}\nu\right)\Big|_{\text{SM}} = (0.46 \pm 0.05) \times 10^{-5}, \tag{7.22}$$

where we sum over neutrino flavors.

This process has been searched for at Belle and BaBar by tagging the second B meson in either a hadronic or semileptonic decay [1110–1112]. Similar searches exist for $\mathrm{BR}(B \to K^* \bar{\nu}\nu)$, e.g. see [1110, 1111]. Each of these channels leads to the same qualitative conclusions; thus, for the rest of this work we will focus on $\mathrm{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right)$ measurements, for simplicity. A combination of these results yields a 90% C.L. upper limit on the branching ratio of

$$\mathrm{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right) < 1.6 \times 10^{-5}. \tag{7.23}$$

Recently, Belle II has searched for the same decay using an inclusive tagging technique, which allows them to partially compensate for their smaller dataset and larger backgrounds [698]. Though not yet statistically significant, a combination of these results (assuming their uncertainties are uncorrelated) leads to a best fit value of

$$\mathrm{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right) = (1.1 \pm 0.4) \times 10^{-5}, \tag{7.24}$$

which leaves room for a BSM contribution on top of the SM prediction in Eq. (7.22). The uncertainties in all of these estimates—both the tagged and inclusive searches—are predominantly statistical, and are expected to improve and become comparable to the theoretical uncertainty in Eq. (7.22) with the forthcoming full Belle II dataset [1113]. Therefore, while it remains to be seen if any signals of new physics exist in this channel, it provides an interesting application of our wrinkled FN setup.

The S_1 leptoquark contributes to $b \to s\bar{\nu}\nu$ transitions via the tree-level diagram shown on the right in Fig. 7.6. It generates a Wilson coefficient

$$C_L^{ij} \propto \frac{v^2}{m_{S_1}^2} \Delta_{QL}^{3i} \Delta_{QL}^{2j\,*} \tag{7.25}$$

for the effective theory of Eq. (7.19). Since this is the same operator as generated in the SM, it is convenient to capture these effects by considering the ratio:

$$R_K^{\nu\nu} \equiv \frac{\mathrm{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right)}{\mathrm{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right)\big|_{SM}}. \tag{7.26}$$

The contribution from S_1 is given by [1114, 1115] (see also [1072, 1108])

$$R_K^{\nu\nu} = 1 - y\,\mathrm{Re}\left[\frac{(\Delta_{QL}^{3i}\Delta_{QL}^{2i\,*})}{V_{tb}V_{ts}^*}\right] + \frac{3y^2}{4}\frac{(\Delta_{QL}^{3i}\Delta_{QL}^{3i\,*})(\Delta_{QL}^{2j}\Delta_{QL}^{2j\,*})}{\left|V_{tb}V_{ts}^*\right|^2}, \tag{7.27}$$

with a sum over repeated lepton indices in each term, and

$$y \equiv -\frac{2\pi v^2}{6 C_L^{SM} \alpha_{em} m_{S_1}^2} \simeq \left(\frac{1.2\,\mathrm{TeV}}{m_{S_1}}\right)^2. \tag{7.28}$$

Observable	S_1 Yukawa Couplings	Experimental Result	Future Bounds
$BR(B^+ \to K^+ \bar{\nu}\nu)$	$\Delta^{3i}_{QL} \times (\Delta^{2j}_{QL})^*$	$(1.1 \pm 0.4) \times 10^{-5}$ [698]	-
electron EDM	$(V^* \Delta_{QL})^{31} \times (\Delta^{31}_{\bar{u}\bar{e}})^*$	$< 4.1 \times 10^{-30}$ e cm [1116]	$< 10^{-31}$ e cm [1117, 1118]
$BR(\mu \to e\gamma)$	$(V^* \Delta_{QL})^{32} \times \Delta^{31}_{\bar{u}\bar{e}}$ $\Delta^{32\,*}_{\bar{u}\bar{e}} \times (V^* \Delta_{QL})^{31\,*}$	$< 4.2 \times 10^{-13}$ [1119]	$< 6 \times 10^{-14}$ [1120]
$CR(\mu \to e)_N$	$(V^* \Delta_{QL})^{11\,*} \times (V^* \Delta_{QL})^{12}$	$< 7.0 \times 10^{-13}$ [1121]	$< 2.5 \times 10^{-18}$ [1122, 1123]
$BR(\tau \to \mu\gamma)$	$(V^* \Delta_{QL})^{33} \times \Delta^{32}_{\bar{u}\bar{e}}$ $\Delta^{33\,*}_{\bar{u}\bar{e}} \times (V^* \Delta_{QL})^{32\,*}$	$< 4.2 \times 10^{-8}$ [1124]	$< 6.9 \times 10^{-9}$ [1125, 1126]
$BR(K^+ \to \pi^+ \bar{\nu}\nu)$	$\Delta^{2k}_{QL} \times (\Delta^{1k}_{QL})^*$	$< 1.88 \times 10^{-10}$ [1127]	$(8.4 \pm 0.4) \times 10^{-11}$ [1128]
Δm_{B_s}	$(\Delta_{QL}\, \Delta^{\dagger}_{QL})^{32}$	$\Delta C_{B_s} \leq 0.09$ [1129]	$\Delta C_{B_s} \leq 0.026$ [1129]

Table 7.2: Here we show the experimental results for $BR(B^+ \to K^+ \bar{\nu}\nu)$ and a few other constraining observables; we also show the predominant S_1 Yukawa couplings contributing to each. Note that for B-mixing, we use the experimental uncertainty on the quantity C_{B_s} as defined in Eq. (7.33). For $K^+ \to \pi^+ \nu\bar{\nu}$, the future bound corresponds to reaching a 5% experimental uncertainty on the SM branching ratio [1130]. The muon to electron conversion rate in nuclei, $CR(\mu \to e)_N$, gets contributions from both dipole and four-fermion operators; we show the Yukawas entering the four-fermion operator that is dominant in the FN ansatz (associated with a left-handed vector current) here, while the complete set is given in Appendix E.2. The current (future) bound listed for it is on the conversion rate in a gold (aluminum) nucleus.

In terms of $R^{\nu\nu}_K$, the 90% C.L. limit and 68% C.L. preferred values of the branching ratio in Eqs. (7.23) and (7.24) translate to

$$R^{\nu\nu}_K < 3.4, \qquad R^{\nu\nu}_K \in [1.5, 3.3], \tag{7.29}$$

respectively. The interpretation of these bounds in the context of the leptoquark depends on the assumptions made about the hierarchies in Δ^{ij}_{QL}, to which we now turn.

7.4.2 Constraints with Different Flavor Ansätze

In addition to $b \to s\bar{\nu}\nu$ transitions discussed above, the S_1 leptoquark can contribute to a number of flavor-changing processes or precision observables that are constrained by experiments. These include electric and magnetic dipole moments of SM particles, LFV decays, leptonic and semi-leptonic meson decays, flavor-violating decays of gauge bosons, and neutral meson mixing. Some of the most powerful observables, and their dependence on the leptoquark Yukawa couplings are summarized in Table 7.2.[8] As is apparent from the table, the observables depend on numerous different combinations of the leptoquark couplings. More

[8] For simplicity, we work with flavor basis neutrinos, so no dependence on the PMNS matrix appears.

details about the observables, including the dependence on the leptoquark couplings and references to more complete treatments in the literature, are given in Appendix E.2.

Because the contributions to various observables are correlated, we need to pick a particular ansatz and study it in order to understand these constraints. In the rest of this section, we study these constraints in the context of three different flavor ansätze: flavor anarchy, vanilla FN, and FN with wrinkles. In particular, we explore how adding wrinkles can alleviate constraints while maintaining consistency with $\mathrm{BR}\,(\mathrm{B}^+ \to \mathrm{K}^+ \bar{\nu}\nu)$ measurements.

Without any assumptions about the underlying structure, a minimal assumption is that all elements of Δ_{QL} and $\Delta_{\bar{\mathrm{u}}\bar{\mathrm{e}}}$ are $\mathcal{O}(1)$. This assumption is commonly referred to as "flavor anarchy". Under this assumption, the mass of the leptoquark consistent with the $\mathrm{BR}\,(\mathrm{B}^+ \to \mathrm{K}^+ \bar{\nu}\nu)$ measurements is $\mathrm{m}_{\mathrm{S}_1} \in (9,\ 18)$ TeV. On the other hand, measurements of the electron EDM and other flavor-changing processes constrain the mass of the leptoquark to be above $\sim 10^5$ TeV. The resulting limits for some of the observables considered are shown as yellow bars in Fig. 7.7. To calculate these ranges for observables that are already measured experimentally, we demand the leptoquark contribution to be within one standard deviation of the measured value, while for others we use the reported upper bounds from Ref. [1131].[9] For the electron EDM, a CP-odd observable, we assume a purely imaginary coupling to show the maximum reach of the experimental results.

It is clear that without any flavor texture on the leptoquark Yukawas, observables such as the electron EDM, LFV decays, or meson-mixing parameters rule out the leptoquark mass range relevant for $\mathrm{BR}\,(\mathrm{B}^+ \to \mathrm{K}^+ \bar{\nu}\nu)$.[10] We have also checked the contribution of our setup to many other similar observables (electron and tau MDM, $\tau \to \mathrm{e}\gamma$, $\mathrm{K} \to \mathrm{e}\nu$, various other D meson decays, $\mathrm{D_s} \to \mathrm{e}\nu$, $\mathrm{B} \to \mathrm{e}\nu$, $\pi \to \mathrm{ee}$, $\pi \to \mu\mathrm{e}$), but find that the constraints they place are not as competitive for our model.

Thus we are led to consider embedding the S_1 leptoquark in a FN model of flavor. This has the benefit of not only alleviating some of the experimental constraints discussed above, but also relating it to the SM flavor puzzle.

As discussed in §7.2, aside from the general shifts in the lepton and quark horizontal charges, there are only a handful of possible charge assignments that give rise to the correct pattern of SM masses and mixing angles. For concreteness, we choose horizontal charges from Table 7.1 with $\mathrm{q}_0 = 0$, $\mathrm{l}_0 = -1$, and $\mathrm{X} = -\mathrm{Y} = -1$. This yields

[9] The exceptions to this are $\mathrm{R_D}$ and $\mathrm{a_\mu}$, where we take the maximum leptoquark mass consistent to within 3σ and 4σ, respectively, of the experimental measurement for the anarchic coupling case, and use the 2σ ellipse for the preferred mass range in the wrinkled case.

[10] Our model can also contribute to $\mathrm{a_\mu}$ at one loop to explain the observed anomaly [683, 684], although recent lattice calculations [687, 1132–1137] and measurements [688] hint toward a smaller discrepancy with the experimental data. However, other observables already rule out the leptoquark mass range that has a large enough contribution to $\mathrm{a_\mu}$. See [1072, 1084, 1114, 1138–1150] for other solutions to this anomaly, including attempts at embedding the solution in a FN construction.

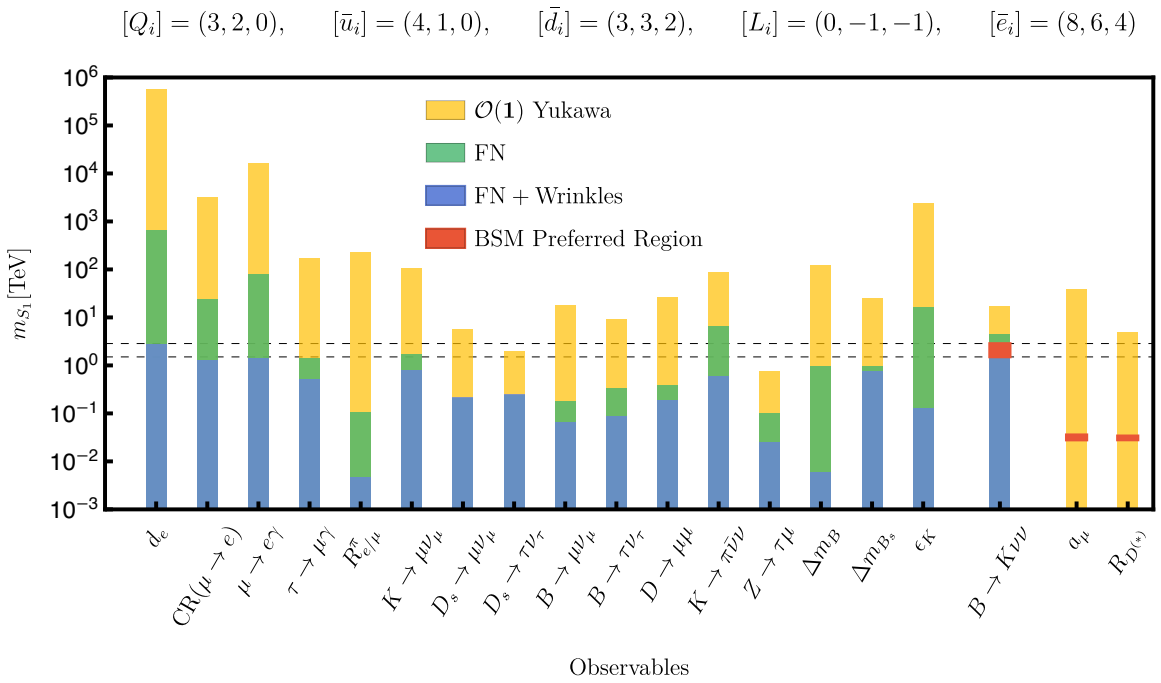


Figure 7.7: The leptoquark mass range probed by various observables if the Yukawa couplings of the leptoquark are either $\mathcal{O}(1)$ (yellow), follow the vanilla FN setup in Eq. (7.31) (green), or the same FN setup plus the wrinkles from Eq. (7.32) (blue). The preferred range for explaining some existing anomalies are shown in red, assuming the wrinkled setup. The undetermined $\mathcal{O}(1)$ factors in the Yukawas (folded in r_{ij} in Eq. (7.1)) can further affect the leptoquark contribution and slightly change the mass range probed by each observable. We see that in our wrinkled setup, the mass range that explains the current discrepancy in $\mathrm{BR}(B^+ \to K^+\bar{\nu}\nu)$ measurement (between the horizontal dashed lines) can also be probed by the LFV processes $\mu \to e\gamma$ and $CR(\mu \to e)$, and the electron EDM in near future measurements.

$$([Q_1], [Q_2], [Q_3]) = (3,2,0), \qquad ([\bar{u}_1], [\bar{u}_2], [\bar{u}_3]) = (4,1,0), \qquad ([\bar{d}_1], [\bar{d}_2], [\bar{d}_3]) = (3,3,2),$$
$$([L_1], [L_2], [L_3]) = (0,-1,-1), \qquad ([\bar{e}_1], [\bar{e}_2], [\bar{e}_3]) = (8,6,4). \tag{7.30}$$

With these charge assignments, the FN ansatz for the leptoquark couplings is:

$$\Delta_{QL} \sim \begin{pmatrix} \lambda^3 & \lambda^2 & \lambda^2 \\ \lambda^2 & \lambda & \lambda \\ 1 & \lambda & \lambda \end{pmatrix}, \qquad \Delta_{\bar{u}\bar{e}} \sim \begin{pmatrix} \lambda^{12} & \lambda^{10} & \lambda^8 \\ \lambda^9 & \lambda^7 & \lambda^5 \\ \lambda^8 & \lambda^6 & \lambda^4 \end{pmatrix}. \tag{7.31}$$

The resulting bounds, neglecting $\mathcal{O}(1)$ Yukawa factors, are shown as the green bars in Fig. 7.7. Compared to the anarchic ansatz, the bounds on the leptoquark mass are significantly relaxed.

Nevertheless, it is clear that the mass range consistent with the $\mathrm{BR}(B^+ \to K^+\bar{\nu}\nu)$ measurements at Belle II is still excluded by other observables under the FN ansatz. We have checked that—while the exact bounds for different observables can change significantly—this conclusion remains unchanged for the other possible charge assignments enumerated in Table 7.1. If any deviation from SM is observed in

$\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$, the S_1 leptoquark embedded in a vanilla FN model cannot explain the anomaly while respecting bounds from other measurements.

Adding wrinkles to the FN ansatz as discussed in §7.3 can ameliorate the tension with these observables. Using the scaling of the observables with the leptoquark Yukawas shown in Table 7.2 as a guide, we add the following wrinkles (as defined in Eq. (7.11)) to the leptoquark Yukawa matrices:

$$W_{\bar{u}e}^{ij} = \lambda^3, \qquad W_{QL} = \begin{pmatrix} \lambda^3 & \lambda^3 & \lambda^3 \\ \lambda^3 & 1 & 1 \\ \lambda^3 & 1 & 1 \end{pmatrix}. \tag{7.32}$$

This is the largest number of wrinkles we can add to suppress the leptoquark contribution to the most constraining observables (especially electron EDM, $\mu \to e\gamma$, $\tau \to \mu\gamma$, and meson mixing observables), while retaining consistency with the naïve constraint $\omega \gtrsim \lambda^3 \sim 1/16\pi^2$ from §7.3.2 and leaving the contribution to $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$ mostly intact. Further suppression with additional powers of λ may be possible, but must be carefully checked with all of the consistency conditions in Appendix E.1.

It is worth emphasizing that it is not obvious how to get the pattern of wrinkles in Eq. (7.32) from the example UV completions discussed in §7.3.3. Nevertheless, we can treat them consistently in an effective field theory approach, and leave the model-building to future work. Note also that with the additional suppression of the right-handed Yukawa couplings, the phenomenology of this model resembles that of the RPV down squark as discussed in §7.2.2.

The contribution of this wrinkled FN setup to various observables is shown by blue bars in Fig. 7.7. We find that the set of wrinkles from Eq. (7.32) sufficiently suppresses the contribution to other observables, so that they are all compatible with the mass range of interest for $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$. In particular, bounds from meson mixing observables and leptonic meson decays are circumvented. Within this wrinkled setup, the viable leptoquark mass range that can account for a signal in $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$ is slightly above the current direct search bounds at the LHC (see [1151, 1152]) and could be detected in future searches at the LHC or future hadron [1153–1156] or lepton [1157–1162] colliders.

There are several observables which probe a similar mass range to $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$ which will see significant improvement in experimental measurements soon. In particular, these observables include $\mu \to e\gamma$, $\mathrm{CR}(\mu \to e)$, and electron EDM, though the precise mass range depends on $\mathcal{O}(1)$ Yukawa couplings in the UV completion. As a result, they could be the smoking gun signal of an FN-like S_1 leptoquark solution to any future excess observed in $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$. Since the experimental precision on both of these (and several other) observables is expected to improve significantly in the near future, we will dedicate the next subsection to discussing potential discovery prospects for this wrinkled FN scenario.

7.4.3 Predictions for Future Measurements

We have already seen that adding wrinkles to a FN ansatz allows for greater flexibility in simultaneously accommodating experimental deviations from the SM while satisfying constraints from other observables and explaining the observed pattern of SM masses and mixing angles. As we will now emphasize, despite this added flexibility, these choices still make concrete predictions for other observables, which can be tested in future experiments. The importance of these tests lies in being able to probe indirect information about the underlying UV model which is hidden in the charge assignments and wrinkles in the IR.

Several upcoming experiments will provide concrete tests of our wrinkled ansatz. When assuming the wrinkled FN ansatz from Eq. (7.32) for the leptoquark Yukawa couplings, several classes of observables—including LFV processes, the electron EDM, meson-mixing measurements, and the decay $K \to \mu\nu$—have a present sensitivity to roughly the same mass scale as $\mathrm{BR}\,(B^+ \to K^+\bar{\nu}\nu)$. Moreover, the mass reach of many of these observables is expected to improve significantly with forthcoming experimental data. Since we have suppressed our model contribution to these observables as far as possible while satisfying the bound from the consistency condition in Eq. (7.16), these correlated signals allow for a definitive test of these types of wrinkled models within the FN mechanism.

At the moment, the strongest bound on LFV processes involving muons is the 90% C.L. limit, $\mathrm{BR}(\mu \to e\gamma) < 4.2 \times 10^{-13}$ set by the MEG experiment [1119]. In the future, however, the most powerful probes of this model will come from searches for $\mu \to e$ conversion in atomic nuclei. As discussed in more detail in Appendix E.2, the conversion rate depends not only on the dipole operator relevant for $\mu \to e\gamma$ and $\mu \to 3e$ decays, but also on four-fermion operators including the first generation quarks generated by integrating out the leptoquark. Future prospects for detecting $\mu \to e$ conversion include the COMET experiment, which will set a limit on the conversion rate of 7×10^{-15} (2.6×10^{-17}) in Phase-I (Phase-II) [1163, 1164], and at Mu2e, which aims at a final sensitivity of 2.5×10^{-18} [1122, 1123][11], both in aluminum nuclei. For more discussion on current and forthcoming searches for LFV, see [1045, 1165–1169].

In the top left panel of Fig. 7.8, we show the predicted $\mu \to e$ conversion rate in aluminum nuclei as a function of the leptoquark mass, with the wrinkled FN ansatz taken for the Yukawa couplings. The $\mathrm{BR}\,(B^+ \to K^+\bar{\nu}\nu)$-preferred region discussed in §7.4.2 is highlighted in red, while the dashed horizontal lines show the future sensitivities for the conversion rate. We see that even Phase-I of the COMET experiment will be sensitive to the mass range preferred by $B^+ \to K^+\bar{\nu}\nu$ measurements, while Mu2e will decisively test all of the relevant parameter space predicted by this model of flavor.

For the electron EDM, the bounds from the ACME II and JILA experiments [1116, 1170] are at the level

[11] This sensitivity might be achievable at Mu2e-II, a proposed upgrade of Mu2e using the PIP-II accelerator at Fermilab, potentially with a target material other than aluminum [1165, 1166].

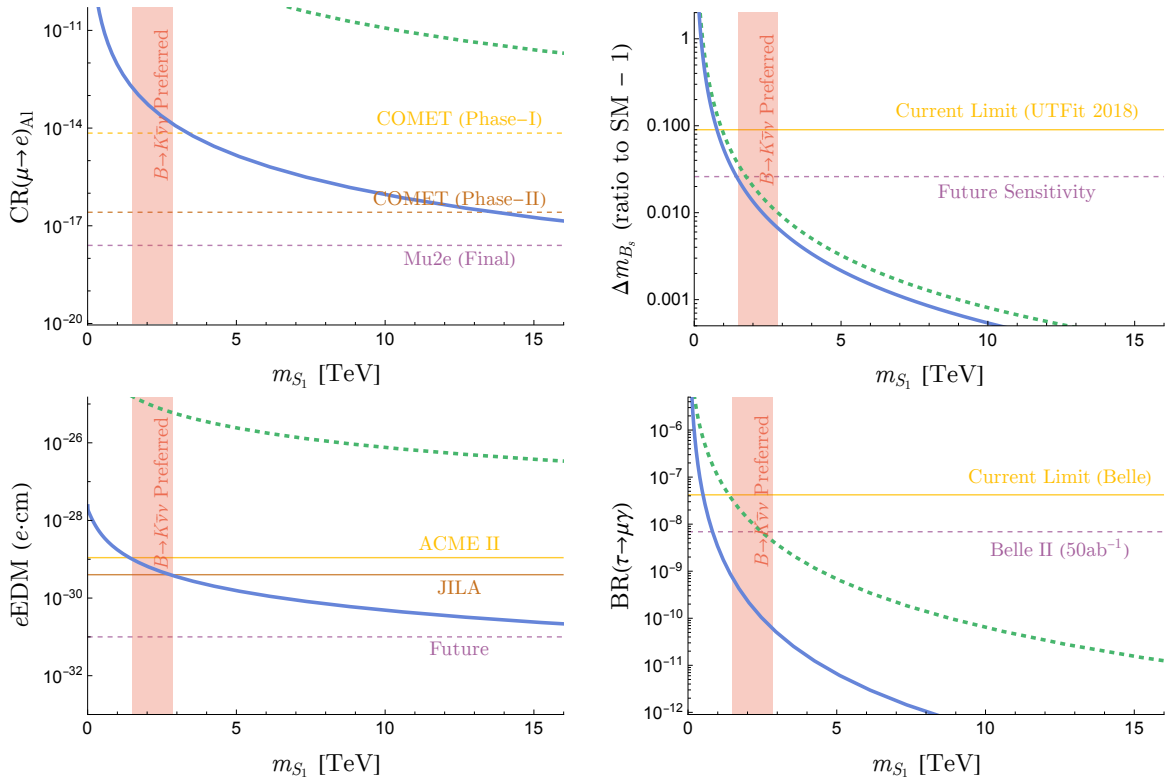


Figure 7.8: Predictions for the S_1 leptoquark contributions to precision observables with the wrinkled (blue, solid) and vanilla (green, dashed) FN ansätze described in §7.4.2. We show the $\mu \to e$ conversion rate in an aluminum nucleus (top left), the electron EDM (bottom left), the relative new physics contribution to Δm_{B_s} (top right), and $\mathrm{BR}(\tau \to \mu\gamma)$ (bottom right), using solid (dashed) lines for current (future) experimental bounds or sensitivity. We do not show the best current bounds on $\mu \to e$ conversion rate, $< 7 \times 10^{-13}$, from SINDRUM II [1121] since it was made with a different nucleus (gold). The red band indicates the mass range of interest for $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$, as in Fig. 7.7.

$d_e < 1.1 \times 10^{-29}$ and 4.1×10^{-30} e cm, respectively. For anarchic flavor couplings, this excludes masses up to $\sim 10^5$ TeV. A vanilla FN ansatz relaxes this constraint to $\sim 10^2$ TeV, and with the additional wrinkles invoked in Eq. (7.32), this bound weakens to $m_{S_1} \gtrsim 2.7$ TeV. Random factors of $\mathcal{O}(1)$, neglected throughout our calculations, can slightly affect the reach on m_{S_1}. The fact that this is the same mass range as favored by $\mathrm{BR}(B^+ \to K^+ \bar{\nu}\nu)$ measurements, and that the reach in m_{S_1} scales faster with improvements to electron EDM measurements compared to other observables, underscores the importance of future electron EDM experiments in probing our model. In the coming years, experimental advances and new technologies promise to increase the sensitivity of EDM experiments by an order of magnitude or more [1117, 1118, 1171]. In the lower-left panel of Fig. 7.8, we show the predicted value of the electron EDM as a function of the leptoquark mass, alongside current bounds and a projected constraint of 10^{-31} e cm, assuming an $\mathcal{O}(1)$ CP-violating phase. As is clear from the figure, future EDM experiments will decisively test this model, up to scales $m_{S_1} \sim \mathcal{O}(10)$ TeV.

For the meson mixing observables, we focus in particular on the neutral B_s meson mass difference, Δm_{B_s},

whose matrix element is directly related to the $B^+ \to K^+ \bar{\nu}\nu$ process for the S_1 leptoquark. To understand the current sensitivity to new physics of $B_s - \bar{B}_s$ mixing, we follow the UTFit analysis [1129, 1172, 1173] and compute the quantity C_{B_s}, defined as

$$C_{B_s} e^{2i\varphi_{B_s}} \equiv \frac{\langle B_s | \mathcal{H}_{\rm mix}^{\rm SM+NP} | \bar{B}_s \rangle}{\langle B_s | \mathcal{H}_{\rm mix}^{\rm SM} | \bar{B}_s \rangle}, \tag{7.33}$$

where $\mathcal{H}_{\rm mix}$ includes the four-fermion operators responsible for $\Delta F = 2$ transitions, as defined in Appendix E.2.6. The SM is defined as the point $C_{B_s} = 1$, $\varphi_{B_s} = 0$, and the allowed size of the new physics contribution is determined by a global fit to the flavor sector, with the range determined primarily by the uncertainties on the input parameters, such as the CKM matrix elements. To be conservative, we consider only the absolute value of the matrix elements above, and avoid making any assumptions about the relative phase between the SM and leptoquark contributions, which is constrained by φ_{B_s}.

The resulting current and future sensitivities (where we assume the current central value is at the SM, for consistency with future projections) are shown on the top right in Fig. 7.8. The projected future sensitivity of $\Delta C_{B_s} = 0.026$ is taken from Ref. [1129], based on projections of HL-LHC results and Belle II results with $50\,\text{ab}^{-1}$ integrated luminosity. We see that the improved sensitivity will start to probe the leptoquark mass range preferred by the $\text{BR}\,(B^+ \to K^+ \bar{\nu}\nu)$ measurements. It is also worth emphasizing that these projections do not account for potential improvements in lattice inputs, and thus could be quite conservative. A statistically significant signal in any of the aforementioned channels would also warrant a much more careful analysis of these B_s-mixing constraints and projections, including phase information that depends in more detail on the flavor ansatz, which could improve sensitivity even further.

A number of additional flavor-changing or flavor-violating decays will be probed with increasing sensitivity at Belle II. A notable example is the LFV decay $\tau \to \mu\gamma$, for which the current bound set by Belle is $\text{BR}(\tau \to \mu\gamma) < 4.2 \times 10^{-8}$ [1124]. Belle II is projected to improve this bound to 6.9×10^{-9} [1125, 1126]. In the lower-right panel of Fig. 7.8, we show the predicted branching ratio of $\tau \to \mu\gamma$ as a function of mass. We see that, for the mass range preferred by the $\text{BR}\,(B^+ \to K^+ \bar{\nu}\nu)$ measurements, the addition of wrinkles in our flavor ansatz suppresses what would otherwise be a predicted signal from assuming the FN mechanism.

Finally, the $K \to \pi\nu\bar{\nu}$ decays, which would rule out the preferred mass range for $B^+ \to K^+ \bar{\nu}\nu$ without wrinkles, have a sensitivity $\sim 1\,\text{TeV}$ in the wrinkled FN ansatz. The $K^+ \to \pi^+ \nu\bar{\nu}$ decay was only recently measured (with a significance of 3.4σ) at the NA62 experiment [1127]. A 10 – 20% precision on this branching ratio is necessary to start excluding $m_{S_1} \sim 2 - 3\,\text{TeV}$, and the requirement for $K_L \to \pi^0 \nu\bar{\nu}$ is similar. Both of these may be achievable with future runs at NA62, or at future experiments planned at the NA62 hall at CERN [1128, 1174] and at J-PARC [1175], and would be an interesting complementary probe of the same physics considered here.

The preceding discussion demonstrates that all of these powerful, forthcoming measurements could have a similar sensitivity to new mass scales for an appropriate choice of wrinkles. Exactly which search channel is ideal depends on the precise pattern of charges and wrinkles in the IR. However, the expectation that we will probe these other correlated signals is relatively robust since the wrinkles in Eq. (7.32) were chosen to saturate the bound in Eq. (7.16) without diminishing the $B^+ \to K^+ \bar{\nu}\nu$ signal. While this enhancement to $B^+ \to K^+ \bar{\nu}\nu$ was only for illustration, and not a fit to a true, significant deviation from the SM, it reveals that for some motivated UV models of flavor, upcoming experiments can simultaneously test explanations for the SM flavor puzzle.

7.5 Discussion

When new physics is embedded in the FN mechanism, the FN ansatz determines the size of both the SM and new physics couplings. In this chapter, we have put forward a systematic extension of this ansatz which can change the expected scaling of the new physics and SM couplings. These changes, referred to as wrinkles, deviate from the FN pattern that is dictated by the horizontal symmetry charges. Wrinkles allow us to demand consistency with other experimental measurements and searches: modifying the relative size of couplings restores some theories that would otherwise be unfeasible due to the correlations between different observables from the FN ansatz. Therefore, they vastly increase the FN mechanism's versatility in accommodating solutions to flavor anomalies. However, owing to radiative corrections, we have also argued that wrinkles can not give rise to arbitrarily large deviations from vanilla FN predictions. There are consistency conditions which must be obeyed by the size of the new wrinkled Yukawas.

While the primary purpose of wrinkles is to give a consistent IR description for various flavor observables, we have also explored how they can be UV completed by various different models. Specifically, in this chapter we have given some simple schematic examples of possible UV realizations. In future work, it would also be interesting to understand more about what patterns of wrinkles can be realistically realized in the UV and the various models that can be used to realize them.

Throughout this chapter, we focused on the phenomenological example of the S_1 leptoquark. We discussed the implementation in the IR when the leptoquark is embedded in a FN model. We also provided a detailed example of how an enhancement of the leptoquark contribution to $\text{BR}\left(B^+ \to K^+ \bar{\nu}\nu\right)$ can consistently respect other experimental bounds, but only if wrinkles are invoked. This wrinkled setup also motivates future measurements, since several signals would be on the verge of discovery in this model, even when the number of wrinkles is enlarged to saturate the simplest consistency condition. In particular, we showed predictions for the most sensitive upcoming probes, namely $\mu \to e$ conversion and the electron EDM.

While we limited our exploration to a specific example with the S_1 leptoquark in this work, it would be

interesting to explore how wrinkles can be applied more broadly. For instance, in our example we fixed the horizontal charges of the SM particles, but there are many other possible choices that reliably yield the SM masses and mixing angles. One could explore how changing the charges affects the correlations and hence the allowed wrinkled ansatz, and see which observables remain correlated to the same mass scale more generally. It would also be intriguing to include other flavor spurions or to add wrinkles to the SM couplings in addition to the new physics couplings. Moreover, it would be useful to do a broad methodical study on the effect of $\mathcal{O}(1)$ numbers in different spurions to explore naturalness in these types of models; see [1081, 1082, 1085] for previous studies of naturalness in such models.

Aside from the flexibility permitted by wrinkles, it is worthwhile to emphasize a separate point about FN models in general: there is more than one charge assignment that can naturally generate the observed SM masses and mixings, beyond just the overall shift in the quark and lepton charges. In particular, we find that the charges of first generation fermions can be either larger than or smaller than other two generations. This is in contrast to a criterion in Ref. [1046], where it was demanded that charges increase monotonically between generations. However, this general FN charge assignment is still not anomaly free and requires some cancellation mechanism, such as Green-Schwarz.

With a number of precision flavor experiments gathering data in the near future that could probe the underlying mechanisms for the flavor structure of the SM, it is the right moment to think about sophisticated UV flavor structures beyond the vanilla FN setup. Wrinkles—a systematic deviation from the vanilla FN prediction for the relationship between different couplings—are one such example that significantly increase the versatility of FN constructions in confronting potential signs of flavorful new physics. We encourage their use in embedding solutions to anomalous signals in UV complete models of flavor.

8

Parameter Inference from Event Ensembles and the Top-Quark Mass

8.1 Introduction

The number one goal of collider physics experiments is to determine the existence and properties of particles in nature. In some rare cases first-principles theoretical calculations can be compared directly to data. More commonly, theory is used to construct sophisticated simulations with adjustable parameters that are then fit to data. Some of these simulation parameters, like coupling constants or masses, have straightforward physical interpretations while other parameters, such as elements of Pythia's string fragmentation model [1176], are required to provide enough flexibility for the data to be described. Often the parameters are highly correlated: varying one can sometimes be entirely compensated by varying another. Typically the uncertainty generated by profiling the unphysical parameters is smaller than other sources of uncertainty, however for precision studies it can be important.

The example of parameter extraction studied in this chapter is the determination of the top-quark mass. The top mass is one of the few parameters in the Standard Model for which a measurement with improved precision is both extremely important and feasible at the LHC. For example, our current best estimate of the lifetime of our metastable vacuum in the universe is limited by precision on top quark mass and the

strong coupling constant [806, 1177–1179]. Moreover, the lifetime is exponentially sensitive to the top quark mass. Using $m_t^{\text{pole}} = 173.1$ GeV, our universe is predicted to last 10^{167} years, but if the top mass were 0.6 GeV higher it would last only 10^{111} years, and if it were 0.6 GeV lower, the universe would last 10^{252} years [1179]. Another example, is searches for certain supersymmetry (SUSY) models, in which stop squarks that are nearly degenerate with the top quark are difficult to constrain because the signal is so similar to $t\bar{t}$ background. This similarity allows the stops to contaminate precision measurements of the top quark, so the consistency of top measurements can be used to search the SUSY parameter space [1180–1185].

It is possible to measure the top-quark mass by direct theory-data comparison, for example through total cross section measurements [1186]. The cross-section approach has two main advantages: it allows for a direct comparison between data and precision theory and the top mass extracted has a clean short-distance definition (typically the $\overline{\text{MS}}$ mass). However, current mass determination by this method has an uncertainty of 1-2 GeV [1182, 1187–1191]. The method for extracting the top-quark mass from LHC data that currently has the smallest uncertainty is fitting the invariant mass peak from the decay products of top quarks in $t\bar{t}$ events [1192–1196]. While such fits typically have errors at the sub-GeV level, there are systematic and theoretical uncertainties associated with such a procedure that are not present in the cross section method. The main complication is that one is more reliant on simulation. For example, there is an uncertainty about how to translate the mass extracted this way, called the Monte Carlo mass, to a scheme like the $\overline{\text{MS}}$ mass which is more theoretically sound.

It is important to separate the challenges in converting between a Monte Carlo mass parameter and a short distance scheme like $\overline{\text{MS}}$ from the extraction of the Monte Carlo mass parameter itself. Typically, the conversion to $\overline{\text{MS}}$ is done by equating the Monte Carlo mass with the pole mass. One could attempt to systematically improve this mapping, for example by comparing precision theory and simulation directly (without data) [1197–1203]. Regardless of how or whether this is done, one cannot hope to begin converting from the Monte Carlo mass to another scheme if different Monte Carlo tunes lead to a different value of the top mass when fit to the same data. Thus, a prerequisite for considering the conversion between Monte Carlo and pole mass is to reduce the tune-dependence of the extracted mass. This reduction is the primary target of this chapter.

The problem of reducing tune uncertainty of the top mass was examined in [1204]. There it was estimated that using classical histogram fitting, the tune-uncertainty on the top Monte Carlo mass was around 500 MeV. This number results from a comparison among the masses extracted using a standard set of tunes in Pythia. It was then shown that the uncertainty could be reduced to 200 MeV by calibrating to the W mass (as is often done by the experiments), and further reduced to 140 MeV by applying soft-drop jet grooming [1205] to the data before fitting. In this chapter, we reproduce the main results of [1204] and

explore whether further reduction is possible using machine learning or with linear regression on ensembles of events. Additionally, we also compare these methods to a profile likelihood fit that is similar to what is currently done in the best experimental measurements.

Using machine learning (ML) to fit a parameter like the top Monte Carlo mass involves complementary challenges to typical collider physics ML applications. Typical collider ML applications such as top-tagging or pileup removal have essentially a right answer: which event was a top and which was background, or what does an event look like with pileup removed? For top mass measurement, there is no right answer: a perfect oracle would not be able to determine the mass from a single event. Instead, only after a collection of events are observed can the top mass be extracted.

There are a number of approaches that have been suggested for learning from ensembles of events. For example, the JUNIPR framework uses a jet-physics inspired architecture to construct the likelihood distribution [42, 258]. This can be done as a function of the top mass, or other training parameters which can then be regressed on data. This application for JUNIPR was suggested in [42] but has not yet been implemented to our knowledge.

Another approach is the DCTR method proposed in [314]. DCTR works by learning the relative weights of a distribution of events as a function of some reference tuning parameters. Then it can be used for regression by minimizing the loss over the tuning parameters to find the best fit. DCTR takes as input events processed through a Particle Flow Network [1206], which is an adaptation of the "Deep Sets" framework developed in [1207] to particle physics. In [314], it was shown to be able to fit simultaneously three Monte Carlo tuning parameters in $e^+e^- \to$ jet events with good results. Thus it is a natural candidate method to test on top-mass extraction where there is a clear metric for what a "good" fit would be. Although the top mass is physical, the top mass parameter in the Monte Carlo can be treated as a tuning parameter and fit in the same way as other Monte Carlo parameters. A discussion of the DCTR method is given in §8.5.

While the DCTR method is promising, it is somewhat cumbersome to implement and train. Moreover, learning the full likelihood ratio as a function of a very high-dimensional input (such as ParticleFlow) may not be necessary if the goal is the regression of a single parameter, like the top mass. Thus we also consider a simpler approach, where ordered sets of high-level observables are input to a dense neural network. We discuss this approach in §8.4. The dense network is very effective, even if the activation functions connecting the nodes are linear. Thus, the entire network is a linear function acting on a sorted ensemble of events. We compare the linear network performance to an ordinary least squares regression, finding similar performance. Moreover, the linear mapping can be examined to see how it depends on the tune and the various elements of the input ensemble. This analysis is also included in §8.4. A summary of our main findings is given in Fig. 8.1.

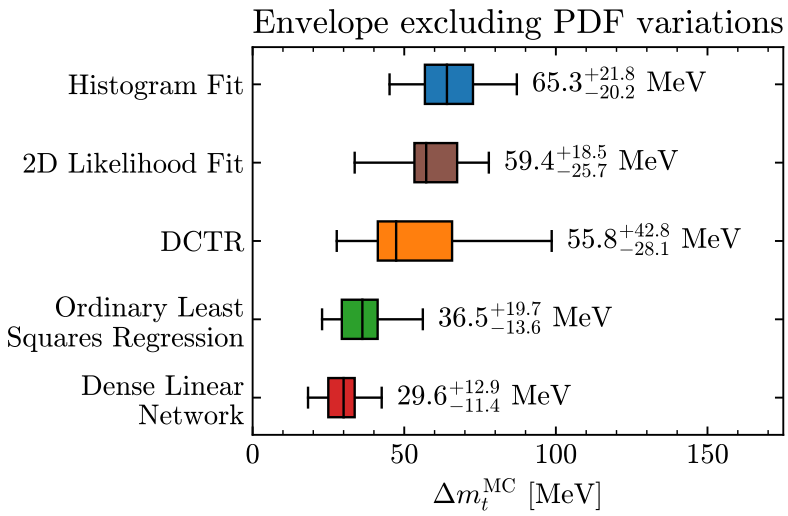


Figure 8.1: Summary of main results. The bars show the estimate of the Monte-Carlo tuning uncertainty in top-quark mass extraction from top events at the LHC. The errors on the uncertainties include uncertainties from training and statistical variations. The top row is a histogram fit, using soft drop and normalizing to the W-mass (following [1204]). The second row uses a 2D likelihood fit to profile over correlations between the top and W masses. The third row uses the DCTR technique of [314]. The fourth row is an ordinary least squares linear regression on an ensemble of 30,000 events. The fifth row shows the result of using a linear network for regression, also on an ensemble of 30,000 events. Numbers here correspond to the total envelope excluding PDF uncertainties, as in Fig. 8.14.

The chapter is organized as follows. Event generation and general elements of our fitting procedure are discussed in §8.2. §8.3 describes our implementation of classical fitting methods that are similar to what is often done in experimental work, including a histogram fit modeled after [1204] and a 2D profile likelihood fit, to benchmark our samples and fits. §8.4 discusses the regression approach, using both a linear network and an ordinary least squares regression. §8.5 discusses the DCTR approach of [314]. Our conclusions and a discussion are given in §8.6.

8.2 Event Generation and Uncertainty Estimation

For this study, events with pairs of top quarks are produced using Pythia 8, including both $q\bar{q}$ and gg production channels in $\sqrt{s} = 13$ TeV proton-proton collisions. We restrict to semi-leptonic events with $t \to b\ell^+\nu_\ell$ and $\bar{t} \to \bar{b}q\bar{q}'$, where ℓ stands for electrons or muons. The events are showered to final state particles, which are then clustered into jets using the anti-k_t [1208] algorithm in Fastjet [1209] with R=0.5. For simplicity, we do not attempt to include any realistic detector effects or experimental efficiencies. Thus, we mark any jet within $\Delta R < 0.4$ of one of the b's from the top decays as a b-jet and assume the lepton is correctly tagged. A more realistic study would of course need to incorporate b-tagging, jet energy scale resolution, pileup, backgrounds, and so on. Each of these effects will necessarily increase the top-quark mass

uncertainty. However, since our goal is mainly to understand the relative performance of different ensemble regression methods, we do not believe our simplifying assumptions will affect the qualitative conclusions.

The event selection is as follows. We require a final state ℓ with $p_T^\ell > 20$ GeV and $|\eta| < 2.4$. We only keep jets if they have $p_T^j > 30$ GeV and $|\eta| < 2.4$ and demand that there are exactly 2 b-tagged jets and at least 2 un-tagged jets. The invariant mass of pairs of un-tagged jets is scanned to find the pair with a mass closest to $m_W = 80.3$ GeV. If this two-jet invariant mass, m_{2j} is not within $(70\text{ GeV}, 90\text{ GeV})$, the event is discarded. Next, we find the three-jet invariant mass for the two jets of the W and b-tagged jet coming from the $\bar{b}$. This is overly simplified and ignores combinatoric background.[1] However, we take a tight cut on the three-jet invariant mass, and only accept events with $150\text{ GeV} < m_{3j} < 200\text{ GeV}$[2], which reduces such contamination. This still allows for a comparison of the different methods.

The uncertainty in the regression of the Monte Carlo top mass from each of the methods is computed using the A14 Pythia 8 Tunes of the ATLAS 7 TeV data [1211]. The 14 tunes cover 4 different families of variations: VarPDF, Var1, Var2, and Var3. VarPDF covers variations in the parton distribution functions with `tunepp:19-22` corresponding to the CTEQL1 [1212], MSTW2008LO [1213], NNPDF2.3LO [1214], and HERAPDF1.5LO [1215] PDFs, respectively. Var1, Var2, and Var3 all use the NNPDF2.3LO [1214] PDF but vary other parameters, with Var1 covering underlying event effects, Var2 accounting for jet substructure, and Var3 covering different aspects of extra jet production. Var3 includes three separate variations (Var3a, Var3b, and Var3c) since it could not be reduced to a single pair. The tuning parameters for all A14 variations are shown in Table 8.1.

There are of course many more tunes one can consider. But again, since the main purpose of this study is to compare the relative strengths of different approaches, not to produce a final numerical value of the uncertainty, we believe this set should be sufficient.

We attempt as much as is possible to use the same fitting procedure to compare different methods. In all cases, after a method is fit or trained, it provides a mapping from an ensemble of events to a regressed mass. To assess the uncertainty of the regressed mass, we first assess its variation for fixed Monte Carlo mass within each tune family. We denote the maximum, minimum and mean regressed mass within the family for the fixed mass by m_{fit}^{max}, m_{fit}^{min}, and $\bar{m}_{fit}$, respectively. We compute the uncertainty for the given m_t^{MC} and tune family as

$$\Delta m_t^{MC} = \frac{1}{2}\left(m_{fit}^{max} - m_{fit}^{min}\right)\frac{m_t^{MC}}{\bar{m}_t^{fit}}. \tag{8.1}$$

The factor of $m_t^{MC}/\bar{m}_t^{fit}$ reflects that the fit mass (especially in the histogram fit with soft drop) can be

[1] Ref. [1210] introduces a machine learning method to identify the correct combination of jets in $t\bar{t}$ without the factorial scaling of scanning each combinatorial permutation.

[2] In the dense network section, we also generate events without this cut to see the effect of the m_{3j} range on Δm_t^{MC}.

Variation	Tunes	ColorRec	α_S^{MPI}	p_{T0}^{Ref}
VarPDF	19-22	1.71	0.126	1.56
Var1	21, 23, 24	[1.69,1.73]	[0.121,0.131]	1.56
Var2	21, 25, 26	1.71	0.126	[1.50,1.60]
Var3a	21, 27, 28	1.71	[0.125,0.127]	[1.51,1.67]
Var3b	21, 29, 30	1.71	0.126	1.56
Var3c	21, 31, 32	1.71	0.126	1.56

Variation	$p_T^{dampFudge}$	α_S^{FSR}	$p_T^{maxFudge}$	α_S^{ISR}
VarPDF	1.05	0.127	0.91	0.127
Var1	1.05	0.127	0.91	0.127
Var2	[1.04,1.08]	[0.124,0.136]	0.91	0.127
Var3a	[0.93,1.36]	[0.124,0.136]	[0.88,0.98]	0.127
Var3b	[1.04,1.07]	[0.114,0.138]	[0.83,1.00]	[0.126,0.129]
Var3c	1.05	0.127	0.91	[0.115,0.140]

Table 8.1: Table shows the relevant parameters for the A14 tune variations. Var1 through Var3 tunes are listed in order of central, +, then -. The relevant tuning parameters are the strength of the color reconnection (`ColourReconnection:range`), the strong coupling constant for multiparticle interactions α_S^{MPI} (`MultipartonInteractions:alphaSvalue`), the initial state radiation (ISR) p_T cutoff p_{T0}^{Ref} (`SpaceShower:pT0Ref`), the factorization/renormalization scale damping $p_T^{dampFudge}$ (`SpaceShower:pTdampFudge`), the strong coupling constant for final state radiation (FSR) α_S^{FSR} (`TimeShower:alphaSvalue`), the multiplicative factor on the max ISR evolution scale $p_T^{maxFudge}$ (`SpaceShower:pTmaxFudge`), and the ISR strong coupling α_S^{ISR} (`SpaceShower:alphaSvalue`).

linearly offset from the true Monte Carlo mass.

An example of this procedure is shown Fig. 8.2 for the Var1 tunes. The blue, orange, and green data points denote the fits from the central, +, and − variations, respectively. The x-axis shows the true Monte Carlo mass of the sample and the y-axis gives the fit value. We compute Δm_t^{MC} for five different values of the top mass: m_t^{MC}=172.0, 172.5, 173.0, 173.5, and 174.0 GeV. The spread between the maximum and minimum fit mass at each point is marked, and the average mass is reported along with Δm_t^{MC} for each Monte Carlo mass. Note that the value of Δm_t^{MC} is different for each m_t^{MC}.

In order to get a statistical estimation of the the uncertainty, we repeat the analysis on the same five masses using four more independent data sets generated with new random seeds. The maximum and mean uncertainties from the 25 samples (5 masses times 5 data sets) are presented in the following figures. To visualize these uncertainties, we show box-and-whisker plots. These start by placing a box covering the 25th-75th percentiles of the Δm_t^{MC} values. The whiskers then extend as a line out to the maximum and minimum, unless these are further away from the box than 1.5 box lengths, in which case the points are considered to be outliers and denoted by open circles. The line in the box denotes the median. An example of this statistical estimate is also shown in Fig. 8.2.

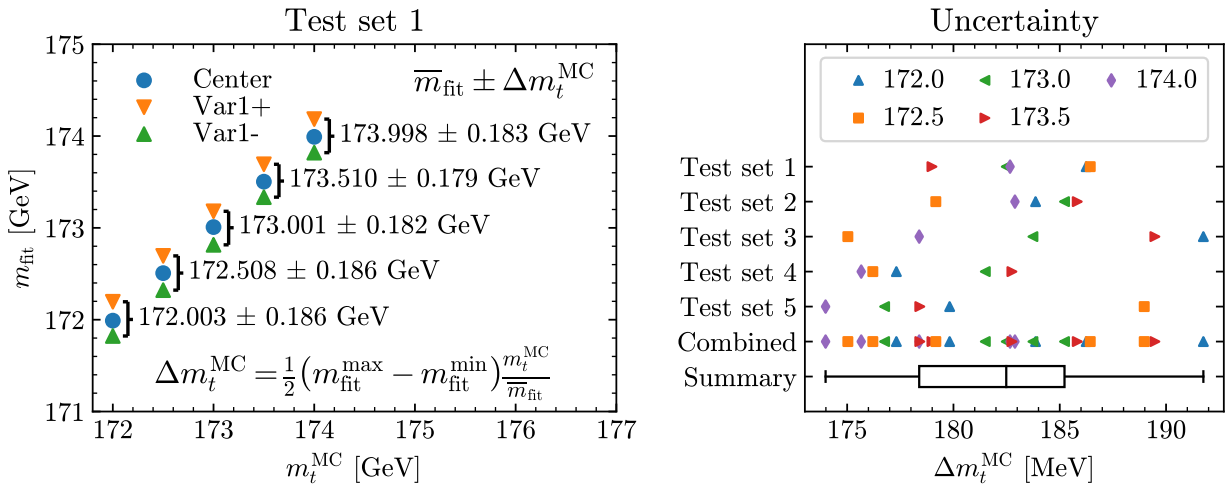


Figure 8.2: How uncertainties on m_t^{MC} are estimated. These particular numbers are from the uncorrected histogram fitting method using the Var1 tune, but the same error estimation is used throughout. Left: we show the fitted mass for different truth m_t^{MC} values and + and − variations of the tune parameters. For each m_t^{MC}, the uncertainty Δm_t^{MC} is computed using (8.1). We repeat the fits with 5 test sets. Right: the values of Δm_t^{MC} for each test set and each mass are shown. The markers correspond to specific m_t^{MC} samples. The distribution of these uncertainties are shown in the box-and-whiskers plots ("summary" row of right panel).

8.3 Classical Fitting Methods

In order to benchmark the tune uncertainties for the top mass, we first implement two template fitting procedures modeled roughly on what is often done for actual experimental data.

8.3.1 Histogram Fitting

We employ an iterated Gaussian histogram fit, similar to that used in [1204]. For each test set at each tune and mass, we create a histogram of the three-jet invariant mass, m_{3j}, using anti-k_T R = 0.5 jets. We then fit a Gaussian to the distribution along the full range 150 GeV $< m_{3j} <$ 200 GeV. The fit range is then adjusted to include one standard deviation on either side of the mean of this Gaussian and a new Gaussian is fit to this new range. We continue to iterate this fitting procedure until the mean and width of the Gaussian converge to stable values. The mean is then used as the fitted top mass m_t^{fit} and the width discarded. The left panel of Fig. 8.3 depicts this procedure. We also tested iterated fits of different functions such as a crystal ball function and a skewed gaussian, but do not display the results since they do not improve the top mass fit compared to the Gaussian case.

The top mass m_t^{fit} extracted from this method is very nearly linearly proportional to the top Monte Carlo mass m_t^{MC}. This linear fit is shown on the right panel of Fig. 8.3. We then use the fitted mass and Monte Carlo mass to compute an uncertainty as described in the previous section.

We consider three variants of this method, again following [1204]. We first fit directly to the 3-jet mass

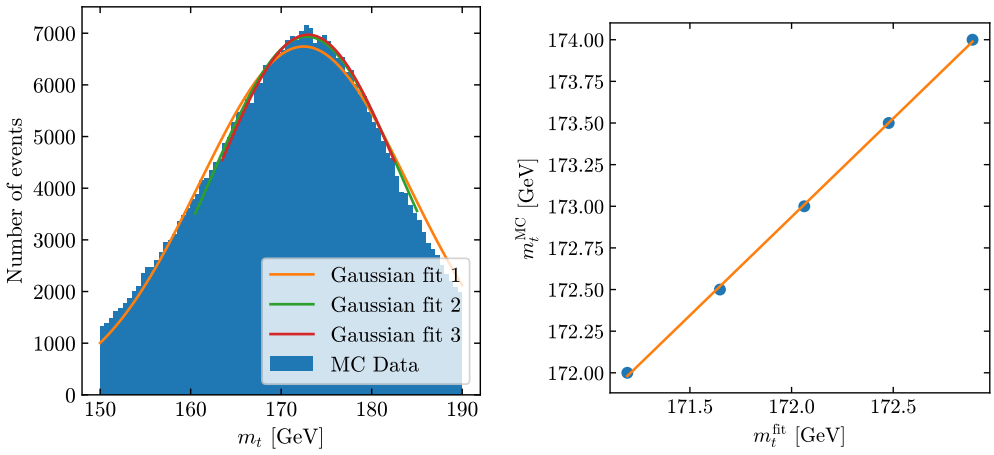


Figure 8.3: Shows the iterative fitting procedure used to fit a top mass to the 3-jet mass distribution. The left panel shows the distribution and several fits. In each iteration the fit range is adjusted to include one standard deviation on either side of the mean of the previous fit. The right panel demonstrates the linear relation between the fitted values of $\mathrm{m_t^{fit}}$ from tune 21 data and the Monte Carlo mass $\mathrm{m_t^{MC}}$ used to generate the events. The fit ranges shown are $150 - 200$ GeV, then $159.4 - 185.6$ GeV and finally $162.3 - 183.1$ GeV.

histograms. Second, we calibrate to the W mass. To do this, we rescale the 3-jet mass so that the 2-jet mass, $\mathrm{m_{2j}}$, is equal to the W mass:

$$\mathrm{m_{calibrated}} = \mathrm{m_W} \frac{\mathrm{m_{3j}}}{\mathrm{m_{2j}}}. \tag{8.2}$$

Finally, we apply the soft drop algorithm [1205] with parameters $\beta = 0, 1, 2$ on the jets before computing the histogram.

A summary of the resulting uncertainties from the histogram-fitting approach is presented in Fig. 8.4. We find that the best variant, including both W calibration and soft drop with $\beta = 0$, yields a mean total envelope uncertainty of about 65 MeV and an uncertainty of about 100 MeV when the variations are added in quadrature. This is roughly consistent with the values in [1204], and similar to values found by ATLAS [1194].

8.3.2 Profile likelihood fitting

The histogram fitting method does not easily extend to more than one observable and does not include information from every event. A second method which is used by the experiments is to perform a profile likelihood fit. The idea behind this method is to find the mass which is most likely to have generated the observed events. To do so, a likelihood function is used to model the distribution. The likelihood function is able to incorporate more than one observable, allowing for more flexibility than the histogram fitting.

Here, we model the top and W resonances as Gaussian distributions. The mean value of the top distribution will be fit, and the mean of the W distribution is set to 80.3 GeV. The standard deviations of the distributions are determined from fitting the resonances across all tunes simultaneously and are

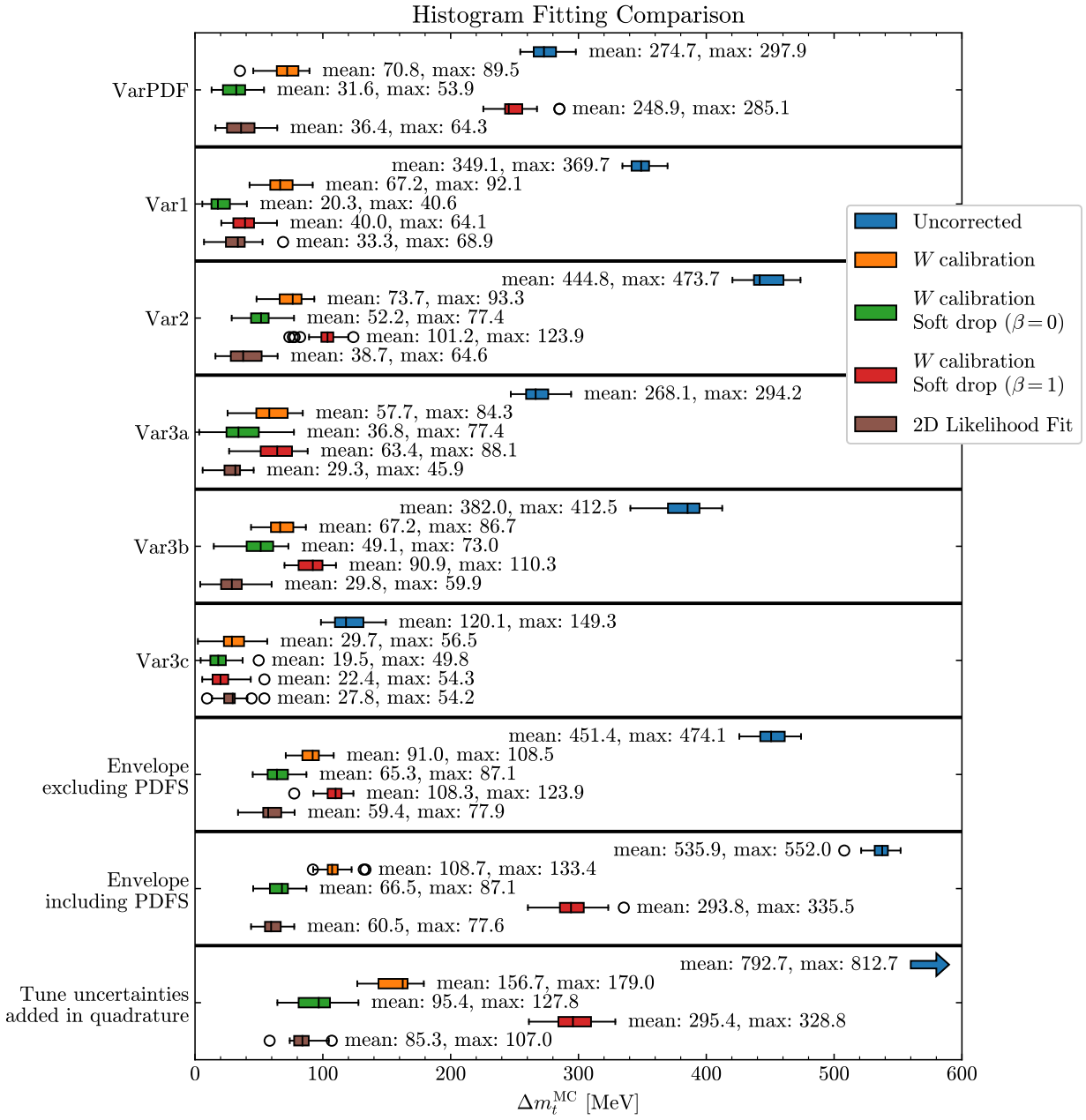


Figure 8.4: Box-and-whisker plot for the uncertainties on the top mass histogram fit and profile likelihood fit within each group of variations in the A14 set of tunes, as in Fig. 8.2. The open circles denote outliers which are further away from the box than 1.5 box lengths. The uncertainties are calculated with and without W calibration and soft drop grooming methods. The bottom three rows show several ways of combining the error for the different variations.

$\sigma_t = 6.5$ GeV and $\sigma_W = 3.5$ GeV. In addition, we include a nuisance parameter, c, in the model to help account for fluctuations in the ratio of the reconstructed top and W masses coming from differences in the tune parameters. Explicitly, the likelihood is given as

$$\mathcal{L}\left(m_t^{fit}, c\right) = \prod_{i \in Events} \left(\mathcal{G}\left(m_{3j}\, c \,|m_t^{fit}, \sigma_t\right) \mathcal{G}\left(m_{2j}\, c \,|m_W, \sigma_W\right) \mathcal{G}\left(c \,|1, \sigma_c\right) \right), \tag{8.3}$$

where $\mathcal{G}(x|\mu, \sigma)$ is the probability density evaluated at x for Gaussian distribution with mean μ and standard deviation σ, and σ_c is the standard deviation of the fitted value of the m_{2j} peaks across a range of tunes. We use $\sigma_c = 0.13$.

For a set of events with a fixed tune, the value of the top mass is extracted by maximizing the likelihood function over both m_t^{fit} and c. The value of m_t^{fit} that maximizes the likelihood does not equal m_t^{MC}, but is linearly correlated. The linear relation between m_t^{MC} and m_t^{fit} is used for the inference of top mass.

Fig. 8.4 shows the results using this method as the brown bars denoted by "2D Likelihood Fit". Overall, this method does similar to the histogram fitting with grooming and calibration, even though these are not done explicitly here. The likelihood fit improves the mean Δm_t^{MC} by around 10% when taking the envelope of the tunes or adding the uncertainties in quadrature. This improvement comes as a result of using the values of m_{2j} and m_{3j} from every event and including a nuisance parameter. In principle, it is possible to implement a nuisance parameter for each of the tuning parameters, but this is challenging in practice, as the effects of each tuning parameter may not be well modeled by a Gaussian. Instead, we advocate for the method presented in the next section, which still includes the values of m_{2j} and m_{3j} from every event, but allows for a flexible function–unlike the fixed form of Eq. (8.3)–which uses changes in the distributions to account for tune variations.

8.4 Regression on Sorted Ensembles

In this section, we study whether doing regression on ensembles of events can improve on the traditional template histogram fit. We consider both using a dense neural network (DNN) to do the regression with machine learning, and alternatively an ordinary least squares (OLS) linear regression. The two methods give comparable results. The DNN with linear activations is slightly better, but the OLS regression is simpler and faster (but uses more memory).

8.4.1 Inputs

We take as inputs to the regression ensembles of observables computed from simulated events, sorted by one of the observables. We use these sorted ensembles as inputs to regress out the top-quark Monte Carlo mass, m_t^{MC}.

For training, we use events simulated with m_t^{MC} ranging from 170 to 176 GeV in intervals of 0.2 GeV. For each mass we generate 300,000 training events for each of the A14 tunes. We have also tested using a larger total number of events in each sample and finer spacings between the masses, but this does not improve our results. There is no apriori reason why a uniform prior necessarily gives the best performance, but we find it to be sufficient for our purpose. To train the regression, we use a random ensemble of 30,000 events (with replacement) from the total set of 300,000 at a given tune and m_t^{MC}. The number 30,000 is chosen because taking a smaller number of events per ensemble gives a larger error, while taking a larger number is prohibitively slow (at least in the DNN case) and does not lead to noticeable improvement. Both the DNN and OLS regression see many different ensembles from each training sample, but the total number of ensembles and which samples they are from differs between the two regression methods. For the DNN, batches of 100 ensembles are seen in each training step, and each ensemble is from a randomly selected mass and tune. In contrast, the best OLS regression uses 20 ensembles for each mass and tune.

The basic observables we consider are the 3-jet invariant mass (i.e. top), m_{3j}, the 2-jet invariant mass (i.e. the W boson mass), m_{2j}, and their ratio $R_{32} = \frac{m_{3j}}{m_{2j}}$ [1216, 1217]. The inputs to the regression are the values of these observables, sorted by one (or more) of them. Sorting the ensemble is important because it determines which parts of each observables' distribution the different weights are applied to, and allows the regression to exploit correlations in different observables across tunes. Example input distributions sorted by m_{3j} are shown in Fig. 8.5. We tested different orderings and several different observables as inputs (discussed more in section 8.4.2), and find that sorting by increasing m_{3j} tends to give the best results.

To extract the uncertainty from the regression, we generate five more statistically independent samples of 400,000 events at each mass between 172.0 and 174.0 GeV (in intervals of 0.5 GeV) and for each of the A14 tunes. From each of these test samples (at fixed mass and tune), we take an ensemble of 30,000 events and evaluate the network to get an output value. We repeat this 100 times for each sample and take the mean of those values to get a final predicted value for a given trial, mass, and tune. We then use those predicted values to compute the error as described in §8.2.

8.4.2 Dense Network

First, we discuss using a linear network. We use a two layer network implemented in keras [1218], with one node in each layer and linear activation functions between nodes.[3] This is shown in Fig. 8.5. While more than one layer is not strictly necessary since our activation functions are linear, additional layers can help with training and hyperparameter optimization. We also tested more complicated neural networks with different filter configurations (including deeper networks), removing various node connections, and nonlinear

[3]The exceptions to this are the R_{32} only networks, which train better when we use a third layer.

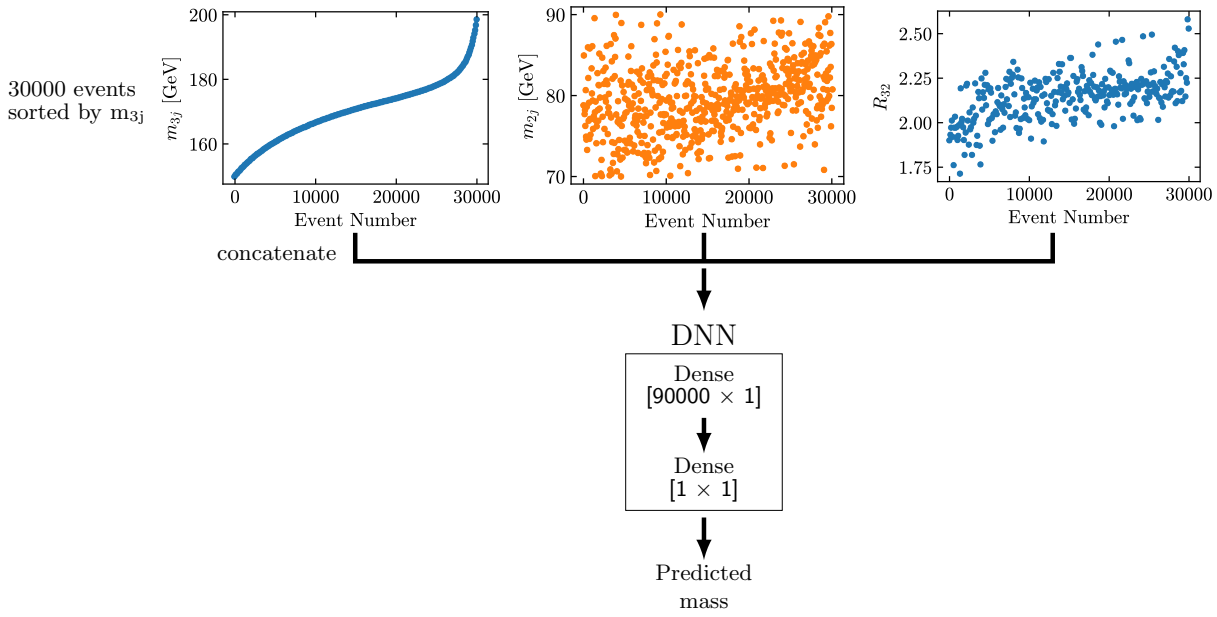


Figure 8.5: Example DNN and its inputs. Example inputs are m_{3j}, m_{2j}, and R_{32} for 30,000 events, sorted according to increasing m_{3j}.

activation functions, none of which seemed to improve performance. Networks were trained with the Adam algorithm [1219] for 600 epochs of 750 steps each, with an early stopping patience of 60 and a batch size of 100. The initial learning rate was 0.0005, with a learning rate decay of 0.7 after 8 epochs without improvement. We did not exhaustively optimize these hyperparameters, so it is possible that there would be further performance gains with a more systematic hyperparameter search. We also normalize all inputs by subtracting a constant so that the mean of each sorted ensemble is small compared to its spread, which helps ensure consistent results when the network is trained multiple times. This amounts to subtracting 173 GeV from m_{3j}, 80 GeV from m_{2j}, and 2 from R_{32}. We also tested other normalization methods, but found they do not improve performance noticeably. We tested several loss functions and determined that the network is mostly insensitive to which loss function was used and performs equivalently for loss functions such as logcosh and mean squared error. The results presented use the logcosh loss.

We tested multiple different sets of observables as inputs. We considered including combinations of m_{3j}, m_{2j}, R_{32}, and $m_{\ell b}$ (the invariant mass of the lepton and b-quark on the leptonically decaying top quark side of the event). For each of these inputs, we considered sorting the ensembles in different ways before putting them into the network. Sometimes, we sorted the events the same way for all the observables, and sometimes we sorted the events differently for different observables. In either case, the orderings were determined by making sure the sorted variables were strictly increasing, and then applying one of these orderings to each of the other, unsorted variables. We find sorting is necessary to train the network, and that the results depend

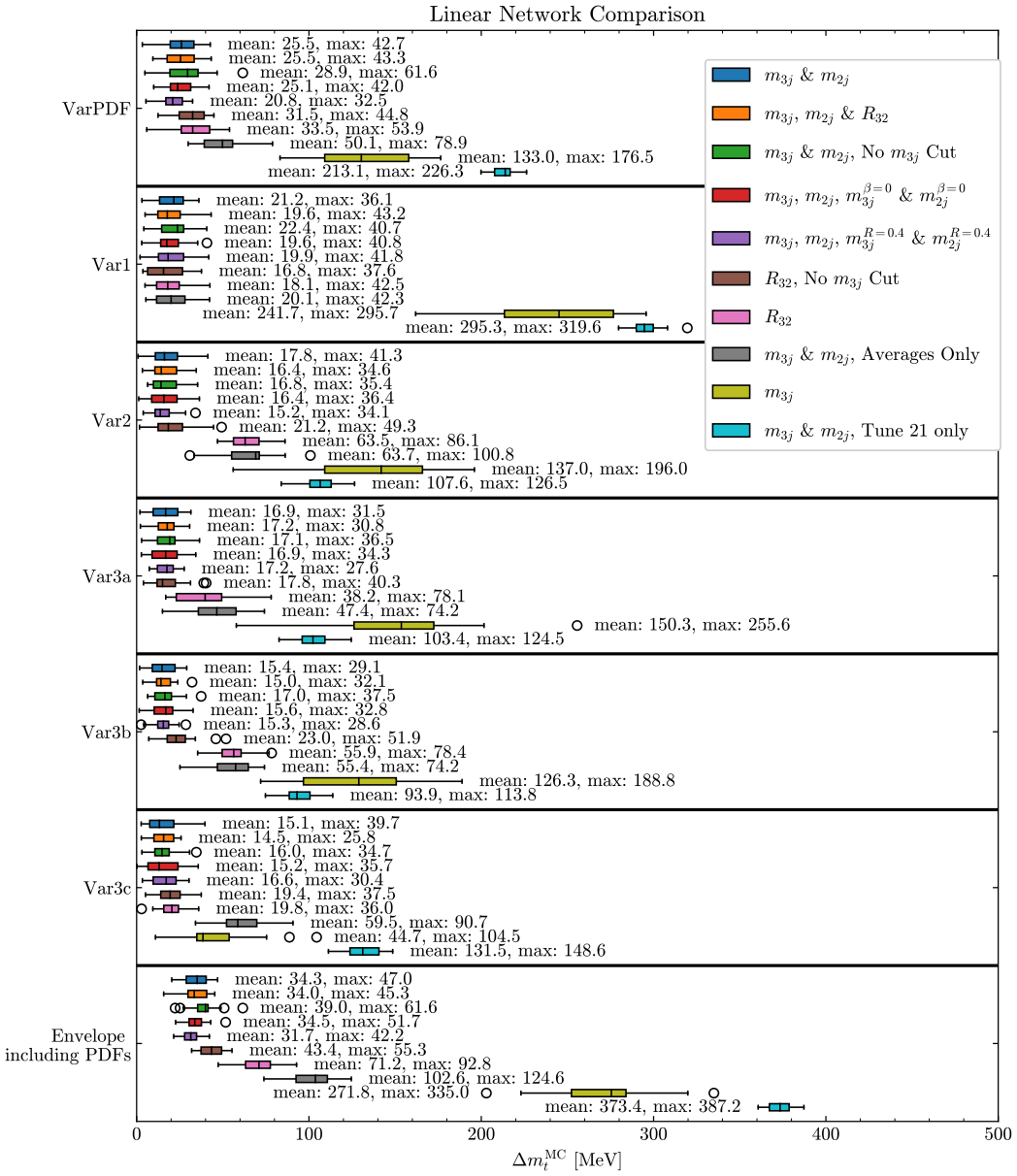


Figure 8.6: Uncertainties on the top mass linear network fits within each group of variations in the A14 set of tunes, with different observables as inputs. The Total Envelope section contains the envelope of all tunes. For every network displayed (except that trained only on the means of the distributions and the soft drop example), the distributions are sorted by increasing m_{3j}. For the soft drop example, the distributions are sorted by m_{3j} with $\beta = 0$ rather than the original m_{3j}. We have restricted m_{3j} to between 150-200, except for those networks labeled "No m_{3j} cut".

on how the inputs were sorted. Generically, we find that sorting by m_{3j} works best.

Additionally, we recomputed these observables for subjets determined in different ways and tested combining these subjet observable ensembles with those for the original jets. We used subjets obtained by applying soft drop [1205] with $z = 0.1$ and $\beta = 0, 1, 2$ to the initial jet, as well as telescoping subjets at different radii [1220–1222]. We found that our results are not sensitive to the value of z used in soft drop as long as it is small enough. Changing β has a small effect which is not noticeable in the best case of m_{3j} & m_{2j} inputs. For single observable networks, $\beta = 0$ often does best. For both types of subjets, our results depend on the specific network inputs, but none of these networks perform noticeably better than the best network without subjet observables.

A subset of our results for various different inputs is shown in Fig. 8.6. As previously mentioned, a combination of m_{3j} and m_{2j}, both sorted by increasing m_{3j}, is sufficient to give our best results. This is shown in blue. From the figure, we can see that most of the contributions from the different variations are a similar order of magnitude, in contrast to the histogram fitting case. The largest contributions to the error are from Var1 and VarPDF.

For completeness, Fig. 8.6 contains several other results, considering both different input variables and different m_{3j} ranges. We find many networks with additional variables perform similarly to the m_{3j} & m_{2j} combination (which can be seen by comparing the blue errors to the orange and green ones), while networks that do not include both m_{3j} and m_{2j} tend to perform worse. Like in the case of the histogram fit, using mass alone (shown in yellow) gives the worst results, while R_{32} alone (shown in pink and brown) improves upon the mass, though both do better than the histogram fit when the same inputs are used.[4] Examples including soft drop variables and subjets at different radaii are shown in red and purple respectively. Including soft drop variables or subjets at smaller radii can help when compared to networks trained on single variables, but there is no further improvement on the m_{3j} & m_{2j} combination. For reference, we also include two other networks that do not use the full distributions for every tune as input. In gray, we show the case of taking the average of the ensemble before inputting to the network. In turquoise, we show a network trained on tune 21 only (but still tested on all the A14 tunes). Unsurprisingly, we find that in both of these cases the networks perform worse than the full sorted ensembles marginalized across tunes.

Next, we would like to understand why these networks are able to perform better than the histogram fit. In the histogram fit, the mass is given by the center of a Gaussian which is similar to our average-only network. Therefore, we look for improvements over the average-only network as a proxy for understanding why the dense network does better than the histogram fit. In order to understand how different parts of the

[4]The improvement between R_{32} and using multiple observables is dependent on the range of m_{3j} used; for m_{3j} between 150-200 the difference is larger than when the full m_{3j} distribution is used. This can be seen in the difference between the pink and brown errors.

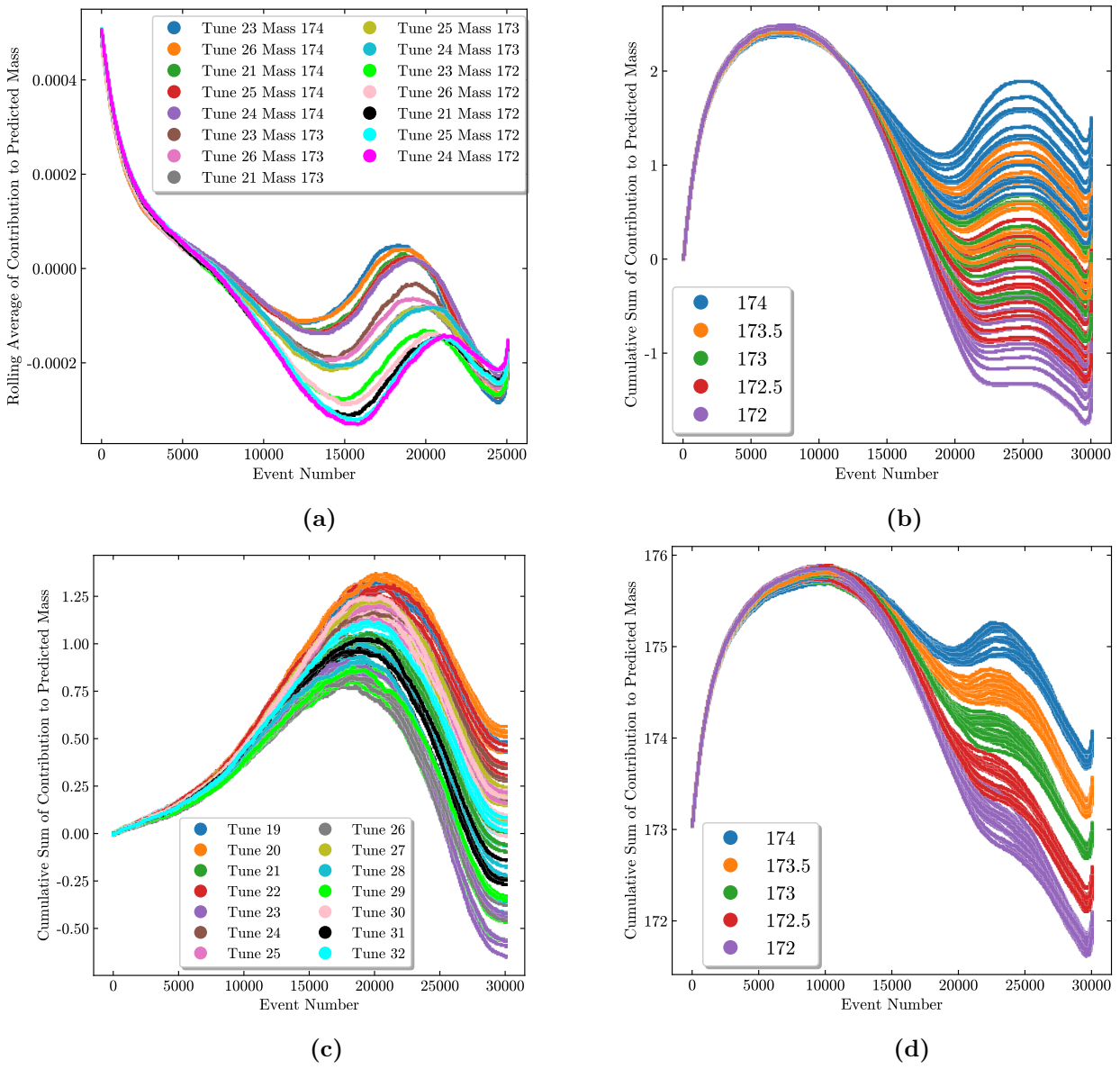
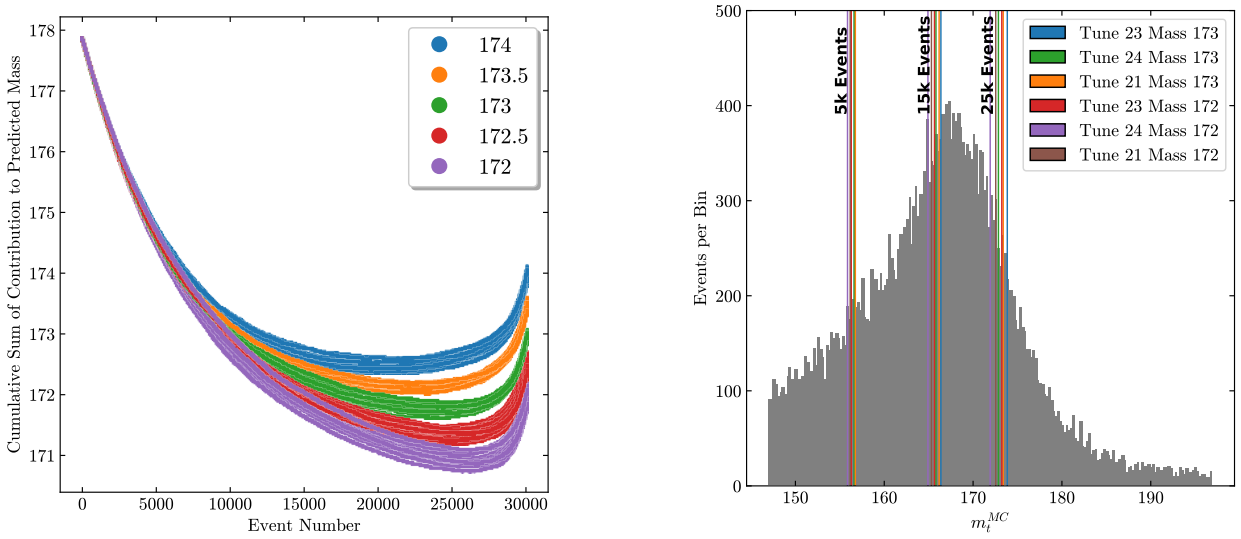


Figure 8.7: Graphical representations of the contribution to the predicted mass as a function of entry number in the ensemble using the m_{3j} & m_{2j} network. In (a), we show a rolling average of the product of the (normalized) m_{3j} input multiplied by the network weight for each event number. Each color denotes a different mass and tune. Only a subset of masses and tunes have been shown for clarity. In (b)-(d), each point is the cumulative sum up to that event number of the (normalized) input value multiplied by the network weight. (b) includes only the m_{3j} contribution, with colors denoting masses and each mass including 1 sample from each of the A14 tunes. (c) includes only the m_{2j} contribution, with colors denoting tunes. One sample at each mass is included for each tune. (d) sums over both m_{3j} and m_{2j} with colors again denoting masses. The bias is included in the zeroth entry.



(a) **(b)**

Figure 8.8: (a): Graphical representation of the contribution to the predicted mass as a function of entry number in the sorted ensemble using the network that depends only on the average of the ensemble. The zeroth entry includes the bias. Each point is the cumulative sum up to that entry of the (normalized) input values multiplied by the network weight and divided by the total number of samples. Color denotes the mass. There are fourteen different lines for each color; one example for each tune. (b): Histogram showing an example distribution. The solid lines drawn show 5k, 15k, and 25k events for several masses and VAR1 tunes.

ensembles contribute, we examined the weights of the networks.

We use these weights to construct Figs. 8.7 and 8.8a. Since these weights are shared by all masses and tunes, we multiply the weights by example input ensembles to construct the plots. Specifically, we plot a rolling average of the input times the weights (as in Fig. 8.7a) or a cumulative sum of the inputs times the weights (as in Figs. 8.7b-8.7d and 8.8a) as a function of event number. For Figs. 8.7d and 8.8a, which include all input observables, we have also added the constant bias learned by the network and accounted for the normalization of the labels by adding 173 to the predicted outputs. In these two plots, we can read off the predicted m_t^{MC} value from event number 30,000. We include Fig. 8.8b for reference, to see which parts of the top mass distribution are contributing the most in each network.

In general, we want to design a procedure that is sensitive to the Monte Carlo mass but not the tuning parameters. The difficulty with this is that most variables that are strongly affected by m_t^{MC} (such as m_{3j}) are also strongly affected by the tuning parameters, which we can see from Fig. 8.7b. This can be partially corrected by including other variables (such as m_{2j}, seen in Fig. 8.7c) which are more sensitive to the tune than the MC mass. When we just fit the mean of each distribution, there is not much more that we can do, aside from trying to clean up the distributions themselves. However, in the case of directly inputting a sorted ensemble into the regression, the network can look for other combinations that are less sensitive

to the tune than the mean. The network can learn to use a particular part of an observables' distribution to differentiate the masses, and a different part to partially correct for the difference in tunes. This can be seen in Fig. 8.7a, where the middle of the ensemble distinguishes the mass, whereas the upper tail is more strongly correlated with the tune. We can also see this from comparing the average only network in Fig. 8.8a with the full $m_{3j}\&m_{2j}$ network in Fig. 8.7d. In 8.8a, most of the difference in masses comes from events between 5,000-15,000, where the original distributions differ most, and the width of each mass band in the upper half of the ensemble remains mostly constant. In contrast, in 8.7d, most of the difference in regressed mass comes from event numbers greater than 15,000, and the difference in tunes shrinks substantially at the top tail of the m_{3j} ensemble.

For completeness, we also tried generating new samples uniformly spaced in the other tuning parameters and regressing out these tuning parameters in addition to the mass. For this test we restricted to the VAR1 tunes, but an equivalent test could be conducted across all variations. We might think this type of network would improve our results since the loss function explicitly depends on tuning parameters in addition to m_t^{MC}. However, we found that in the case of a linear regression, a multidimensional output did not help improve the predicted top mass (in contrast to what we found with the DCTR method, discussed in §8.5). In particular, we find that sorting the inputs encodes enough information about the other tuning parameters that additional outputs are unnecessary. This can be seen from the solid lines in Fig. 8.8b. While the value of the 15,000th event near the peak depends primarily on the Monte Carlo mass, the value of the 25,000th event is also strongly dependent on the tune.

8.4.3 Ordinary Least Squares Regression

Since non-linearities and a deep network structure do not seem to improve results, it is natural to ask if we can reproduce the same results with something simpler. Therefore, we test the case of using a projection matrix to do the ordinary least squares linear regression exactly (rather than using the Adam algorithm to do the minimization). We implement this regression using scikit-learn [1223], and consider three separate cases: using the full ensembles of 30,000 events, using the means across the ensemble only, and using the ensemble means and their standard deviations. For the full ensembles we use 20 random samples from the 300,000 event training sets with each mass and tune; for the other cases we resample 200 times from each mass and tune. We find that with full ensembles the results are worse when fewer samples are used, but including more than 20 samples decreases performance. We suspect this is due to overparameterization, but that it could be improved through regularization or dimensionality reduction techniques. Additionally, the matrix operations in the OLS regression become memory intensive (using over 32 GB) with many samples. The other regression methods are mostly insensitive to the number of samples, as long as there enough for

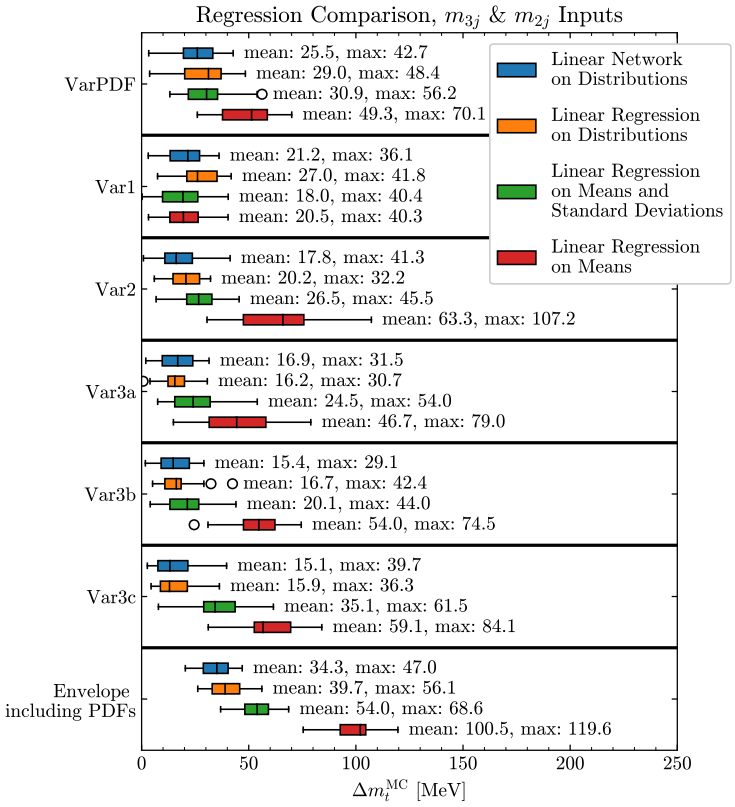


Figure 8.9: Uncertainties on the top mass fit within each group of variations in the A14 set of tunes for the OLS regressions, compared to the best linear network.

fitting. Results from the OLS approach are shown in Fig. 8.9. We find roughly similar, but slightly inferior, performance to the linear network for comparable inputs. Though the linear network does slightly better, it takes longer to train than OLS regression and is not deterministic.

8.5 DCTR with ParticleFlow

An alternative machine learning method developed to fit parameters is DCTR [314]. This method is based upon parameterized neural networks [1224] and exploits a relationship between the loss function and the likelihood ratio [118, 333–338, 375, 1225–1227].

The DCTR method works as follows. Suppose we have some parameters θ and some observables x. The probability distribution $p(x|\theta)$ of the observables depends on the values chosen for θ. An ambitious goal is to learn a function $f(x, \theta)$ which gives the full likelihood distribution of the observables x for any θ (as in JUNIPR [42, 258]). In practice, DCTR learns this distribution relative to the distribution over x for a fixed reference value θ_0. To do so, we give it observables x_{θ_0} drawn from the distribution at fixed $\theta = \theta_0$ (the reference sample) as well as observables x_{θ_S} drawn from the distribution using many values of $\theta \in \theta_S$

(the scanned sample). We do not tell the network the value of θ_0, however. Instead, we pretend that θ_0 is equal to θ and the network will learn that this is inconsistent. In practice, we train the network over pairs of events $\{x_i, x_i^0\} \in \{x_{\theta_S}, x_{\theta_0}\}$ chosen over a distribution of θ values and compute the binary cross-entropy loss

$$f = \underset{f'}{\operatorname{argmin}} \left(-\sum_{\theta} \sum_{\{x_i, x_i^0\} \in \{x_{\theta_S}, x_{\theta_0}\},} \left[\log\left(f'(x_i, \theta)\right) + \log\left(1 - f'(x_i^0, \theta)\right) \right] \right) \tag{8.4}$$

It is important that f' in the second term takes θ and not θ_0, otherwise the classification would be trivial.

Using such a loss function, the DCTR process for inferring model parameters from an ensemble of events involves two steps.

1. Train a parameterized classifier $f(x, \theta)$. In the application to top mass extraction, the reference sample has θ_0 corresponding to a fixed mass and tune. The scanned sample has θ which varies among values of m_t^{MC} and many values for the tune parameters.

2. Use the function $f(x, \theta)$ for regression. To do so, we re-minimize the loss for an unknown sample compared with an independent sample drawn using the same parameters as the reference sample. Now the network is fixed, but the parameters θ are varied to minimize the loss. The values which minimize the loss are the prediction.

To give a better sense of how DCTR works, we include a toy example with a one-dimensional Gaussian in Appendix F.1. For more details on DCTR, see [314] or [332].

8.5.1 Network architecture

In order to use DCTR to infer the top-quark mass, we need a parameterized neural network, $f(x, \theta)$ which is flexible enough to learn the likelihood ratio. The parameter(s) θ must include m_t^{MC}, but can also include the other tune parameters, depending on whether we try to regress those tune parameters or marginalize over them and only extract m_t^{MC}. We find the most effective network takes as input both low-level and high-level observables. The architecture of the network is sketched in Fig. 8.10.

For the high-level variables we take m_{3j} (the "top mass") and m_{2j} (the "W mass"), as in previous sections. We consider optionally applying soft-drop jet grooming to the jets before constructing the invariant masses. These are indicated by the blue portion of the figure.

For low-level observables, following DCTR [314] we use four-vectors of the constituent particles of the jets represented with a ParticleFlow Network [1206]. This is shown in the green region of the figure. For the inputs to ParticleFlow, we include up to 75 particles, with a maximum of 25 from each of the three jets. Each particle contains eight input variables: four variables are the four vector in (p_x, p_y, p_z, E), three variables are the momenta in a transformed coordinate system (p_T, η, φ), and a discrete tag for which jet the

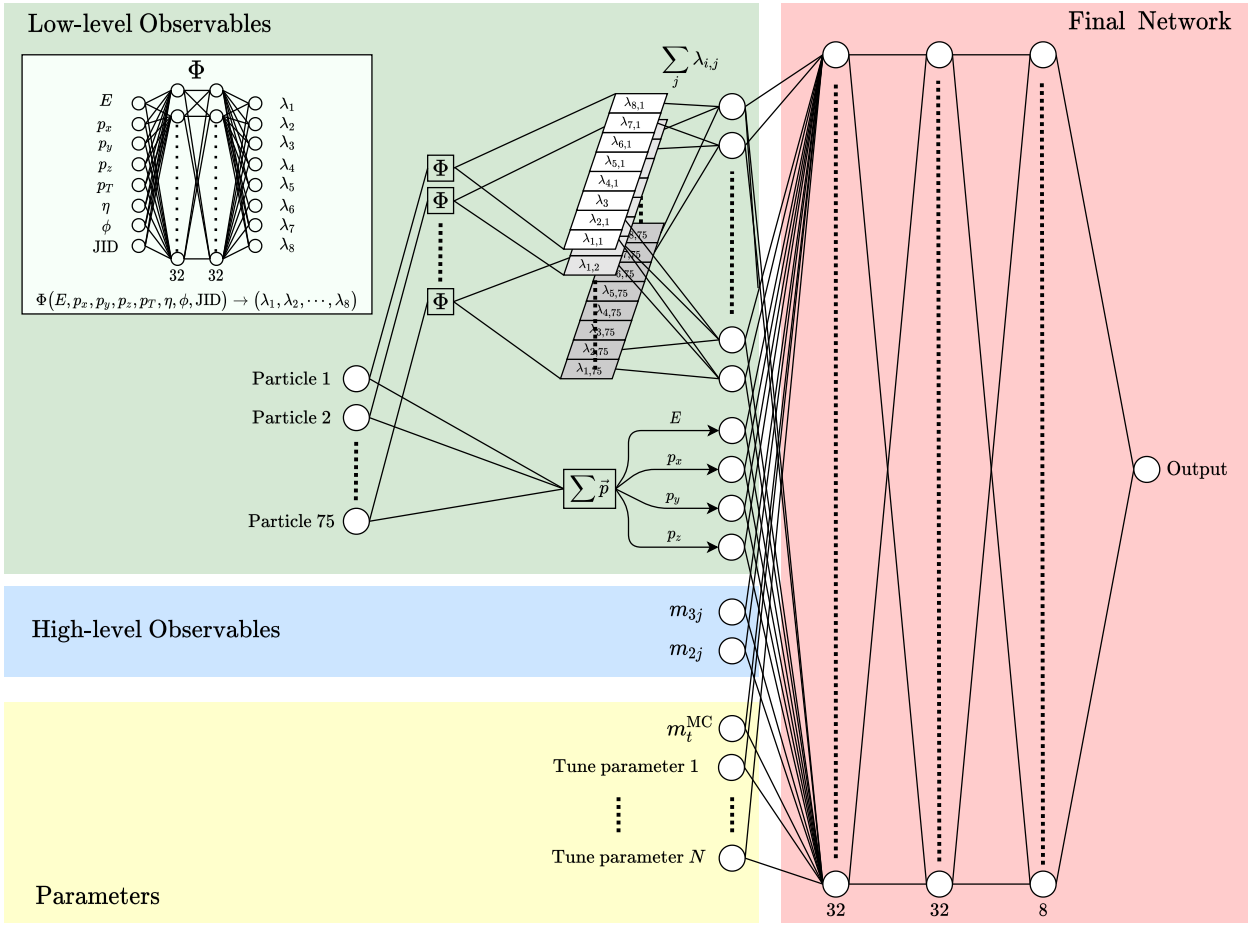


Figure 8.10: Architecture used to infer the top-quark mass with DCTR. The green portion shows the low-level information from the constituents of the jets. These are combined with a ParticleFlow Network, denoted by Φ acting on each constituent, with the resulting output being summed across the particles. The Φ network is shown in the breakout box. Next are the high-level inputs of the three-jet and two-jet invariant masses, shown in the blue portion. The last elements are the Monte Carlo parameters to be inferred, shown in yellow. All of these are combined in the latent space, which is then connected to the final output with a dense neural network shown in red.

constituent came from (0 for the b-tagged jet, 1 for the hardest un-tagged jet, and 2 for the softer un-tagged jet). A function Φ is applied to each of the particles in the event, mapping from an eight-dimensional input to a k-dimensional output. To ensure that the ordering of the particles is unimportant, each k-dimensional output is symmetrized (summed) over the particles. These are marked by the λ_k nodes in the figure. For Φ we use a neural network with two hidden layers. Each hidden layer contains 32 nodes with the ELU activation function and use a dropout rate of 10% during training. The final layer of Φ contains eight nodes, also using the ELU activation.[5] Many applications of ParticleFlow find that a larger latent space is needed, however, we found our results to be much more stable with 8, rather than 16 nodes. We also tried not including the ParticleFlow part of the network, but found better performance when it is included. As an additional input, we sum the four-vectors of each of the constituent particles and pass the sum top-quark 4-vector directly to the latent space.

The combined information from the Monte Carlo parameters, the high-level inputs, and the low-level inputs are concatenated together. From this space, another neural network is applied to generate the final event level classification, shown in the red region of the figure. We use three hidden layers with 32, 32, and 8 nodes, respectively. We again use the ELU activation function and apply a 10% dropout rate during training. The output is a single node activated with the sigmoid function. This results in networks that have between approximately 1400 and 3600 weights, depending on the number of Monte Carlo parameters included in the parameterization.

8.5.2 DCTR on a single tune

First we test DCTR's ability to regress m_t^{MC} for a fixed tune (A14 tune 21). For the fixed reference sample θ_0, we chose the top-quark mass to be 175 GeV.[6] In the scanned sample θ, we randomly choose m_t^{MC} for each event from a uniform distribution between 170-176 GeV. We use 1 million events each for the fixed and reference samples. The data set is split with 25% for validation and 75% for training.

As part of the study, we want to see if more information than just the three jet invariant mass can help the network extract the top mass better. To do so, we allow the network to use only m_{3j}; to use m_{3j} and m_{2j}; or to use m_{3j}, m_{2j} and the low-level inputs (as described above). In addition to these observables, the network is also given a value for $\theta = m_t^{MC}$. For the scanned sample, this is simply the value chosen in the random draw for the event. In the fixed reference sample, where $\theta_0 = 175$ GeV, the value of θ input is masked to a random value, chosen from the same range as in the scanned sample.

The networks are trained using the Adam optimizer to minimize the binary cross entropy loss function.

[5] The weights of Φ are trained along with the rest of the network, but could be pre-trained from a similar application.

[6] We found empirically that the DCTR procedure works better in practice if the Monte Carlo mass of the reference sample is larger than the values to be inferred.

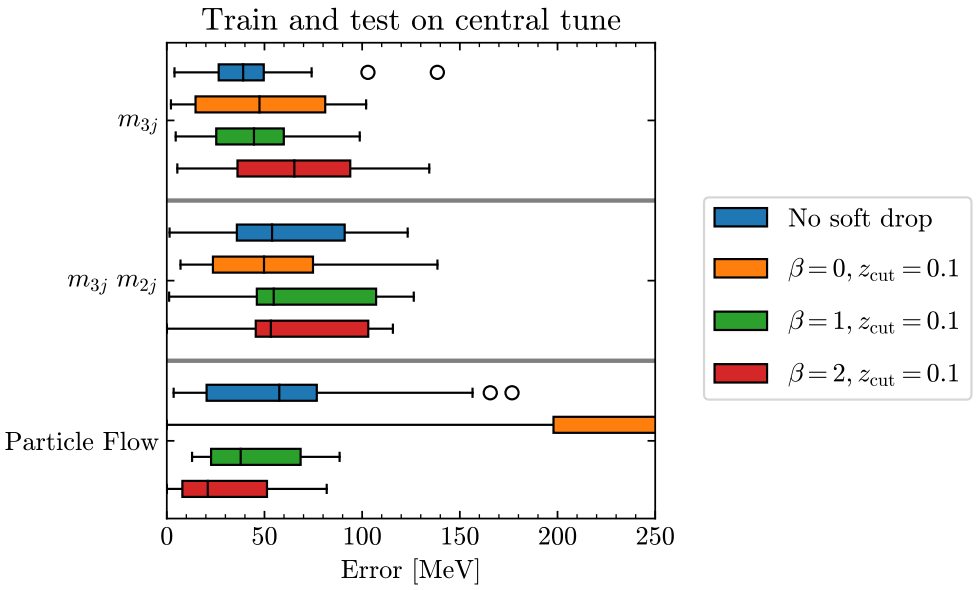


Figure 8.11: Errors on the mass of the top quark returned by the DCTR methods when trained and tested only on the central tune, for different soft drop parameters. Adding extra information (such as the mass of the W jet) does not seem to increase the accuracy. The uncertainty for the method is on the order of 50 MeV.

We set the initial learning rate to 10^{-3} and use the default β values for Adam. When the loss on the validation set has not improved for 10 epochs, the learning rate is decreased by a factor of $\sqrt{10}$, with a minimum rate of 10^{-6}. We also implement early stopping; if the validation loss has not improved for 25 epochs, training is halted. Training typically takes around 80 epochs.

After training the networks, the mass is extracted by computing the loss of the classifier between a test set and an independent reference set. We repeat this with the same five masses and five iterations of the test sets as in the regression methods presented earlier, with 4×10^5 events in each data set. The loss is minimized for the combined (test and reference) data as a function of m_t^{MC}.

In Fig. 8.11 we summarize the results for different soft-drop grooming parameter, and with and without the low-level inputs. Numbers shown are the absolute difference between the true and extracted Monte Carlo mass. The median error for nearly all of the methods here is $\lesssim 50$ MeV, while the maximum error is around 100-150 MeV. For this exercise, where the tune is fixed and only m_t^{MC} varies, there does not seem to be an advantage to using extra information (such as the mass of the jets from the W) or performing jet grooming. This conclusion will change when we include tune variations.

8.5.3 DCTR on Var 1 tunes

We saw that with no tune uncertainties, the DCTR method can regress m_t^{MC} with an uncertainty of order 50 MeV. When other parameters related to tunes are varied, such as in the showering or hadronization models,

DCTR offers multiple ways to proceed. We could train only on a single tune, trying to learn m_t^{MC}; we could train on multiple different tunes, again trying to learn only m_t^{MC}; or we could train over different tuning parameters and try to learn those as well as as m_t^{MC}.

To asses which of these options works the best, we train networks on data using the Var1 tunes. We again use 10^6 samples for the reference set and 10^6 sample for the scanned set with 75% of these samples for training and 25% for validation. For the reference set, we use samples drawn from the central tune (tune 21). The scanned set uses a uniform distribution for the mass (m_t^{MC}), the color re-connection range, and the strong coupling constant for multiple parton interactions (α_S^{MPI}). To remove edge effects, the sampling space is larger than the tune variations we eventually test against. Explicitly, the ranges are given by

$$\begin{aligned} m_t^{MC} &\in [170\text{ GeV}, 176\text{ GeV}], \\ \text{Color re-connection range} &\in [1.67, 1.75], \text{ and} \\ \alpha_S^{MPI} &\in [0.116, 0.136], \end{aligned} \tag{8.5}$$

and there is no correlation in the random samples. The training procedure is the same as above.

After training the network, we use DCTR to infer the mass (and possibly the color re-connection range and strong coupling) on three different tunes: 21, 23, and 24. These are the central, up, and down tunes of Var1. For reference the color re-connection range and the strong coupling constant for the tunes are (1.71, 0.126), (1.73, 0.131), and (1.69, 0.121) for 21, 23, and 24, respectively. For each test mass, we evaluate the spread in the inferred mass from the different tunes. This process is repeated for five separate test sets, each with 4×10^5 events for the reference and test set.

The results of the spreads are summarized in Fig. 8.12 with box-and-whisker plots. The results for the ungroomed jets are in the upper left panel, using soft drop with $\beta = 0$ in the upper right panel, using soft drop with $\beta = 1$ in the bottom left, and using soft drop with $\beta = 2$ are in the bottom right panel. In the top row of each panel, the only observable given to the classifier is the three jet invariant mass. The networks of the middle row have access to the two-jet invariant mass in addition, and the bottom row also includes a ParticleFlow network for the constituents of the three jets.

There are many noteworthy trends in these results. First, we examine how the different grooming methods affect the reconstruction. We saw before that the using soft drop for the histogram fitting methods greatly reduced the uncertainty. A similar pattern is observed here, especially when looking at the first two rows (not using ParticleFlow). For instance, all of the color bars for both the $\beta = 0$ and $\beta = 1$ panels have significantly lower mean and maximum Δm_t^{MC} than the corresponding colors for not using soft drop. The option of soft drop with $\beta = 2$ still does better than no soft drop, but not as good as the others.

The next noteworthy trend is that adding more information to the network helps to reduce the uncer-

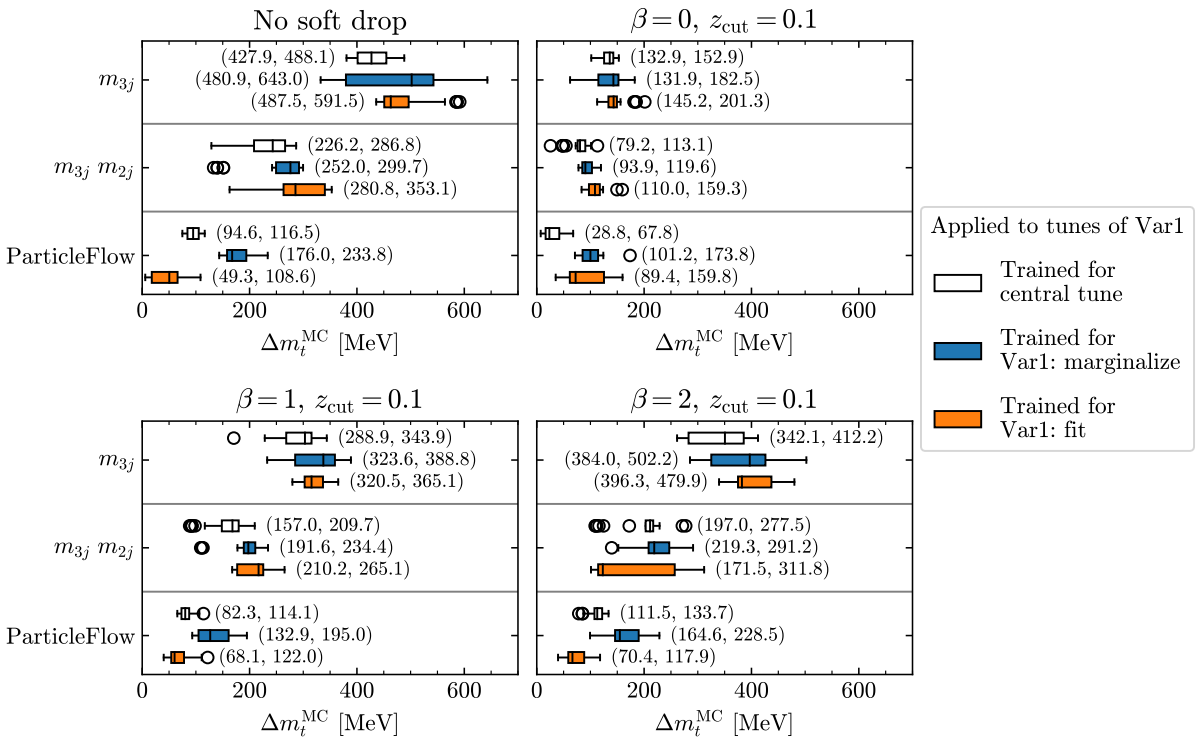


Figure 8.12: Results testing the input data for DCTR to use as well as the method of inference on VAR1. Including both m_{3j} and m_{2j} along with the low-level information captured by ParticleFlow results in the lowest uncertainty. DCTR works best when inferring all of the tune parameters (orange) as opposed to marginalizing over them (blue).

tainty. In each panel, the uncertainty is largest when only using m_{3j} and improves when adding in m_{2j}. The uncertainty is further reduced when including the ParticleFlow information in most panels. However, these networks are more challenging to train and often do not work for the full mass range. This is why the mean (and median) values drop, while sometimes still having large maximum uncertainties.

The last important observation is that the networks with ParticleFlow do better when they are also fitting to the tune parameters. The white data is for networks trained on the central tune alone, and thus only capable of inferring the mass. The blue data sees the scan across the tune parameters, but only tries to infer the mass, while the orange data also infers the tune parameters. For the m_{3j} alone or m_{3j} and m_{2j} rows, marginalizing over or fitting the tune parameters actually tends to make the uncertainties worse. With such a small amount of information (either one or two observables), the network does not learn how to correlate the changes in the tune parameters to changes in the observables. However, when the network also includes ParticleFlow, it can learn these correlations, and thus the uncertainty is reduced when fitting the tune parameters.

To summarize these results, DCTR works better with more input observables. Using the information contained in the four vectors of the constituents of the jets coming from the top quark decay allows DCTR

to correct for differences in the distributions caused by changing the tune parameters. Therefore, for the full set of tuning parameters in the next section, we restrict to the case of including ParticleFlow in the inputs and fitting each of the Monte Carlo parameters.

8.5.4 DCTR on full set of A14 tunes

We now apply the DCTR methodology to the A14 variations. The PDF variations are not included in the training, although we do evaluate on them. The reason for this is that PDF selection is a discrete choice, and DCTR is designed to work on continuous parameters. That is, there are not specific Monte Carlo parameters for DCTR to infer from the different PDFs. One can nevertheless assume that PDF variations are within the range of other tune variations and test how well DCTR works on samples generated with different PDF sets.

When considering multiple tunes, we must also decide which tune to use as the reference sample θ_0. Using the central tune would be the most obvious choice. However, since we found that the DCTR algorithm works better when we use higher reference masses, we allow for the possibility that it will work better using non-central tunes. We therefore test taking as the reference sample both the central tune as well as each of the "+" tunes for each of the variations.

For the scanned sample, the Monte Carlo parameters are randomly sampled for each event. The mass is drawn from a uniform distribution with a range of 170 GeV to 176 GeV. The tune parameters are sampled from $(\min \mathrm{z} - 0.5\Delta\mathrm{z}, \max \mathrm{z} + 0.5\Delta\mathrm{z})$, where z represents the value of an individual tune parameter, $\min \mathrm{z}$ is the minimum value across the variations, $\max \mathrm{z}$ is the maximum value across the variations, and $\Delta\mathrm{z}$ is the difference between the maximum and the minimum. The sampling space is larger than the values we will be testing at to remove possible edge effects.

The results are summarized in Fig. 8.13. The left panel shows the total envelope of $\Delta \mathrm{m}_{\mathrm{t}}^{\mathrm{MC}}$ caused by changing the Pythia tune parameters across the 11 variations. The right panel additionally includes the 3 remaining PDF variations. The different rows show different amounts of grooming, with no grooming on top and the three different soft drop options in the remaining rows. Each color denotes a different tune used as the reference set.

Overall, there is not an obvious best choice for the reference tune. In some choices of grooming, one tune will do better, but then it will not do as well on the different grooming choice. Similarly, some reference tunes that do well without the PDF variations do not generalize as well to including the variations from the PDF. However, we do note that using soft drop with $\beta = 0$ seems to consistently lead to worse results.

With an unclear best option, we chose to use no soft drop trained on the central tune to compare with the other regression models. This option generalizes well from training without the PDFs to including the PDF variations, only increasing the uncertainty by around 10 MeV.

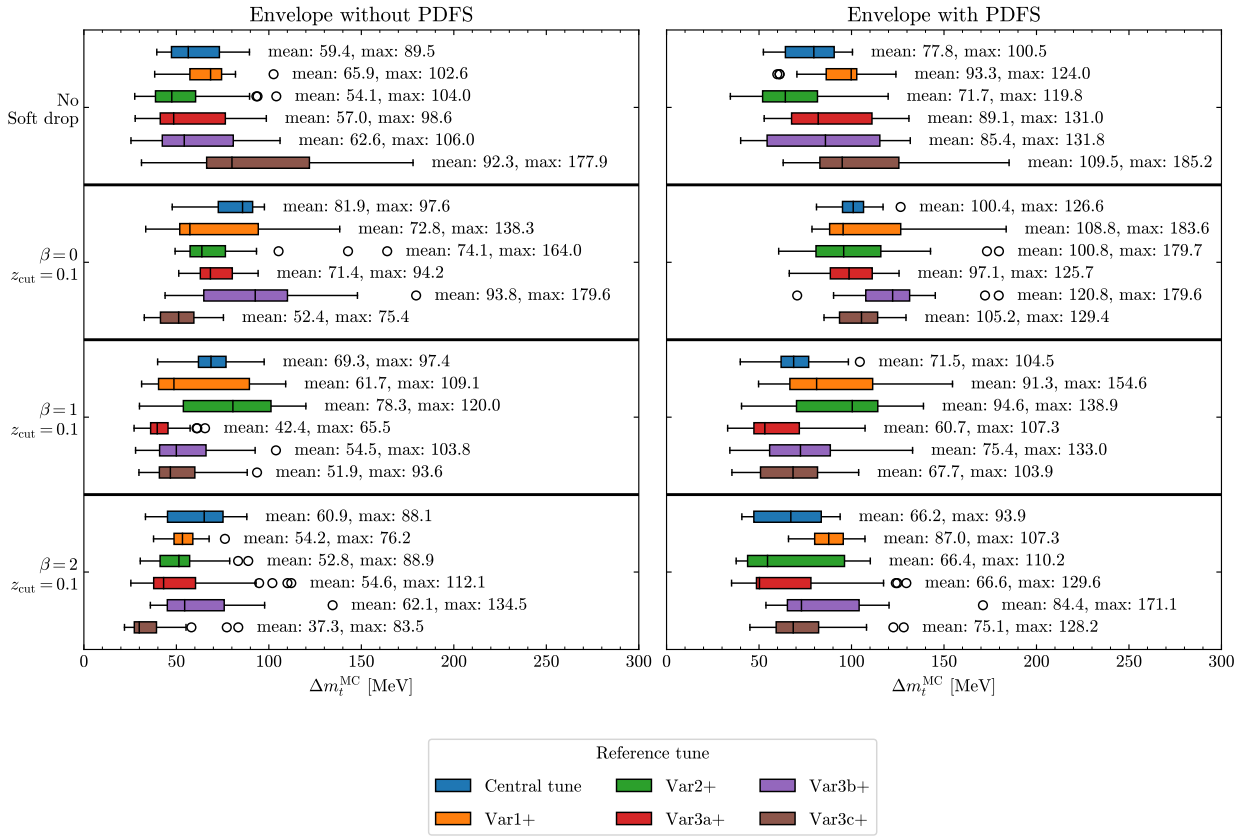


Figure 8.13: Total correlated uncertainty on the DCTR extraction of m_t^{MC} coming from only the Pythia tunes (left) and additionally including the PDF variation (right). The different rows show various grooming options and the colors denote different tunes used as the reference sample (θ_0) for the DCTR training. The no soft drop option trained using the central tune is chosen to compare with the other regression methods.

8.6 Discussion

In this chapter we have investigated an example of how machine-learning methods could help with measurement tasks at particle colliders. In particular, we explored a situation in which regression is assisted by learning simultaneously on an ensemble of events rather than on individual events. Despite the fact the individual events are totally uncorrelated, we find the best performance when variables constructed from the events are concatenated into an array, sorted, then input to the regression algorithm.

The case we explored is when the measurement is done by curve-fitting to simulated data to regress a single simulation parameter marginalizing over other parameters. In particular, we looked at the top-quark mass measurement. The traditional method is to extract the top-quark Monte Carlo mass m_t^{MC} by fitting to histograms, and then to estimate an uncertainty Δm_t^{MC} on this extracted value due to Monte-Carlo tuning uncertainty. This tuning uncertainty might be of order 500 MeV, which is comparable to statistical uncertainties, experimental systematic uncertainties, and theoretical uncertainties (such as converting the

top m_t^{MC} mass to a short-distance mass). In this chapter we focus only on reducing the tuning uncertainty.

We explored 4 classes of methodology to regress m_t^{MC}. First, we looked at histogram curve fitting. Uncertainties from this method are around 500 MeV but reduce to around 100 MeV if jet substructure techniques are used to clean the data before fitting (as shown in [1204]). Second, we used a 2D likelihood fit using the raw m_{2j} and m_{3j} observables incorporating a nuisance parameter to account for Monte Carlo tune differences. Third, we looked at linear regression techniques, both using a dense but shallow linear network and using ordinary least squares regression. Fourth, we used a machine-learning method called DCTR. DCTR is a two-step method: first the weights of a distribution are learned as a function of tuning parameters relative to a fiducial sample, and second the tuning parameters are optimized for a given test data sample.

The results of our study are summarized in detail in Fig. 8.14, with more details of each method in the appropriate section, and fewer details in the concise summary plot shown in Fig. 8.1. Fig. 8.14 shows the box-and-whisker plots for the different families of variations, while the final three rows show different methods of combining the variations. Probably the most realistic estimate of error is the "envelope including PDFs", which means we take the maximum and minimum values for the fit top mass across the A14 tunes. Such an approach assumes that the tunes are correlated and that actual data will lie somewhere within the complete range of variation. For completeness, we also include a more conservative estimate where each variation is assumed to probe completely different physics and is uncorrelated with other tune variations. In this case we add the errors in quadrature. The PDF variations are special because they are discrete: there is no way to interpolate between different A14 PDF sets as we could for other parameters such as α_s.[7] It is unclear how to train DCTR for PDF variations because of this complication. We thus include also numbers for the total envelope not including PDFs. Note that for most tune parameters there is no "right" answer: approximations such as the parton shower are made so the data can never be described perfectly. Thus in the context of a particular Monte Carlo simulation there is an irreducible uncertainty on how well the data could ever be described. In contrast, there is, in principle, a right answer for the PDFs, although in practice they are always used and fit in conjunction with calculations at a fixed perturbative order. In any case, our purpose is not advocating any particular choice of how to combine errors for an experimental analysis. We are simply providing various metrics by which the different methodologies can be compared.

The main take-home lesson from the summary in Fig. 8.14 is that in practice regression on sorted event ensembles does better than the classical histogram fitting approach or the profile likelihood fit. Using such methods reduce the Monte-Carlo-based uncertainty on current extractions on the top-quark mass from LHC data, perhaps even by a factor of 2. This is on top of the reduction gained from using jet grooming as

[7] There are other ways to study continuous variations of the PDF sets, but they are beyond the scope of this study.

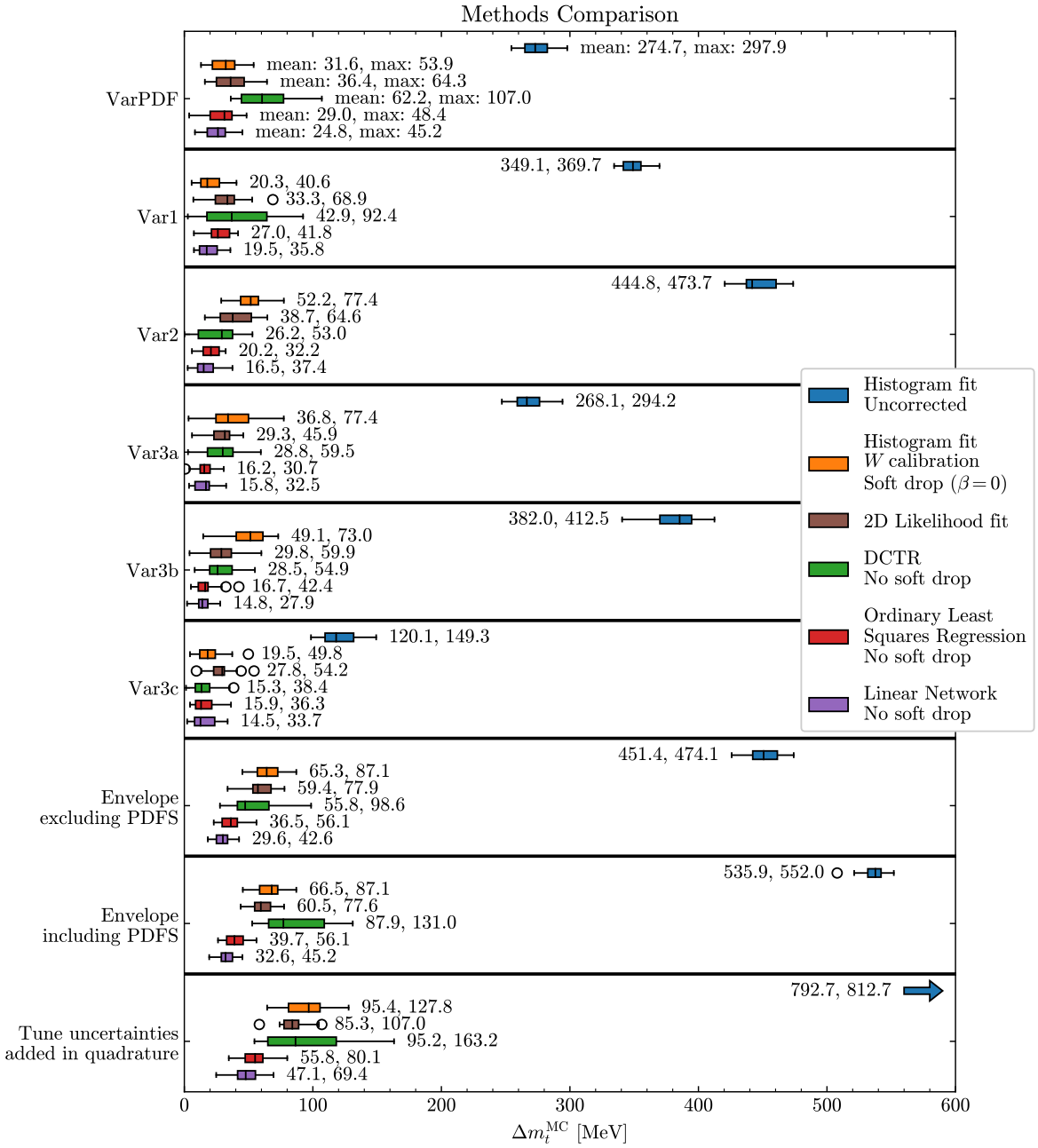


Figure 8.14: Summary comparison of the different methods. Upper six boxes show top quark tuning uncertainty among six different families of variations. The bottom boxes combine the separate uncertainties either by taking the envelope over the variations or by adding the uncertainties in quadrature. Here, DCTR uses ParticleFlow to simultaneously extract the mass and the tune parameters, and both the ordinary least squares regression and linear network are trained on ensembles of re-sampled events. The various methods are discussed more in the text.

advocated in [1204]. We found that DCTR works fine for the families of variations for which it was trained (not the PDFs), and has similar uncertainties to the histogram method. It has the potential benefit of being able to fit other tuning parameters well, but if one is only interested in a specific measurement, such as the top mass, then DCTR may be over-kill. Indeed, DCTR is somewhat challenging to implement and train, sensitive to how the inputs are refined, and requires some hyperparameter adjustment to get reasonable results at all. Its killer application may be more along the lines of [332] than direct parameter estimation.

Having established that a regression on sorted event ensembles is more effective than curve fitting a histogram, we also looked into what features of the ensemble the regression uses. In contrast to the histogram fitting approach, which focuses on the center of each observable's distribution, ensemble learning methods can weigh various parts of each observable's distribution differently. In §8.4, we showed how the ensemble methods can use the center of the m_{3j} distribution to learn the difference in masses, while using the upper tail of m_{3j} and the other observables to correct for the difference in tunes.

It is worth emphasizing that the point of this study is not a total numerical estimate for the uncertainty. Values throughout this chapter do not include any estimate of experimental systematic effects on the Monte Carlo tuning uncertainties, such as smearing due to jet energy resolution or detector effects. Thus one should not take the absolute size of the numbers as indicative that the tuning uncertainty could be reduced to the 30 MeV level. We do, however, conclude that linear regression, either through a linear network or an ordinary least squared regression on an ensemble of events, is a promising technique that has the potential to significantly reduce the dependence of the measured top-quark mass on Monte Carlo tuning parameters beyond the methodology already being employed. Although the uncertainty from marginalizing over unphysical parameters in simulation is smaller than other sources, it can be important for precision studies such as SUSY searches or vacuum stability, as discussed in the introduction.

In conclusion, we have shown that machine learning regression methods can work most effectively when trained on a sorted ensemble of uncorrelated events. We found these methods can improve upon a traditional histogram-fitting procedure for determining the top-quark Monte Carlo mass. In particular, performing linear regression using a shallow but dense network trained on sorted ensembles of events (30,000 at a time in our study) seems to combine excellent performance with simplicity.

9

Challenges for Unsupervised Anomaly Detection

9.1 Introduction

While many searches for physics beyond the Standard Model have been carried out at the Large Hadron Collider, new physics remains elusive. This may be due to a lack of new physics in the data, but it could also be due to us looking in the wrong place. Trying to design searches that are more robust to unexpected new physics has inspired a lot of work on anomaly detection using unsupervised methods including community wide challenges such as the LHC Olympics [123] and the Dark Machines Anomaly Score Challenge [150]. The goal of anomaly detection is to search for events which are "different" than what is expected. When used for anomaly detection, unsupervised methods attempt to characterize the space of background events in some way, independent of signal. The hope is then that signal events will stand out as being uncharacteristic.

Anomaly detection techniques can be broadly split into two categories. For some signals, the signal events look similar to typical background events and one must exploit information about the expected probability distribution of the background to find the signal. Many anomaly detection techniques have been developed to find signals of this type [109–128, 384, 1228, 1229]. Alternatively, some signals are qualitatively different from prototypical background and then methods that try to characterize an individual event as anomalous can be used [100, 108, 129–157]. Here, we restrict to the latter type of anomaly detection, where an anomaly score for individual events can be determined from the background ensemble and used for

discrimination, without needing to characterize the full probability distribution of the signal ensemble. With an effective method, events with a small score are likely to be a part of the background, while events with a larger score are not. There are many different ways of defining an anomaly score. Some rely on traditional high-level observables, like mass or N-subjettiness [1230], (see e.g. [109, 131, 384] which use traditional variables in anomaly detection). Others attempt to directly learn how likely a given event or object is using low-level information, like individual particle momenta (see e.g., [152]). Some methods that search for outliers rely on abstract representations to try to characterize the event space, such as the latent space of an autoencoder [125, 148]. Others give the event space itself a geometric interpretation in terms of distances [40, 1231, 1232]. Given the complexity and high-dimensionality of data at the LHC, many anomaly detection techniques employ machine learning.

In this chapter, we begin by exploring the use of autoencoders for anomaly detection on individual fat jets within events. Autoencoders were initially introduced for dimensionality reduction, similar to principal component analysis, to learn the important information in data while ignoring insignificant information and noise [1233]. Autoencoders contain an encoder, which reduces the dimensionality of the input to give some latent representation, and a decoder, which transforms the latent space back to the original space. In particle physics, autoencoders were first used for anomaly detection in [130, 132, 133], where they are meant to reconstruct certain types of data (background) but not others (signals). In order to work as an anomaly detector, an autoencoder should have a small reconstruction error for background events and a large reconstruction error for signal events. To do so, the autoencoder must establish a delicate balance in achieving a reconstruction fidelity which is accurate, but not too accurate. There are several cases where training a network with adequate discriminating power is especially difficult, such as when the signal looks very similar to the background, when the dataset has certain topological properties [124], or when innate characteristics of the samples make the signal sample simpler than the background sample to reconstruct [148, 154].

A generalization of autoencoders called variational autoencoders (VAEs) were introduced in [1234]. Unlike an ordinary autoencoder, where each input is mapped to an arbitrary point in the latent space, in a VAE, the latent space is a probability distribution which is sampled and mapped back to the original space by the decoder. In addition to the usual reconstruction error, the VAE loss also includes a Kullback-Leibler (KL) divergence component that pushes the latent space towards a Gaussian prior and regularizes network training. The latent space of the VAE encodes the probability distribution of the background training sample, which can be used in the anomaly score. VAEs were first used in anomaly detection in computer science in [1235], and first used for particle physics anomaly detection in [130, 131]. They have been widely studied since then [125, 142–148, 265, 1236].

The task of an autoencoder, variational or not, for unsupervised anomaly detection is to provide a

strong universal signal/background discriminant for a variety of signals having access only to background for training. In principle, this approach is advantageous because it opens the possibility to bypass Monte Carlo simulations and work directly with experimental data, which is almost completely background.[1] The autoencoder paradigm is based on the vision that there is trade-off between efficacy and generality: the ideal discriminant for a given signal and given background would be ineffective for a different signal and different background while a general discriminant, like the autoencoder, would work decently on a broad class of signals and backgrounds. The ideal assumes, first, that such a general discriminant exists with an appropriate use case, and second that it can be found by training purely on one or more background samples without any direct information about the signal. However, one has reason to be suspicious: machine learning methods work great at optimizing a given loss, which is meant to correlate strongly with the problem one is trying to solve. For autoencoder anomaly detection, the optimization (background only) is not aligned with the ultimate problem of interest (signal discovery over background), so it should not be surprising if the autoencoder does poorly. In §9.4, we explore the challenges induced by trying to optimize a VAE in a model agnostic way.

In order to understand what a VAE is learning, we study its latent space. In particular, we look at the distance between events in VAE latent space (see [148,1237] for other studies of VAE latent spaces in particle physics). Since we can think of the VAE anomaly score as a "distance" encoding how far any given event is from the background distribution, it is also natural to ask about the distances between individual events. We find there is a significant correlation between the Euclidean distance between events represented in the VAE latent space and the Wasserstein optimal transport distance between events represented as images. We study Wasserstein distances in particular because they were physically motivated in [40, 1231, 1232].

The correlation we observe between distances in the VAE latent space and between the event images motivates us to explore using optimal transport distances between events to define an anomaly score in §9.5. One method for using distances directly is to identify representative events in the background sample, and use an event-to-event distance between a given event and the representative event as the score. The advantages of this method we propose are that it does not require training a neural network and that it is easily adaptable to different background samples.

This chapter is organized as follows. In §9.2, we provide information about the dataset used in our study. In §9.3, we provide relevant background information on the metrics used (§9.3.1), and the details of the VAE architecture (§9.3.2). In §9.4, we explore the effectiveness of an image based convolutional autoencoder for anomaly detection, including its sensitivity to hyperparameters. We also explore correlations

[1] Of course, in the realistic situation where the training is performed on data, there could be signal events contaminating the background sample. Nevertheless, a number of studies have demonstrated that this has little effect on the autoencoder performance (e.g., [132, 133]). We will ignore this complication when we refer to "background only" samples in what follows.

between Euclidean distances in the autoencoder's latent space and optimal-transport distances among the event images in §9.4.2. This motivates the development of methods that directly use the optimal transport distances among events as an alternative to VAEs in §9.5. We conclude in §9.6.

9.2 Anomaly Detection Datasets

We begin by describing the datasets we use for our analysis. For concreteness, we focus on anomaly detection in simulated jet events at the LHC. We will consider QCD dijet events as the background, and consider both top and W jets as representatives of anomalous signal events. Although in practice anomaly detection techniques would not be used for top and W jets since there are dedicated experimental searches for these objects, these jets provide a simple benchmark for studying unsupervised methods. The authors of [146] have provided a suite of jets for Standard Model and beyond the Standard Model particle resonances which are available on Zenodo [1238]. A sample of QCD dijet background events are also provided on Zenodo using the same selection criteria, showering, and detector simulation parameters [1239]. The datasets were generated with MadGraph [973] and Pythia8 [1176] and used Delphes [1240] for fast detector simulation. Jets were clustered using FastJet [1209, 1241] using the anti-k_T algorithm [1208] with a cone size of $R = 1.0$. The event selection requires two hard jets, with leading jet having $p_T > 450$ GeV and the sub-leading jet having $p_T > 200$ GeV. The QCD jets are created using the $pp \to jj$ process in MadGraph, while the top and W jets we examine are produced through a Z' which decays to $t\bar{t}$ or a W' which decays to a W and invisibly decaying Z. Samples are available in [1238] for a variety of top and W masses, but we use only those with the SM values. There are around 600,000 QCD dijet events and 100,000 events for the "anomalous" top and W events. We reserve 100,000 QCD events for testing and use 50,000 QCD events for validation when training the VAE.

The leading jet in each event is used for the analysis. We pre-process the raw four-vectors into an image following the procedure presented in [72]. Using the EnergyFlow package [1242], we boost and rotate the jet along the beam direction so that the p_T weighted centroid is located at $(\eta, \varphi) = (0, 0)$. Next, the jet is rotated about the centroid such that the p_T weighted major principal axis is vertical. After this, the jet is flipped along both the horizontal and vertical axes so that the maximum intensity is in the upper right quadrant. Only after the centering, rotations, and flipping do we pixelate the data [72]. We use 40×40 pixel images covering a range of $\Delta\eta = \Delta\varphi = 3.2$. The final step of the pre-processing is to divide by the total p_T in the image. Note that we do *not* standardize each pixel by, e.g., subtracting the mean and dividing by the standard deviation for the entire training dataset, because optimal transport requires positive values in every pixel. It is important to note that the individual images are very sparse and do not resemble the average of the dataset. For instance, out of the 1600 pixels, only 10.4 ± 5.3, 13.5 ± 4.3, and 10.1 ± 3.3 pixels account for more than 1% of the total p_T of the image for the QCD, top, and W jets, respectively.

9.3 Defining the Anomaly Score

Anomaly detection, in general, requires an anomaly score: we want to determine if an event is anomalous by measuring how far away it is from a typical background event. This anomaly score can also be thought of as the "distance" between an event and an ensemble. In order to define an event-to-ensemble distance it is helpful first to explore event-to-event distance measures. For instance, given an event-to-event metric, one could compute the distance from an event to some fiducial background event, and use this as a proxy for the event-to-ensemble distance. To understand both types of distances, we need to review the metrics used to define the distance, which we will do in §9.3.1. We can also use an autoencoder to generate an implicit construction of an approximate event-to-ensemble distance, in the form of an anomaly score. We will provide background and discuss the architecture of our autoencoder in §9.3.2.

9.3.1 Metrics

First, we define the metrics that can be used to compute event-to-event distances. One of the simplest event representations is to treat an event as an image, with pixel intensities representing the particles' transverse momentum [1243].[2] A simple event-to-event metric, the "mean power error" (MPE), can then be written as:

$$\mathrm{d}^{(\alpha)}_{\mathrm{MPE}}(\mathcal{I}_1, \mathcal{I}_2) = \frac{1}{\mathrm{N}_{\mathrm{pixels}}} \sum_{\mathrm{i} \in \mathrm{pixels}} |\mathcal{I}_{1,\mathrm{i}} - \mathcal{I}_{2,\mathrm{i}}|^{\alpha} \ . \tag{9.1}$$

where $\mathcal{I}_{1(2),\mathrm{i}}$ is the pixel intensity (transverse momentum) in pixel i of the image 1(2), and α is a parameter that governs the relative importance of pixels with high/low intensity differences. This type of metric is often used for doing regression. Frequently, the choice $\alpha = 2$ is made, inspired by the χ^2 statistic, in which case $\mathrm{d}^{(2)}_{\mathrm{MPE}}$ is known as the mean-square error (MSE). The mean-absolute error (MAE) is another well-known choice, corresponding to $\alpha = 1$.

While $\mathrm{d}^{(\alpha)}_{\mathrm{MPE}}$ makes sense in regression, using it on images does not make much sense from a physics point of view.[3] For instance, let $\mathcal{I}_1$ be the image of a particle with energy E in a single pixel and $\mathcal{I}_2$ be the image of a particle with same energy E in the neighboring pixel. These events are nearly identical physically, but will have a very large MSE distance. Moreover, we will still get the same MSE distance if we move one of the two pixels much further way. Physically similar events do not necessarily result in small MSE distances.

[2] In principle, it would be interesting to consider the complete set of four-vectors of the particles in an event as a representation, rather than the pixelated image, and define a metric on these. The p-Wasserstein distances described later in this section are well-suited for such a representation, but building an autoencoder architecture on the full set of four-vectors is more challenging. It is also important to comment that our image representation is dependent on its preprocessing. Although recent studies have shown that the processing done to events before anomaly detection is inherently model dependent [154], we work with the images as described.

[3] In contrast, if one designs a neural network with higher-level variables as the input data representation, using MPE as the metric is a sensible choice.

A completely different way to assign distance between two events is to compute the minimum "effort" needed to transform one image into the other, known as the optimal transport distance. There are many possible optimal transport algorithms (see [1244] for a broad review). Finding the minimum effort is an optimization problem: given a cost function c_{ij}, where i and j label elements (e.g. pixel labels) of the two events, we optimize over the transport plan, f_{ij}. The cost can be thought of as how much work it takes to transport a single unit of intensity a given distance, and the plan describes how much intensity to transport and where to transport it to. In terms of the cost and plan, the total optimal transport cost d_{OT} is then defined as

$$d_{OT} = \min_{f} \sum_{i,j} f_{ij}\, c_{ij}\,. \tag{9.2}$$

In some cases, the cost function c_{ij} is itself a positive definite distance, in which case d_{OT} is also a distance. One example is the set of p-Wasserstein distances:

$$d^{(p)}_{Wass} = \left(\min_{f} \sum_{i,j} f_{ij}\ (c_{ij})^p\right)^{1/p}, \tag{9.3}$$

Depending on the problem, the set of f_{ij} may have to satisfy additional constraints.

We define the underlying cost c_{ij} as the Euclidean distance in the (η, φ) plane between pixel i in image $\mathcal{I}_1$ and pixel j in image $\mathcal{I}_2$. The transport plan f_{ij} is defined by the amount of p_T that is moved from pixel i in image $\mathcal{I}_1$ to pixel j in image $\mathcal{I}_2$. The transport plan is constrained such that the amount of p_T moved from a pixel cannot be more than what was there, $\sum_j f_{ij} \leq p_{T,i}$. Similarly the amount of p_T moved into a pixel cannot exceed the amount in that pixel in $\mathcal{I}_2$: $\sum_i f_{ij} \leq p'_{T,j}$. Here, we consider normalized images, preprocessed such that the total intensity summed over all pixels is equal to unity, so that there is no extra cost of creating or destroying p_T. In mathematical language, we are considering "balanced optimal transport".

In particle physics applications, unbalanced optimal transport with the choice $p = 1$ is commonly referred to as the Energy Movers Distance (EMD) [1231, 1232], as it has the interpretation of work required to rearrange an energy pattern. This interpretation makes the EMD a natural choice for a metric on collider events. This has prompted further work on using the EMD to define event shape observables characterizing the event isotropy [1245], which can be useful in searching for signals that are very non-QCD-like [1246,1247]. Sometimes, $p > 1$ has been considered [1232], while the case of $0 < p < 1$ has been less explored. Intuitively, $p < 1$ gives more importance to smaller distances. While the EMD includes an additional term to account for energy differences between jets, in our results, we will restrict to balanced optimal transport, since we normalize the images.

The p-Wasserstein optimal transport metrics are more aligned with what one expects for physical events than MPE. For example, two single-particle events where the particles are nearby will have a much smaller

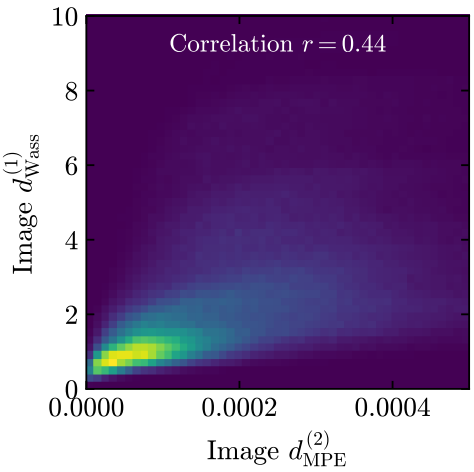


Figure 9.1: The pairwise (event-to-event) distances between event images for different metrics. The mean squared error ($\mathrm{d}^{(2)}_{\mathrm{MPE}}$) is displayed along the x-axis and the 1-Wasserstein distance ($\mathrm{d}^{(1)}_{\mathrm{Wass}}$) is along the y-axis.

p-Wasserstein distance than when they are far from each other, in contrast to their MSE distance. We find that the 1-Wasserstein distance and MSE have mild correlations, as shown in Fig. 9.1.

We reiterate that both $\mathrm{d}^{(\alpha)}_{\mathrm{MPE}}$ and $\mathrm{d}^{(\mathrm{p})}_{\mathrm{Wass}}$ are used to compare the distance between two images (or events). However, for anomaly detection, we want to know how far an event is from the expected distribution. One way to do this is with an autoencoder, which we describe next.

9.3.2 Autoencoders

A popular method for detecting anomalous data is with a neural-network autoencoder (AE). An autoencoder works by first encoding the data in a lower-dimensional *latent space*, and then decoding it back to the original higher-dimensional representation. The idea is that data similar to the training sample will be reconstructed well, whereas data that is not similar to the training sample may be reconstructed poorly. The reconstruction fidelity can then be used as an anomaly score. Often the data are represented as images, and the autoencoder uses the MSE metric (eq. (9.1) with $\alpha = 2$) to compare the input image to the reconstructed image.

In Fig. 9.2, we show an example of an autoencoder architecture that we will use, which we implement in pytorch [1248]. The encoder is made up of some number of *downsampling blocks* (there are two in the figure, each marked by a dashed blue line). Each block contains two sets of 3×3 convolutional layers with a depth of five filters. The stride and padding are set to keep the image size the same and the ELU activation function [1249] is applied after each layer. After the convolutional layers, the data is downsampled through a 2×2 average pooling layer. After the final downsampling block, the data is flattened and then followed by a dense layer with 100 nodes and an ELU activation. Finally the network is mapped to the latent space through another dense layer. We experiment with one, two, and three downsampling blocks, and use a fixed

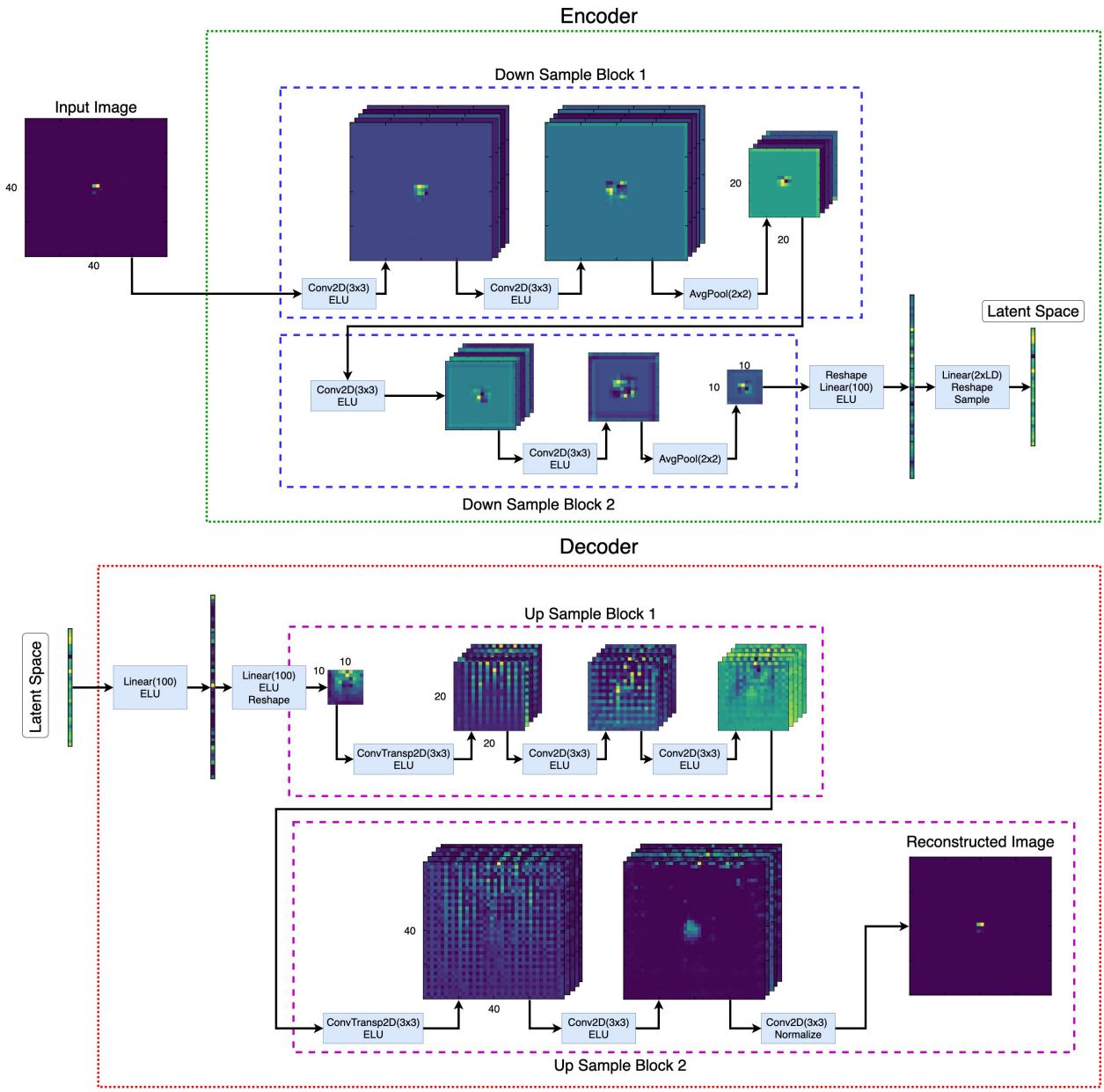


Figure 9.2: Example architecture of an autoencoder, as used for this study. The autoencoder is made of two networks, the encoder and the decoder, each with one, two, or three down(up)-sampling blocks.

latent size of 64 dimensions. Our latent space is substantially larger than what is often used, for example [132] uses a six dimensional latent representation and [133] finds the optimal size to be around 20-34 for their top-tagging data. We chose the latent space size by preforming a small scan over latent dimension sizes [2, 4, 8, 16, 32, 64] on the "The Machine Learning Landscape of Top Taggers" data [39] and found that the larger latent space yielded better top tagging. We then changed to the current data set, as there are more signals to consider (i.e. the W). The new data uses different transverse momentum cuts, so re-optimizing would in general be required. However, part of the point of this chapter is to point out that one cannot

optimize without a signal model in mind.[4]

The second part of the AE is the decoder, which maps the latent space back to the space of the input data. In our setup, the decoder is a mirror of the encoder. The first step is a dense layer with 100 nodes and ELU activation. From here, another dense layer is used, where the number of nodes is set to the number of pixels in the final downsampling block. The ELU activation function is used again, and then the data is reshaped into a square array. From here, the same number of *upsampling blocks* is applied as the number of downsampling blocks. In each upsampling block, the first operation is a 2D transposed convolution which doubles the shape of the image and contains a depth of five filters, followed by the ELU activation. After this, two 3×3 convolutional layers are used with the ELU activation with the stride and padding set to keep the image size the same. The final convolution operation reduces the depth to one channel.

During training and inference, the input image is compared with the reconstructed image via some choice of event-to-event metric. A common method is to use the MSE as the loss function, with the aim of reproducing the exact image. However, it is possible to use other metrics for the comparison. Furthermore, the metric used for training does not need to be the same as the metric used for the anomaly score (see for instance [146]). We will refer to the difference between the input and reconstructed image as the *image distance*, also known as the reconstruction error.

A variational autoencoder enhances the basic autoencoder by adding stochasticity to the latent embedding. In a regular autoencoder, which is a deterministic function, very dissimilar events can be placed near each other in the latent space. Distances in the latent space of an ordinary AE therefore do not have a precise meaning. In a VAE, the stochastic element makes the network return a distribution in the latent space for each input event. Since the same input data can be mapped to several nearby points in a VAE, dissimilar events cannot be placed nearby. Returning a distribution in the latent space is therefore essential for making distances in the latent space meaningful. The stochasticity also connects the loss to the statistics method of variational inference [1234, 1250], as we summarize in appendix F.3 (see also [1250, 1251] for reviews). Specifically, we show that the autoencoder estimates a *lower bound* on the likelihood for any given event given the assumption that the event comes from the background distribution that the network is trained on.

To implement the stochasticity of a VAE, our networks are trained using the standard reparameterization trick [1234, 1252]. A single element of the input data now yields a distribution, and these distributions are treated as a set of D independent Gaussian distributions, where D is the dimension of the latent space. The output of the encoder is then doubled: instead of returning a single point in the latent space, it now outputs both the means μ and the variances σ^2 of the distribution in latent space. The loss function for the

[4]On the other hand, as discusssed below, we can view the KL-divergence (and β) as regularizing the network. There will be a strong correlation with the optimal value of β with the size of the latent space. So making a different selection for the latent dimension would lead to a different value of β, but would not change the story.

network also has to be modified: we want the background sample to be well modeled by a set of Gaussian distributions in latent space. This is done by introducing a Kullback-Leibler divergence (KLD) term (see appendix F.3 for details), which is estimated as:[5]

$$\text{KLD} = -\frac{1}{2}\left(1 + \log\sigma^2 - \mu^2 - \sigma^2\right). \tag{9.4}$$

This KLD term acts to regularize the autoencoder by pushing the means in the latent space to zero and the variances to one. Depending on the metric used to determine the distance between the original and reconstructed data, more or less regularization may be needed. To account for this, we introduce another hyperparameter β, and define the loss function as

$$\text{L} = \left(1 - \beta\right) \times \text{Image distance} + \beta \times \text{KLD} \ . \tag{9.5}$$

We scan over $\beta \in \left\{0, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\right\}$, typically finding the best results for small but nonzero β.

To minimize the loss given in eq. (9.5), we use the Adam optimizer [1255] with the default parameters and an initial learning rate of 10^{-3}. The training data consists of around 550,000 QCD dijet events, and we reserve 50,000 QCD events for an independent validation set. After each epoch of training, the loss is evaluated on the validation set. When the loss has not improved on the validation set for five epochs, the learning rate is decreased by a factor of 10, with a minimum learning rate of 10^{-5}. Training concludes when the validation loss has not improved for 12 epochs. We then restore the weights of the network from the epoch with the best validation loss.

9.4 Autoencoder Results

Here we present the results of our studies of variational autoencoders. We start by studying the metric dependence of VAE performance as anomaly detectors. Then we study the latent space to understand what the VAE is learning.

9.4.1 Autoencoder performance

Now we study the performance of variational autoencoders as anomaly detectors using different metrics. Anomaly detection with an autoencoder requires two metric choices. First, one must choose a **training metric**, used for computing the image distance during training. Next, one must choose an **anomaly metric**

[5] This estimation assumes Standard Normal priors for the likelihood of the latent data, as described in appendix F.3. There is a great deal of ongoing research into methods to improve the likelihood estimate by changing the latent space priors or improving the posterior approximations of the encoder [148, 150, 1253, 1254].

to compute an anomaly score which determines how similar an event is to the training sample. The training metric and anomaly metric can be the same, but do not have to be.

For the training metric, we consider MSE-type metrics $d^{(2)}_{MPE}$ and $d^{(1)}_{MPE}$ and p-Wasserstein metrics $d^{(1)}_{Wass}$ and $d^{(2)}_{Wass}$. Using a p-Wasserstein metric in the loss function to train an autoencoder is not standard, and requires a little bit of extra engineering.[6] The challenge is that the optimal-transport metrics are not well-suited for the back-propagation part of the training procedure of a neural network. To get around this, we used the Sinkhorn approximation within the GeomLoss package [1256]. Even with this, training was slow and sometimes timed out after three days of training on GPU. In contrast, the MSE and MAE networks typically completed training in around 12 hours on the same platform.

For the anomaly metric, we consider either using the full loss (including both the training metric contribution and the KL-divergence part in the variational autoencoder), just the MSE error between the input and output images $\left(d^{(2)}_{MPE}\right)$, the MAE $\left(d^{(1)}_{MPE}\right)$, or the p-Wasserstein distances $\left(d^{(p)}_{Wass}\right)$ with $p = 0.5$, 1.0, and 2.0. The value of each of these is computed for the test samples for the QCD dijet events, the top-jet events, and the W-jet events.

To evaluate performance in anomaly detection, we train the autoencoder on a QCD background using the training metric. Then we evaluate the anomaly score using the anomaly metric for a boosted top jet signal sample and a boosted W-jet signal sample. For a figure of merit of performance we use the Area Under the receiver operating characteristic Curve (AUC). We also include the signal efficiency at a cut which allows only 10% of the QCD events to pass, which is denoted $\varepsilon_S(\varepsilon_B = 0.1)$.

Results are shown in Table 9.1 for the training metric choices $d^{(2)}_{MPE}$ and $d^{(1)}_{Wass}$ and for different numbers of downsampling blocks in the network. For each number of down samplings, we trained the network with different values of the VAE parameter β, and in the table present the results for the value of β which achieved the smallest loss on the validation data. For the $d^{(2)}_{MPE}$ trained networks, the values of β which minimized the loss were 10^{-7}, 10^{-7}, and 10^{-8}, for the one, two, and three down sample block networks, respectively. The $d^{(1)}_{Wass}$ trained results are in the lower part of the table and had optimal values of β of 10^{-5}, 10^{-8}, and 10^{-7} for one, two, and three down sampling blocks, respectively. The entries highlighted in blue indicate the configuration with the best AUC and $\varepsilon_S(\varepsilon_B = 0.1)$ for top jets and W jets across all of our considered VAE architectures, training methods, and anomaly score methods. The top row in the table shows the results (in red) from a supervised approach, for comparison (see appendix F.4 for details of the supervised algorithm).

In general, we find the networks trained with $d^{(2)}_{MPE}$ as the training metric and using the full loss as the anomaly metric has the best AUC. The exception is when only a single down sample layer is used, in which case using $d^{(1)}_{Wass}$ as the anomaly metric does slightly better for the top-jet signal than using the full loss as

[6] [1237] also implements a VAE trained with a p-Wasserstein metric.

Signal Training Metric	Down Samplings	Anomaly Metric	Top jet AUC	Top jet $\varepsilon_S(\varepsilon_B = 0.1)$	W jet AUC	W jet $\varepsilon_S(\varepsilon_B = 0.1)$
Supervised	-	-	**0.94**	**0.81**	**0.96**	**0.91**
MSE	1 ($\beta = 10^{-7}$)	Loss	0.82	0.45	0.61	0.10
		MSE	0.82	0.45	0.60	0.10
		MAE	0.79	0.34	0.48	0.03
		Wass(0.5)	0.82	0.42	0.45	0.04
		Wass(1)	0.83	0.47	0.41	0.05
		Wass(2)	0.81	0.45	0.39	0.08
	2 ($\beta = 10^{-7}$)	Loss	0.83	0.48	**0.65**	0.14
		MSE	0.83	0.48	0.65	0.14
		MAE	0.80	0.37	0.53	0.04
		Wass(0.5)	0.82	0.43	0.51	0.04
		Wass(1)	0.82	0.44	0.51	0.04
		Wass(2)	0.81	0.44	0.54	0.06
	3 ($\beta = 10^{-8}$)	Loss	**0.84**	0.49	0.65	0.12
		MSE	0.84	0.48	0.65	0.12
		MAE	0.81	0.39	0.53	0.04
		Wass(0.5)	0.83	0.46	0.52	0.04
		Wass(1)	0.84	0.51	0.52	0.05
		Wass(2)	0.82	**0.51**	0.54	0.08
Wass(1)	1 ($\beta = 10^{-5}$)	Loss	0.78	0.35	0.44	0.04
		MSE	0.71	0.23	0.57	0.12
		MAE	0.72	0.20	0.49	0.03
		Wass(0.5)	0.75	0.26	0.47	0.03
		Wass(1)	0.78	0.35	0.44	0.04
		Wass(2)	0.76	0.37	0.39	0.05
	2 ($\beta = 10^{-8}$)	Loss	0.79	0.37	0.46	0.04
		MSE	0.76	0.33	0.61	**0.15**
		MAE	0.75	0.26	0.52	0.04
		Wass(0.5)	0.77	0.31	0.49	0.03
		Wass(1)	0.79	0.37	0.46	0.04
		Wass(2)	0.77	0.38	0.40	0.06
	3 ($\beta = 10^{-7}$)	Loss	0.79	0.36	0.41	0.03
		MSE	0.79	0.38	0.60	0.13
		MAE	0.77	0.31	0.51	0.03
		Wass(0.5)	0.79	0.33	0.47	0.03
		Wass(1)	0.79	0.36	0.41	0.03
		Wass(2)	0.7	0.32	0.36	0.06

Table 9.1: Results showing the ability of a VAE trained on QCD only samples to distinguish top and W jets from QCD jets. The Training Metric column shows which distance metric is used in the loss function for training, and the Anomaly Metric column shows the distance metric used at inference time. The bold blue entries mark the highest AUCs and signal efficiencies overall. We indicate the p-Wasserstein distance metric as Wass(p), and the MPE with power $\alpha = 1, 2$ by MAE and MSE, respectively. The number in parenthesis in the Down Sampling column denotes the value of β which yields the lowest total loss on the validation set for the given number of down samplings.

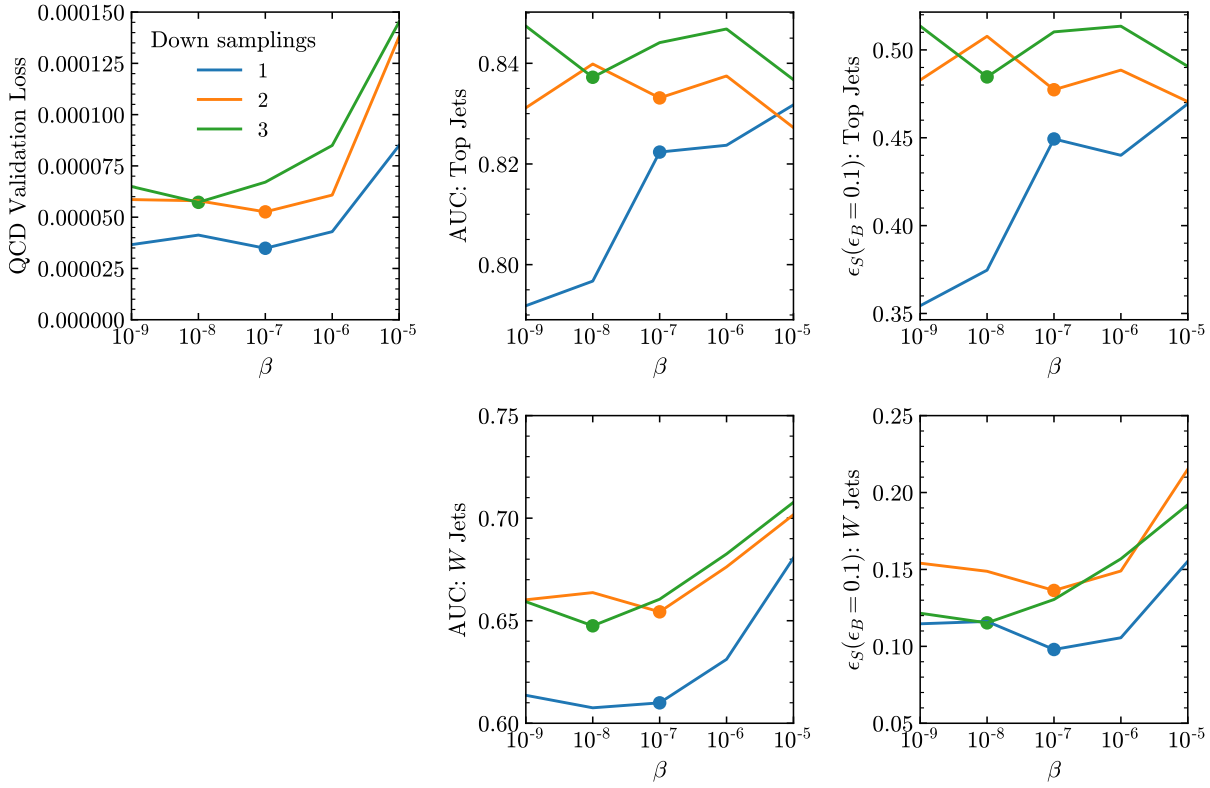


Figure 9.3: Results from scanning over β. The value of β which minimizes the validation loss does not yield the highest AUCs or signal acceptence for fixed background rejection for either the top or W samples. If one were to use one of the signal samples to chose the value of β, it can lead to worse results on the other signal.

the anomaly metric. When $d^{(1)}_{\text{Wass}}$ is used as the training metric, the best performance is with $d^{(2)}_{\text{MPE}}$ as the anomaly metric.

We can see at this stage the proliferation of choices one has to make when deciding what architecture, training metric, and anomaly metric to use. Making these choices is especially hard to do if one wants to remain model agnostic. For instance, Fig. 9.3 shows the results of the network trained with $d^{(2)}_{\text{MPE}}$ as the reconstruction loss. The left panel contains the loss on the QCD validation events. Using the idea that minimizing the loss is getting a better estimate of the probability of an event, one would expect that the network configuration (number of down samplings and value of β) which minimizes the loss will have learned the QCD distribution the best. However, the remaining panels show the ability of the networks to distinguish top and W jets from the QCD background in the upper and lower panels, respectively. The middle panels display the AUC and the right panels show the signal acceptance at a cut that allows only 10% of the QCD background events to pass. In particular, we see that the value of β which minimizes any of the loss curves does not yield the best signal separation. We also point out that the network with a single down sample block has the lowest loss, but is consistently the worst anomaly detector. This figure also highlights the danger

of using the metrics of a particular signal to chose the hyperparameters of a universal anomaly detector. Examining only performance on the top jets, it would be tempting to pick the three down sample networks with a value of $\beta = 10^{-9}$, as this gives the best AUC and signal acceptance for the fixed background rejection for the tops. However, these particular networks have poor score for the W jets. This is the challenge of signal independent searches; without a signal model in mind, optimizing analysis strategies is hard to do in a principled manner.

The network trained with $\mathrm{d}^{(2)}_{\mathrm{MPE}}$ with a small KL divergence term yields the best anomaly detection performance. Therefore, we expect that it is learning a good representation of the underlying background distribution. We next explore this hypothesis by examining event-to-event distances among different metrics.

9.4.2 What has the VAE learned?

In order for a variational autoencoder to be able to judge the likelihood of an event given the assumption that it came from the set of training data, it must have a representation of the probability distribution of events in the training sample. Moreover, since it first maps events to a lower-dimensional latent space, the information about the relative likelihood should be encoded in the latent space in some way. It would make sense if the network places similar events *nearby* in the latent space, and dissimilar events far apart. In this section we attempt to quantify if this is indeed true by comparing to the more physical Wasserstein distance.

Since each input is mapped to a (Gaussian) distribution in the latent space, we use the Euclidean distance between the means of these distributions, which is a simple measure of the distance in the latent space.[7] In Fig. 9.4, we show the correlations between the Euclidean latent space distance and the 1-Wasserstein event distance among all the $\sim 10^6$ pairings of 1000 events in the QCD test set for various values of the VAE parameter β. For this study, the events are passed through the encoder part of a VAE with three down sampling layers, down to a 64 dimensional latent space where the Euclidean distance is computed. As the value of β is increased, the network goes from having little regularization to being forced to approach a Gaussian. Correspondingly, the correlation initially grows as the structure is forced upon the latent representation, and then decreases as β becomes so large that the regularization dominates and the distribution becomes nearly Gaussian. We observe similar results for the networks with one and two downsampling layers that are trained with $\mathrm{d}^{(2)}_{\mathrm{MPE}}$ in the loss function. In this figure, the value of β which gives the minimum loss corresponds to the β with maximum correlation, but we do not find this trend to hold in general. It seems that the VAE with an intermediate value of β that balances the $\mathrm{d}^{(2)}_{\mathrm{MPE}}$ and KLD terms in the loss function creates a latent space where distances between events are correlated with the $\mathrm{d}^{(1)}_{\mathrm{Wass}}$ distance in the image space.

[7] One could also try to take into account the variance of the distributions, by e.g., taking the KL divergence between the two distributions.

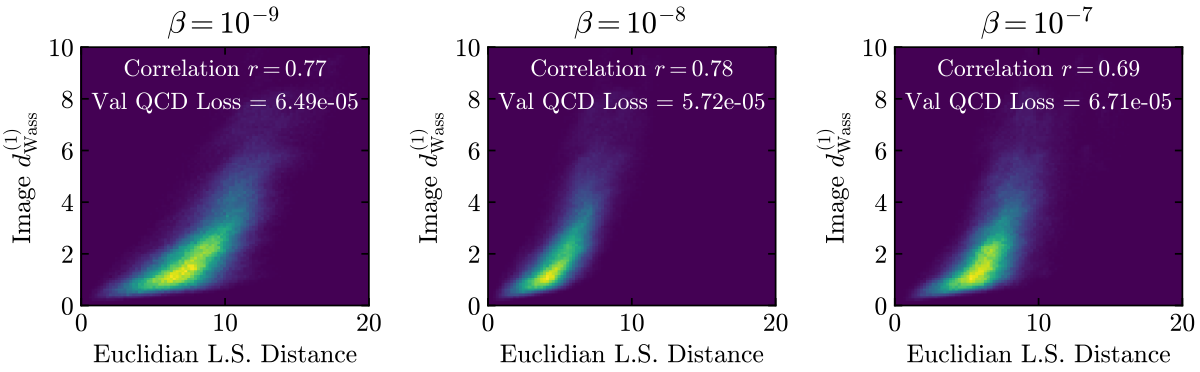


Figure 9.4: Each panel shows correlations between pair-wise distances of events in the QCD test set. The y-axis always denotes the $d^{(1)}_{\text{Wass}}$ distance. The x-axis denotes the Euclidean distance in the latent space. The representation learned by the network is more correlated with the $d^{(1)}_{\text{Wass}}$ distances than the MSE distances (see Fig. 9.1). The latent space distances were computed from networks trained with an MSE loss function, with two downsampling steps.

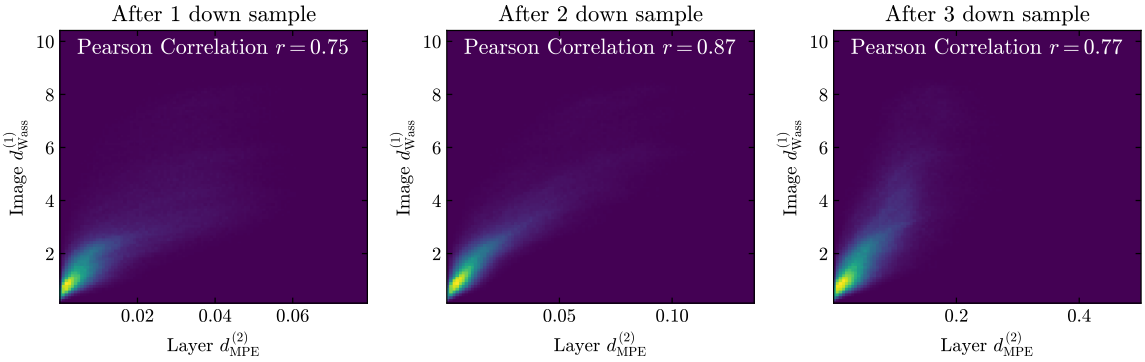


Figure 9.5: Correlations of the pair-wise event distance in the image space with the interior activations of the network after different numbers of downsampling blocks.

The downsampling operations are critical to the production of the latent space. As they combine information from neighboring pixels, they introduce an element of scale which MPE would not exhibit. To verify the importance of downsampling, we show in Fig. 9.5 the pair-wise event distance correlations for the same network at different depths into the encoder. In the first panel, distances on the x-axis are computed in the first downsampling block, where the events are represented as $20 \times 20 \times 5$ tensors and the $d^{(2)}_{\text{MPE}}$ goes across all 2000 "pixels" (see Fig. 9.2). The correlation between the distance in this first downsampling layer and the Wasserstein distance of the events is much larger than the MSE distance between the original events. The correlation further increases from the first down sample block to the second. The correlation then decreases after a third downsampling. Then, when the information is further reduced to the latent space, we get smaller correlations than seen in the early stages of the network.

Although the EMD metric is a p-Wasserstein metric with p = 1, there is no clear reason why p = 1

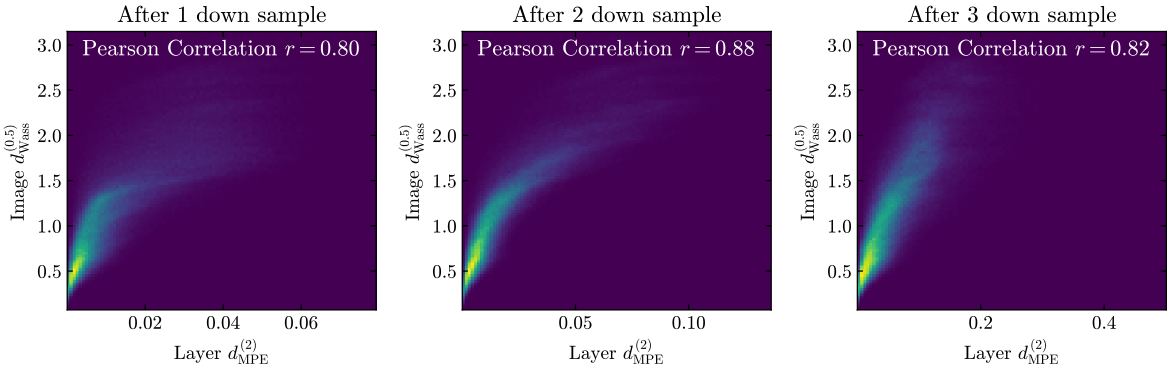


Figure 9.6: Correlations of the pair-wise event distance in the image space with the interior activations of the network after different numbers of down sample blocks. This is the same network as in Fig. 9.5, but now the image distances use a different power of p. The correlation is higher with p = 0.5 than p = 1.

should be preferred to other values. So, next we consider p = 0.5. In Fig. 9.6, we show the correlations for the same network with three down sampling layers but now using $\mathrm{d}^{(0.5)}_{\text{Wass}}$ distances along the y-axis. The distances between events at different layers in the network are $\sim 5\%$ more correlated with $\mathrm{d}^{(0.5)}_{\text{Wass}}$ than $\mathrm{d}^{(1)}_{\text{Wass}}$.

In this section, we have explored the representation of the QCD event distribution that variational autoencoders learn. Our conclusion from this study is that the Euclidean distances between QCD events in the latent space are highly correlated with the p-Wasserstein distances between the events themselves. This is particularly compelling because the VAE is trained with the MSE metric for its loss function and has no direct access to any p-Wasserstein metric. A related question is how the correlations would look if a p-Wasserstein metric were used for training. In that case, we find that the Euclidean distances between events in the latent space of the Wasserstein trained networks are even more correlated with the Wasserstein distances in the image space. Thus, it could be argued that the Wasserstein trained networks learn an even better representation of the QCD distribution than the MSE trained networks. However, the VAE with MSE training worked better for finding the top- and W-jet signals than those trained with a Wasserstein loss. The fact that the method with the "best" latent representation does not yield the best signal separation highlights the challenges of model agnostic anomaly detection.

9.5 Event-to-Ensemble Distance

In the previous section we showed that VAEs tend to work better when MSE loss is used for training than when Wasserstein metrics are used for training and that the Euclidean distance in the latent space correlates strongly with the Wasserstein metric on the data, regardless of the metric used for training. If the power of the VAE for anomaly detection in physical problems stems from it implicitly learning aspects of the p-Wasserstein metrics, we can then ask if there may be a way to use these metrics more directly for anomaly detection, sidestepping the VAE entirely. One way to do this is to use the metric to compute an

event-to-ensemble distance, as we explore in this section.

We would like to use the p-Wasserstein distance, or another metric, to characterize the distance of an event to the background ensemble. There are already several options for using Wasserstein distances to characterize different types of events in the literature, such as k nearest neighbor (kNN) classifiers [1231], "linearized" optimal transport [40], where all the events are compared to a single reference event and this is used to define a new distance, and event isotropy, which compares a given event to an isotropic configuration [1245]. Our goal, using a method like these, is to extract from the background ensemble one or more representative images and to compute the distance of a given signal or background event to those images. This direct event-to-ensemble distance measure can then be compared to the VAE anomaly score, which is also effectively an event-to-ensemble distance.

To compute the direct event-to-ensemble distance we need an algorithm to select or construct fiducial events from the ensemble and a metric with which to compute the distance. As with the VAE architecture, there may be no choice that is optimal for all signals. In choosing the fiducial events, we must decide which sample to choose events from, how to select those events, how many events to use, how to represent the fiducial events (e.g. as images), and how to combine the distances to the different events. To make a fair comparison to the VAE approach, we would like our algorithm for generating fiducial events to depend only on the background sample, independent of what anomalous signal we might search for. Thus we choose the QCD jet event ensemble as our reference sample. To select events from the sample, the simplest possibility is to arbitrarily select some number of random images. However, despite occasionally giving a large AUC for classification, results with random images are very sensitive to fluctuations between the random images. A second possibility that may seem sensible is to take the average of all events in the sample. A third option, which we find to be the most natural, is to use k medoids as we now explain.

With a given metric, which we call the medoid metric, we can compute the pairwise distance $d(x_i, x_j)$ between any two events in the ensemble. Then for each event x we can sum over all the distances to all other events $d(x) = \sum_j d(x, x_j)$. The **medoid** of the ensemble is the event x that minimizes d(x). k **medoids** generalizes this to finding the k events for which the sum of the distances of each event to the closest of the k medoids is minimized. Thus the event fragments into a set of clusters, with each cluster closest to one of the k medoids. k-medoids clustering is similar to k-means clustering when the medoid metric is chosen to be the Euclidean metric, except that k-medoid clustering actually requires the medoid to be one of the events in the set. Medoids have previously been explored in other contexts in [1231, 1232, 1257, 1258]. To use k medoids, we need to choose a value for k and a medoid metric. Then it is natural to take for the event-to-ensemble distance the distance of an event to its closest medoid. Although one could in principle use a different metric to compute the event-to-ensemble distance, it is also most natural to use the same

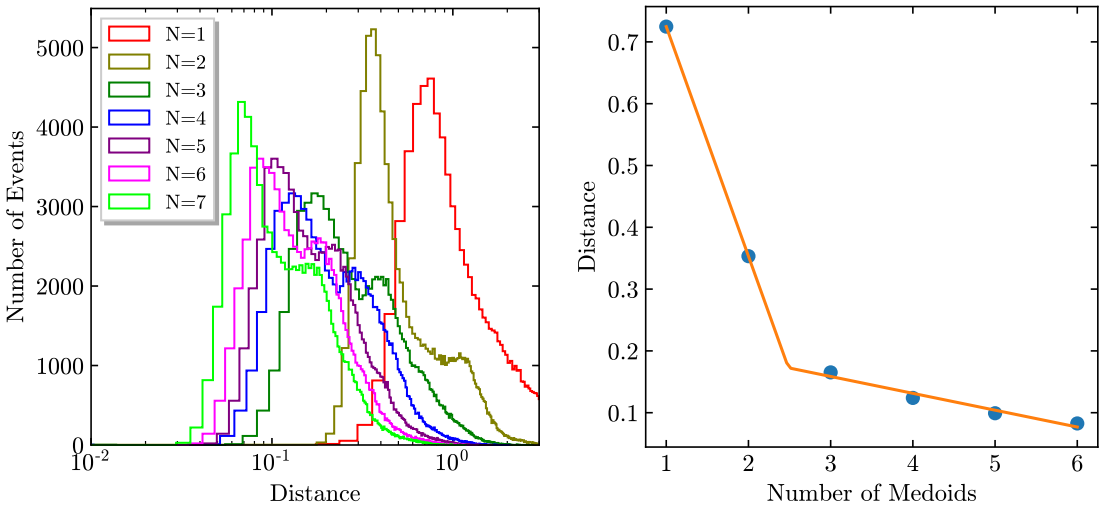


Figure 9.7: Example of the elbow method. Left shows histograms of the 1-Wasserstein distance to the closest medoid, with colors corresponding to the number of medoids. Right plots the peak of each of these histograms, as a function of the number of medoids.

medoid metric that determines the medoids.

Choosing the number of medoids k is challenging to do in a signal independent way. One approach is the elbow method, a common heuristic for determining how many clusters are in a dataset. In our case, to use the elbow method we scan over the number k of medoids, and for each k compute the distances of all the events in our sample to the nearest medoid and histogram the results. There will be a small number of events very close to a medoid and a small number very far from all medoids, so the histograms will have a peak at some finite value of the distance. Moreover, the distance to the peak will decrease monotonically as the number of medoids increases. In many applications the decrease is rapid for small k, but at some k this behavior abruptly slows down. Thus the peak distance as a function of k often has an elbow shape. To determine the elbow location algorithmically we perform a linear regression to an elbow function (two straight lines), and take the first integer value after the elbow as the suggested value of k. The result can be seen in Fig. 9.7. The idea behind the elbow method is that increasing k past the location of the elbow should not give much improvement. Moreover, in the case of anomaly detection, if we have too many medoids, we can get one medoid that looks "signal-like" rather than "background-like". We find that typically $k \sim 2-4$ medoids is selected according to this elbow method.

The main advantage of the elbow method is that it can be automated and used independent of the sample or the use case. However, there often is not a clear elbow. In Fig. 9.7, the elbow is only apparent because we have fit to a piecewise linear function. The data seems to follow more of a power law behavior. In addition, the location of the elbow can be affected by the maximum number of medoids we include in the fit. Additionally, the elbow can only be computed once we've already made the arbitrary choice of the

medoid metric, and of the metric being used for the comparison between the full sample and the reference sample. Finally, there is no reason to expect that the elbow choice of k, which is determined only by the background sample, would be optimal for anomaly detection tasks. Thus, we also consider values of k not determined by the elbow method for this study.

The top of Table 9.2 shows results for top-jet vs. QCD-jet discrimination and W-jet vs. QCD-jet discrimination when QCD jets are used for the reference sample. We show results for different values of k with medoids, using different medoid metrics. We also show the result from using the distance to a single composite average event determined by averaging each pixel intensity over all events in the reference sample. For each case, we include both the AUC and the signal efficiency at a cut which allows only 10% of the QCD events to pass. When we study the elbow for the most common 1-Wasserstein metric, we see reasonable performance for both QCD and top jets, though it is best for neither of them. This is in line with what we expect for unsupervised anomaly detection. For simplicity, we report results where the metric used to select the medoids is the same one used to compute our observable. We could have chosen two different metrics for the medoid metric and that used to compute the event-to-ensemble distance, but restricting to the case where they are the same does not qualitatively change our results. Table 9.2 shows that the number of medoids and the choice of metric matters substantially.

We find that the QCD medoids typically perform better than the average QCD jet. This is not surprising, since the average QCD jet is much more concentrated in the center of the image than any real QCD jet, as can be seen in Fig. 9.8. In this figure, the color shows the fraction of the total p_T in each pixel on a logarithmic scale. We also find better performance for anomaly detection with 5-6 medoids, rather than the 2-4 medoids suggested by the elbow method. When detecting top jets with QCD reference images, we get the best results when the p-Wasserstein metric with $p = 1$ is used to compare images, though we also find reasonable performance for the p-Wasserstein metric with $p = 0.5$ or $p = 2$ (not shown in the table) and when the MAE metric is used. Although MAE is not a physically motivated metric, the performance in this case is not surprising because MAE between events is highly correlated with p-Wasserstein distances between events for QCD jets (the Pearson correlation coefficient between MAE and the 1-Wasserstein distance is 0.87).

That our results depend on the exponent p is suggestive. For the p-Wasserstein metric with larger p we get substantially decreased performance when comparing to QCD reference jets. This suggests that the ideal value of p is related to the relevant scales in the problem: a smaller value of p places comparatively larger emphasis on pixels with smaller differences. This is consistent with results such as [154], which finds better AE performance when pixel intensities are rescaled to emphasize dim pixels. When we choose a better (smaller) value of p, the results are also slightly less sensitive to exactly which QCD reference images are

				Top jet		W jet	
Reference Sample	Metric	Number of medoids	Method	AUC	$\varepsilon_S(\varepsilon_B = 0.1)$	AUC	$\varepsilon_S(\varepsilon_B = 0.1)$
Supervised	-	-	-	**0.94**	**0.81**	**0.96**	**0.91**
QCD Reference	Wass(1)	-	Avg	0.81	0.33	0.62	0.02
		1	Medoid	0.83	0.28	0.63	0.02
		3 (e)	Medoids (min)	0.85	0.43	0.67	0.04
		5	Medoids (min)	**0.87**	**0.54**	0.60	0.05
		7	Medoids (min)	0.87	0.54	0.61	0.05
	Wass(5)	-	Avg	0.53	0.10	0.60	0.03
		1	Medoid	0.62	0.21	0.36	0.03
		3	Medoids (min)	0.66	0.19	0.41	0.05
		4 (e)	Medoids (min)	0.67	0.22	0.41	0.04
		5	Medoids (min)	0.71	0.24	0.43	0.04
	MAE	-	Avg	0.83	0.47	0.71	0.08
		1	Medoid	0.82	0.40	**0.71**	0.07
		3 (e)	Medoids (min)	0.82	0.49	0.61	**0.08**
		5	Medoids (min)	0.83	0.48	0.67	0.08
		7	Medoids (min)	0.83	0.48	0.65	0.08
Top Reference	Wass(1)	-	Avg	0.69	0.17	0.69	0.04
		1	Medoid	0.58	0.20	0.79	0.31
		3 (e)	Medoids (min)	0.32	0.07	0.79	0.53
		5	Medoids (min)	0.45	0.12	**0.84**	**0.62**
		7	Medoids (min)	0.49	0.13	0.83	0.60
	Wass(5)	-	Avg	0.72	0.18	0.40	0.01
		1	Medoid	0.53	0.12	0.52	0.05
		2 (e)	Medoids (min)	0.72	**0.32**	0.70	0.06
		3	Medoids (min)	0.66	0.20	0.61	0.04
		5	Medoids (min)	0.61	0.16	0.54	0.03
	Wass(5)	3 (e)	Medoids (sum)	0.66	0.27	0.66	0.06
		5	Medoids (sum)	**0.73**	0.30	0.58	0.02
		7	Medoids (sum)	0.75	0.31	0.60	0.02
	MAE	-	Avg	0.48	0.05	0.57	0.05
		1	Medoid	0.29	0.04	0.64	0.23
		3 (e)	Medoids (min)	0.25	0.03	0.36	0.03
		5	Medoids (min)	0.31	0.10	0.58	0.31

Table 9.2: Results for QCD vs. signal classification, with signal labeled in the top row. Top rows use a QCD reference sample, and bottom rows use a top reference sample (assuming W events are more "top-like" than QCD events). When there are multiple medoids, distances are combined either by taking the minimum or the sum of the distances to the k different medoids, as denoted in the table. Medoids are selected with the same metric as the one used to compare images. For each metric, we note which number of medoids corresponds to the elbow by (e). The best AUC and $\varepsilon_S(\varepsilon_B = 0.1)$ values for each reference sample are denoted in blue. We indicate the p-Wasserstein distance metric as Wass(p), and the MPE with power $\alpha = 1, 2$ by MAE and MSE, respectively.

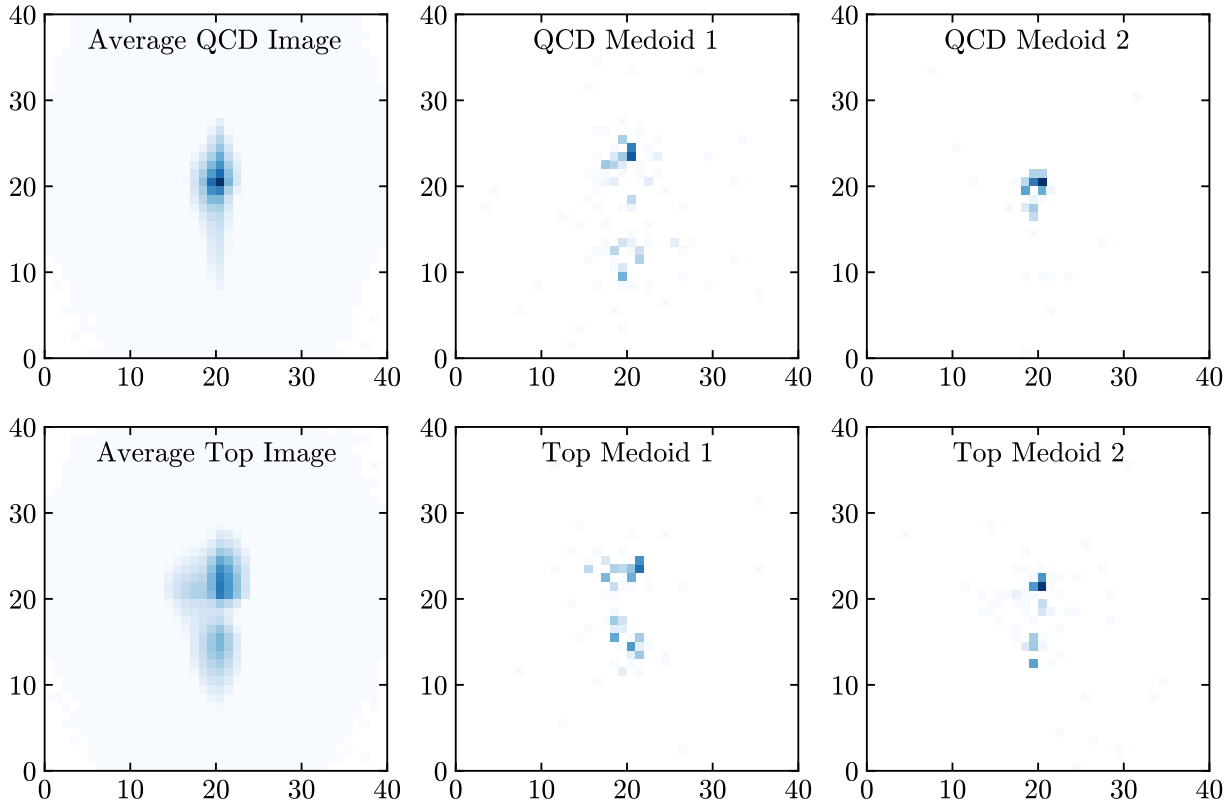


Figure 9.8: Images of example QCD and top events. The top row shows QCD events; the bottom row shows top events. The left column shows the average image in each case, and the other two columns show two medoids computed with the 1-Wasserstein metric. Note medoids are more sparse and varied than average images, and that one of the top medoids appears "QCD-like" when we include multiple medoids.

chosen than when a larger value is chosen.

While autoencoders are often trained on a QCD background, several studies have explored trying to train an AE on alternative samples. One example is [154], which showed that AEs perform poorly when tagging QCD jets if trained on a top jet sample. This can be attributed to top jets being more complex than QCD jets, so that an AE trained on top jets can also reconstruct QCD jets despite them being out of distribution samples. While modifications can be made so that an AE trained on top jets can tag QCD jets [148, 154], requiring sample dependent optimization defeats the point of unsupervised anomaly detection.

Unlike in the case of an AE, event-to-ensemble distances can be directly applied to other reference samples, assuming an applicable metric is selected. We include in the bottom half of Table 9.2 results using a top-jet reference sample for concreteness and brevity, but the method can be easily applied to other reference samples. We find that for top reference samples, the best metric is not the same as for QCD reference samples. In contrast to the case of the QCD reference sample, we find the p-Wasserstein metric with higher p does better for QCD vs. top classification than that with $p = 1$. We suspect that top reference

samples usually prefer a large value of the exponent p because a larger power prioritizes larger differences in pT over smaller ones, and effective classification using top reference samples should deprioritize small pT differences, since the top reference samples are spread out with substantial pixel-to-pixel variation. In contrast, the QCD reference sample case has fewer populated pixels with less variation, so a smaller value of the exponent p is needed. However, this result is signal dependent, in addition to being dependent on the background sample: when trying to use the top reference sample to distinguish QCD vs. W-jet events, we find using the p-Wasserstein metric with p = 1 is better than higher p. We also find that whether the average event or minimum distance to the medoids does better depends on the signal sample — for QCD versus top classification with a top reference sample the average event does better than medoids (unlike the QCD reference sample case), but the opposite is true for QCD vs. W classification. Furthermore, we find the somewhat counter intuitive result that the sum of the distance to the medoids does better than the minimum for top vs. QCD classification with top reference jets, which is surprising because only the distance to the closest event is actually used when determining the medoids. For QCD vs. top classification, using a QCD reference sample still outperforms the top reference sample, but the opposite is true when doing QCD vs. W classification.

Our best results using the event-to-ensemble distance approach are comparable to and even slightly exceed the performance of the VAE in the previous section, which can be seen from comparing Table 9.2 to Table 9.1. This suggests that if we choose a smart, physically motivated metric like the p-Wasserstein distance then we can use the medoid method, which is much faster and simpler than the VAE, and avoids optimizing all the hyperparameters in the VAE network architecture. The trade-off is that we need to put effort into optimizing the metric and choice of k for the medoid approach. This is not surprising, since as we saw in the previous section, distances in the VAE latent space between two separate encoded latent representations are fairly correlated with the p-Wasserstein distance between the original images. This equivalence is further supported by our study of the number of downsampling layers in the previous section. Like in the case of the p-Wasserstein metrics, locality is incorporated in the convolutional VAE on scales other than the arbitrary pixel size due to the convolutions and down samplings.

The ease and speed of using the event-to-ensemble approach is a distinct advantage when compared to AEs, where the architecture, normalization, and parameters all need to be optimized. However, since the ideal metric and fiducial sample selection depend on both the background and signal samples, the signal dependence of the event-to-ensemble approach further suggests that there may be advantages of weakly/semi-supervised learning as compared to unsupervised learning, and that weakly/semi-supervised methods should be explored further. The potential advantages of semi-supervised learning can be further seen from the fact that the best QCD vs. W AUC values come from top reference samples, rather than QCD reference samples.

9.6 Discussion

Using an autoencoder for anomaly detection is particularly challenging, since it must be trained well enough to reconstruct the background, but not so well that it also reconstructs the signal. There are many details about the network configuration that need to be optimized, such as the network size, metric used to compare input and output images, the definition of the anomaly score, and hyperparameters. Many of the results currently in the literature do not sufficiently emphasize these difficulties, so we have attempted in this chapter to characterize and resolve them. To be concrete, we considered detecting boosted hadronic top and W jets over a QCD-jet background. We considered using different metrics for training the VAE; we found that using the more physically-motivated optimal transport-based metrics did not outperform the simpler mean-squared-error metrics, and actually performed slightly worse. We found that the optimal values of various hyperparameters depend on the signal that we are trying to detect and that the optimal hyperparameters for describing the QCD sample are not necessarily those that detect anomalies the best.

In order to understand what the autoencoder has learned, we also studied the autoencoder latent space. The latent space provides a representation of any particular event which can be used to study the background distribution. In order to characterize this latent space, we computed the distance between distinct events. One way to do this is by using the Euclidean distance between quantities in the latent space. Alternatively, if we rely on a more physical, optimal transport based metric, we can compute the distances between images directly. When we compared the two, we found that the event-to-event optimal transport based distances between the background events are highly correlated with the Euclidean distances between events in the latent space of the autoencoder. This suggests that the autoencoder is learning some aspects of optimal transport, despite being trained with only a mean-squared-error based loss function.

This motivated us to develop methods that use optimal transport more directly. By choosing a representation of the QCD background distribution, such as the average QCD image or several medoids of the set of QCD jets, we can directly compute the optimal transport distance to this fiducial sample and use it as an anomaly score. We found that this method is at least as effective as the autoencoder, with the added benefits of being easier and faster to optimize, and generalizing more easily than the autoencoder to more complicated background distributions. We also found that the best choice of optimal transport metric depends on both the new physics signal and the qualities of the expected background distribution. Before using this medoid method in practice, we suggest additional studies of this method, including exploring background sculpting, testing the effects of signal contamination when selecting a background reference sample, and studying the correlation of this anomaly score with known observables.

Although we have shown that the performance of variational autoencoders can be reproduced, and improved upon, by the relatively simpler medoid method, neither approach is very close to optimal for signal

detection. To be quantitative, when trained on a QCD sample, the best autoencoder performance we found gave an AUC of 0.65 for W detection (see Table 9.1). The best performance using medoids with a QCD background gave a slightly better AUC of 0.71 (see Table 9.2). These are both worse than the performance of a fully supervised network which gave a nearly perfect AUC of 0.96. Somewhat surprisingly, we found that when the medoids method was used on a top-jet background sample, it found W jets over QCD better (AUC of 0.84) than when trained on a QCD background. This is comparable to what a supervised network trained to find tops over QCD gives when tested on W vs. QCD (AUC of 0.86). These observations suggest that a path forward might be to use a semi-supervised approach [99, 142, 1259, 1260], where a network is trained with an example signal in mind, and then used for anomaly detection more broadly.

10

Conclusion

The existence of new physics beyond the Standard Model is motivated by an assortment of open questions. A combination of theoretical and data-driven tools will be essential to effectively probing the myriad of BSM models that have been proposed to answer these questions. In this dissertation, we have explored several of these tools.

We have explored many different EFTs, including typical Lorentz-invariant ones, as well as more sophisticated ones where those assumptions break down, such as those with topological defects. First, we discussed several different theories of axion interactions, consisting of EFTs for axions interacting with other axions or non-periodic scalars, EFTs for axions interacting with magnetic monopoles, and EFTs for massive fermions in an axion string background. We found that, even in the presence of mixing, in consistent EFTs axion couplings to gauge fields are always quantized up to violations proportional to the axion mass; that in the presence of abelian gauge fields axion interactions with virtual magnetic monopoles generate a new contribution to the axion potential; and that fermion zero modes delocalize from axion strings when the bulk fermions have a Dirac mass term. Next, we studied several different models which can explain recent experimental deviations from theoretical expectations, including CP violating Higgs portal dark matter, a scalar triplet model, and modified Froggatt-Nielsen models. We considered the viability of CP violating Higgs portal WIMP dark matter models for explaining the galactic center excess of gamma rays and determined their consistency with other experimental constraints. We also performed updated electroweak fits including the

CDF-II W mass measurement and proposed scalar triplet and multiple generation singlet doublet models as explanations of the CDF-II W mass anomaly. Additionally, we defined wrinkles in Froggatt-Nielsen models and demonstrated that, when BSM physics is included in these models, wrinkles generate flexibility in the size of new physics Yukawa couplings.

Finally, we utilized machine learning for precision Top-quark mass measurement and anomaly detection. For Top-quark mass measurement, we showed that complicated ML methods such as DCTR do not improve the precision more than traditional approaches like jet grooming do, but that simultaneously fitting to ensembles of events can lead to additional precision. For anomaly detection, we identify Wasserstein distances to multiple different reference samples as a useful tool and identify challenges with some autoencoder based methods.

There are still many exciting open questions regarding these and other tools to look for BSM physics. Some of these questions have been the subject of additional work since the papers contained in this dissertation were originally published. For examples, see [862,1261,1262] on axion coupling quantization, [1263,1264] on Higgs portal explanations of the gamma ray excess, and [1265] on axion string zero modes. However, our understanding is far from complete, and there are still many opportunities for future work.

A

Additional Information about Axion String Calculations

A.1 Numerical Techniques

We rely on Chebyshev interpolation to numerically solve the zero mode equations of motion (4.28). This method is one of several numerical approaches used for collocation: the discretization of partial differential equations (PDEs) to transform them into matrix equations. Here we give a brief introduction; more details can be found in [1266, 1267].

Chebyshev interpolation provides a method of approximating a smooth function $f(\zeta)$ on the interval $\zeta \in [-1, 1]$ using its values at a finite and discrete set of points. That is, given the function $f(\zeta)$ evaluated at the set of *Cheybshev points*,

$$\zeta_k = \cos\left(\frac{\pi(2k+1)}{2(N+1)}\right), \qquad k = 0, \ldots, N, \tag{A.1}$$

we may approximate it at *any* point in $\zeta \in [-1, 1]$ via the interpolant

$$f_N(\zeta) = \sum_{k=0}^{N} f(\zeta_k) p_k(\zeta) \tag{A.2}$$

where the functions $p_k(\zeta)$, defined such that $p_k(\zeta_n) = \delta_{kn}$, are the degree-N Lagrange polynomials associated to the points (A.1). Importantly, the Lagrange polynomials can be reliably evaluated away from the interpolation points using the *second barycentric form* [1267, 1268],

$$p_k(\zeta) = \frac{w_k}{\zeta - \zeta_k} \Bigg/ \sum_{n=0}^{N} \frac{w_n}{\zeta - \zeta_n}\,, \tag{A.3}$$

where we introduce the *Chebyshev weights*

$$w_k = \sin\left(\frac{2\pi k(N+2)+\pi}{2(N+1)}\right)\,, \qquad k = 0, \ldots, N\,. \tag{A.4}$$

The interpolant (A.2) then provides a degree-N polynomial approximation that exactly agrees with $f(\zeta)$ at the points ζ_k and has an error which decays exponentially in N for all other points on the interval.

The interpolant (A.2) allows us to represent the function $f(\zeta)$ as a $(N+1)$-component vector $\vec{f} = \{f(\zeta_k)\}$. We can similarly represent derivatives of this function working with the derivative matrices $p'_k(\zeta_n)$ and $p''_k(\zeta_n)$, defined by [1267, 1269]

$$\begin{aligned} p'_k(\zeta_n) &= \begin{cases} \dfrac{w_k/w_n}{\zeta_n - \zeta_k} & n \neq k \\ -\displaystyle\sum_{k\neq n} p'_k(\zeta_n) & n = k \end{cases}\,, \\ p''_k(\zeta_n) &= \begin{cases} 2p'_k(\zeta_n)p'_n(\zeta_n) - \dfrac{2p'_k(\zeta_n)}{\zeta_n - \zeta_k} & n \neq k \\ -\displaystyle\sum_{k\neq n} p''_k(\zeta_n) & n = k \end{cases}\,. \end{aligned} \tag{A.5}$$

Any differential equation for $f(\zeta)$ on the interval $\zeta \in [-1, 1]$ can then be approximated by a matrix equation for the function values $\{f(\zeta_k)\}$ by substituting (A.2) for $f(\zeta)$, evaluating the equation at each Chebshev point ζ_k, and using (A.5) to convert differentiation into matrix multiplication. This matrix equation can then be solved by any number of standard numerical techniques.

For example, the equation

$$\left[\frac{d^2}{d\zeta^2} + g(\zeta)\right] f(\zeta) = 0\,, \tag{A.6}$$

for any regular function $g(\zeta)$ on $\zeta \in [-1, 1]$ can be approximated by the matrix equation

$$\sum_{k=0}^{N} (\mathcal{D}_2 + \mathcal{G})_{nk}\, f(\zeta_k) = 0\,, \tag{A.7}$$

where $(\mathcal{G})_{nk} = g(\zeta_k)\delta_{nk}$, $(\mathcal{D}_2)_{nk} = p''_k(\zeta_n)$, which is defined in (A.5), and the number of nodes is chosen to be $N+1$. The solutions to (A.6) are then well-approximated by the interpolant (A.2) constructed from the null space of the matrix $(\mathcal{D}_2 + \mathcal{G})$. As long as N is large enough, this interpolant then provides an approximate solution to (A.6) whose pointwise error decays exponentially in N.

When the equation we aim to solve is defined on the semi-infinite line $r \in [0, \infty)$, we must first map it to the interval $\zeta \in [-1, 1]$ to apply Chebyshev interpolation. We find that the simplest map, the linear

fractional transformation

$$\zeta(\mathrm{r}) = \frac{\mathrm{r} - \mathrm{L}}{\mathrm{r} + \mathrm{L}}, \tag{A.8}$$

works well, where L is a tuneable parameter that broadly controls which r values are sampled by the corresponding Chebyshev nodes (A.1). In practice, we take L to be an $\mathcal{O}(1)$ number times the scale over which the solution varies naturally (set by μ^{-1} for the equations we are interested in) though we ensure that our solutions do not depend on the precise choice of L.

For two-dimensional equations of motion like (4.28), we have a choice in how to apply Chebyshev interpolation. The simplest approach is to expand in Fourier modes, as in §4.3.1, to reduce the problem to a system of coupled ordinary differential equations. That is, we expand each spinor component into their lowest $2\mathrm{J}+1$ Fourier modes

$$\psi_\alpha(\mathrm{r}, \phi) = \sum_{|\ell| \leq \mathrm{J}} \psi_{\alpha,\ell}(\mathrm{r})\, \mathrm{e}^{\mathrm{i}\ell\phi}\,, \tag{A.9}$$

such that (4.28) reduces to a finite set of coupled equations (4.30) involving $2 \times (2\mathrm{J}+1)$ mode functions. We can then interpolate each radial function $\psi_{\alpha,\ell}(\mathrm{r})$ with N+1 nodes to convert (4.30) into a $2\times(2\mathrm{J}+1)\times(\mathrm{N}+1)$-dimensional linear equation. The profiles displayed in Fig. 4.2 were generated with $\mathrm{J} = 41$ and $\mathrm{N} = 61$, though our results are robust to changing both J and N.

Another approach is to interpolate the spinor components $\psi_\alpha(\mathrm{x}, \mathrm{y})$ using a two-dimensional Cartesian grid. This requires first mapping both x and y from $(-\infty, \infty)$ to $[-1, 1]$. We do so via two different maps

$$\zeta(\mathrm{x_i}) = \begin{cases} \arctan(\mathrm{x_i}/\mathrm{L_i}) & 0 \leq \mathrm{m}/\mu \leq 0.8 \\ \tanh(\mathrm{x_i}/\mathrm{L_i}) & 0.4 \leq \mathrm{m}/\mu \leq 0.95 \end{cases}, \tag{A.10}$$

depending on the value of m/μ, where $\mathrm{x_i} = \{\mathrm{x}, \mathrm{y}\}$. Here, the parameters $\mathrm{L_i} = \{\mathrm{L_x}, \mathrm{L_y}\}$ determine the relevant scales in the original x or y coordinate. These maps efficiently sample small (x, y) values while avoiding numerical artifacts at large x and y. Since the solutions elongate in the x-direction as $\mathrm{m} \to \mu$, we used $\mathrm{L_x} \simeq \mathcal{O}(1) \times 1/(\mu - \mathrm{m})$ and $\mathrm{L_y} \simeq \mathcal{O}(1) \times \mu^{-1}$. We find that with these ranges of $\mathrm{L_x}$ and $\mathrm{L_y}$, our Cartesian results agree with each other for overlapping m/μ and that our solutions are independent of our specific choice. For large numbers of nodes, $\mathrm{N_x}, \mathrm{N_y} \geq 90$, our results obtained with a Cartesian grid agree with those obtained using a Fourier mode expansion, but formulating the problem in the Fourier basis is significantly faster.

Finally, to compute the moments of our solutions and properly normalize them, we must numerically evaluate integrals involving these profiles. We do so via quadrature, in which the integral is approximated by a weighted sum of the function's values on the Chebyshev grid. Explicitly, an arbitrary integral of the density $|\psi(\mathrm{r}, \phi)|^2 = |\psi_0(\mathrm{r}, \phi)|^2 + |\psi_3(\mathrm{r}, \phi)|^2$ can be expressed as

$$\begin{aligned}\int d^2r\, a(r,\phi)\,|\psi(r,\phi)|^2 &= \sum_{\alpha,n,\ell,\ell'}\int_{-1}^{1} d\zeta \int_0^{2\pi} d\phi\, r'(\zeta) a_n(\zeta) e^{in\phi} e^{i(\ell-\ell')\phi}\psi_{\alpha,\ell}(\zeta)\psi^*_{\alpha,\ell'}(\zeta) \\ &= 2\pi \sum_{\alpha,k,\ell,n} q_k\, r'(\zeta_k) a_n(\zeta_k)\psi_{\alpha,\ell}(\zeta_k)\psi^*_{\alpha,\ell+n}(\zeta_k)\end{aligned} \tag{A.11}$$

where

$$q_k = \int_{-1}^{1} d\zeta\, p_k(\zeta) \tag{A.12}$$

is the quadrature weight, which can be efficiently numerically evaluated, and

$$a(r,\phi) = \sum_{|n|\le J,k} a_n(\zeta_k) p_k(\zeta) e^{in\phi} \tag{A.13}$$

is an arbitrary weight function, whose n'th Fourier mode has value $a_n(\zeta_k)$ at $r(\zeta_k)$. For instance, to normalize the profile functions in Fig. 4.2, we evaluate (A.11) with $a(r,\phi) = 1$, while to compute the dipole (4.80) we choose $a(r,\phi) \propto r\cos\phi$ and $a(r,\phi) \propto r\sin\phi$.

A.2 Multipole Moments

In this Appendix we provide more details on the derivation of the worldsheet effective action given in (4.77). We follow the method in [871], adapted to the case of a string. Our starting point is the term in the action that couples the gauge field to the four-dimensional current from (4.75), which is pointing along the axion string,

$$S_{em} = -\int d^4x\, j^\mu A_\mu. \tag{A.14}$$

As in the main text, we take $\mu = 0,1,2,3$, $a = 0,3$, and $i,j,k_i = 1,2$.

Our goal is an effective action describing these interactions which is valid when the wavelength of the gauge field is much larger than the characteristic size of the source. In this regime, we can Taylor expand the gauge field about the string source at the origin,

$$A_\mu(t,\vec{x}) = -\sum_{n=0}^{\infty} \frac{1}{n!}\, x^{k_1}\cdots x^{k_n}\left[\partial_{k_1}\cdots\partial_{k_n} A_\mu\right]_{x=(t,0,0,z)}. \tag{A.15}$$

Since our source of interest points along the string (see 4.75), we can drop the terms with $\mu = 1,2$. We are left with the action

$$S_{em} = -\sum_{n=0}^{\infty}\int d^2\sigma \int d^2r\, \frac{1}{n!} x^{k_1}\cdots x^{k_n} j^a(t,\vec{x})\left[\partial_{k_1}\cdots\partial_{k_n} A_a\right](t,z). \tag{A.16}$$

It is useful to rewrite the action in terms of gauge invariant quantities. We first consider the A^0 term, and note that $E^{k_i} = F^{0k_i} = \dot{A}^{k_i} + \partial_{k_i} A^0$. This allows us to exchange radial derivatives of A^0 for time-derivatives

of A^{k_i} and terms involving the electric field, E. We find,

$$\begin{aligned} S_{A_0} = &- \int d^2\sigma \int d^2r\, j^0 \big[A^0\big](t,z) - \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \int d^2r\, j^0 x^{k_1} \cdots x^{k_n} \big[\partial_{k_1} \cdots \partial_{k_{n-1}} E^{k_n}\big](t,z) \\ &+ \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \int d^2r\, j^0 x^{k_1} \cdots x^{k_n} \big[\partial_{k_1} \cdots \partial_{k_{n-1}} \dot{A}^{k_n}\big](t,z) \end{aligned} \tag{A.17}$$

The last term in this expression can be combined with the A^3 term of (A.16) to get a gauge invariant term involving the magnetic field. To see this, we integrate by parts within the $d^2\sigma$ integral, and use current conservation: $\partial_0 j^0 = -\partial_3 j^3$. This allows us to rewrite,

$$\begin{aligned} \left[\int d^2r\, j^0 x^{k_1} \cdots x^{k_n}\right] \partial_{k_1} \cdots \partial_{k_{n-1}} \dot{A}^{k_n} &= \left[\int d^2r\, (\partial_3 j^3)\, x^{k_1} \cdots x^{k_n}\right] \partial_{k_1} \cdots \partial_{k_{n-1}} A^{k_n} \\ &= - \left[\int d^2r\, j^3 x^{k_1} \cdots x^{k_n}\right] \partial_{k_1} \cdots \partial_{k_{n-1}} \partial_3 A^{k_n}, \end{aligned} \tag{A.18}$$

where in the last equality we have integrated by parts again. With this replacement, the sum of the last term of (A.17) and the A^3 term from (A.16) is

$$\begin{aligned} S_B = &- \int d^2\sigma \int d^2r\, j^3 A_3 - \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \int d^2r\, j^3 x^{k_1} \cdots x^{k_n} \partial_{k_1} \cdots \partial_{k_{n-1}} \big(\partial_3 A^{k_n}\big) \\ &- \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \int d^2r\, j^3 x^{k_1} \cdots x^{k_n} \partial_{k_1} \cdots \partial k_{n-1} \big(\partial_{k_n} A_3\big) \\ = &- \int d^2\sigma \int d^2r\, j^3 A_3 - \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \int d^2r\, j^3 x^{k_1} \cdots x^{k_n} \partial_{k_1} \cdots \partial_{k_{n-1}} \varepsilon^{3 k_n j} B_j, \end{aligned} \tag{A.19}$$

where we have used $\partial_i A_3 - \partial_3 A_i = F^{i3} = \varepsilon^{3ij} B_j$. Adding back the remaining terms in S_{A^0} and relabeling indices, we obtain the full action

$$\begin{aligned} S_{em} = &- \int d^2\sigma \int d^2r\, (j^0 A_0 + j^3 A_3) - \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \left[\int d^2r j^0 x^{k_1} \cdots x^{k_n}\right] \partial_{k_2} \cdots \partial_{k_n} E^{k_1} \\ &- \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \left[\int d^2r\, j^3 x^{k_1} \cdots x^{k_n}\right] \partial_{k_2} \cdots \partial_{k_n} \varepsilon^{3 k_1 j} B_j, \end{aligned} \tag{A.20}$$

which can be written in terms of the field strength as

$$S_{em} = - \int d^2\sigma \int d^2r\, (j^0 A_0 + j^3 A_3) - \sum_{n=1}^{\infty} \frac{1}{n!} \int d^2\sigma \left[\int d^2r j^a x^{k_1} \cdots x^{k_n}\right] \partial_{k_2} \cdots \partial_{k_n} F_{k_1 a}. \tag{A.21}$$

This action reduces to (4.77) using the current (4.75) and the normalization condition for $|\mathcal{F}(r,\phi)|^2$.

B

Next-Order Velocity Expansion of the Dark Matter Annihilation Cross Section

As established, to leading order in dark matter velocity, the annihilation signal is set by the pseudoscalar coupling $\mathrm{Im}[y_{h\chi}]$ (and subdominantly by $g_{Z\chi}$), while spin-independent scattering is set by the scalar coupling $\mathrm{Re}[y_{h\chi}]$. However, we would also like to understand whether we can generate the annihilation signal at all in the limit that $y_{h\chi}$ is real. In this limit, the leading velocity independent term vanishes, and we need to consider terms of higher order in the halo velocity v.

For this argument we will neglect the contribution of the Z portal; a $g_{Z\chi}$ consistent with spin-dependent constraints cannot generate a thermal relic annihilation cross section, as it does not have a mass resonance.[1] Thus, for hypothetically viable parameter space it is safe to assume that the Z-mediated annihilation is subdominant.

When $y_{h\chi}$ has vanishing imaginary part, the leading contribution to the spin averaged annihilation amplitude squared is

$$|\mathcal{M}|^2_{\chi\chi\to f\bar{f}} = \frac{4y_{hf}^2 y_{h\chi}^2 m_\chi^2 (m_\chi^2 - m_f^2) v^2}{(m_h^2 - 4m_\chi^2)^2 + m_h^2 \Gamma_h^2} + \mathcal{O}(v^4). \tag{B.1}$$

[1] In fact, the $\mathcal{O}(v^2)$ Z-coupling term does have a mediator resonance, but enhancement is limited by the significantly larger width of the Z boson.

This term is suppressed by the non-relativistic speeds of dark matter, for typical values $v^2 \sim 10^{-6}$, and the magnitude of the purely real coupling is stringently constrained by direct detection. Thus, any allowed parameter space would require precise fine-tuning of the dark matter mass. However, the enhancement obtained from the $m_\chi \to m_h/2$ resonance is limited by the finite width of the Higgs, which is ~ 4 MeV in the SM [975]. Since the branching ratio of $h \to \chi\chi$ near the resonance is vanishingly small due to phase space suppression, we may take 4 MeV as a conservative bound for the Higgs width. Thus, the comparative ratio between annihilation and scattering cross sections, given in (5.5) and (5.13), can be bounded by

$$\left.\frac{\langle \sigma v\rangle_{\rm ann}}{\sigma_{\rm SI}}\right|_{{\rm Im}[y_{h\chi}]=0} \sim \frac{6\times 10^4\ {\rm GeV}^2\ m_\chi^2}{(m_h^2-4m_\chi^2)^2+m_h^2\Gamma_h^2} < \frac{10^5\ {\rm GeV}^2\ m_\chi^2}{m_h^2\Gamma_h^2} \approx 10^9. \tag{B.2}$$

As the current direct detection limits bound the spin-independent scattering rate at $\leq 10^{-10}$ pb, a model without CP-violation may exhibit an annihilation cross section of at most $\mathcal{O}(0.1)$ pb. We emphasize here that these statements are specifically valid for Majorana fermion dark matter, and dark matter models with a different CP-structure could certainly achieve the required hierarchy between annihilation and scattering with sufficient tuning on this resonance.

C

Oblique Parameters in the Singlet-Doublet Model

As mentioned in §5 and §6, the singlet-doublet model primarily contributes to the S and T parameters at loop-level with the new fermions running in the loop. Here, we provide a quick summary of this calculation. We write a generic coupling between a gauge boson X and fermions i, j as $i\gamma^\mu(C^X_{Vij} - C^X_{Aij}\gamma^5)$ where C^X_{Vij} and C^X_{Aij} are the vector and axial vector couplings respectively. In $\overline{MS}$, we find

$$\begin{aligned} i\Pi_{XY}(p^2)g^{\mu\nu} = \frac{-ig^{\mu\nu}}{4\pi^2}\int_0^1 dx \left(\left(C^X_{Vij}C^{Y*}_{Vij} + C^X_{Aij}C^{Y*}_{Aij}\right)p^2x(1-x)\right. \\ \left. + \left(C^X_{Vij}C^{Y*}_{Vij} - C^X_{Aij}C^{Y*}_{Aij}\right)m_i m_j - \left(C^X_{Vij}C^{Y*}_{Vij} + C^X_{Aij}C^{Y*}_{Aij}\right)\Delta\right)\log\frac{\mu^2}{\Delta}, \end{aligned} \tag{C.1}$$

where $\Delta = m_i^2 + x(m_j^2 - m_i^2) - x(1-x)p^2$. The other relevant expression is $\Pi'(p^2)$, which is given by

$$\begin{aligned} i\Pi'_{XY}(p^2)g^{\mu\nu} = \frac{-ig^{\mu\nu}}{4\pi^2}\int_0^1 dx\Bigg\{2\left(C^X_{Vij}C^{Y*}_{Vij} + C^X_{Aij}C^{Y*}_{Aij}\right)x(1-x)\log\frac{\mu^2}{\Delta} \\ + \Big[\left(C^X_{Vij}C^{Y*}_{Vij} + C^X_{Aij}C^{Y*}_{Aij}\right)p^2x(1-x) + \left(C^X_{Vij}C^{Y*}_{Vij} - C^X_{Aij}C^{Y*}_{Aij}\right)m_i m_j \\ - \left(C^X_{Vij}C^{Y*}_{Vij} + C^X_{Aij}C^{Y*}_{Aij}\right)\Delta\Big]\frac{x(1-x)}{\Delta}\Bigg\}. \end{aligned} \tag{C.2}$$

These expressions hold generically for any external electroweak gauge boson. To compute the oblique parameters, as defined in (5.29) - (5.31), we sum over all fermions which contribute to the specific vacuum

polarization and substitute in the relevant masses and couplings. The nonzero Yukawa couplings y and $\tilde{y}$ mix the neutral eigenstates in the low energy effective theory, so S, T, and U are nontrivial functions of the couplings y and $\tilde{y}$ and the masses m_1 and m_2.

D

Correlations in the Electroweak Fit

Our electroweak fit results with different values of M_W were reported in Table 6.2 and in Figs. 6.1 -6.2. For completeness, correlations between different oblique parameters in the fit are reported in Table D.1.

		CDF-II	CDF-II ($2\times$ syst.)	World Average	PDG 2020
Best Fit (U = 0)	(S, T) Correlation	0.925	0.905	0.934	0.914
	(S, T) Covariance Matrix	$\begin{pmatrix} 0.0060 & 0.0043 \\ 0.0043 & 0.0037 \end{pmatrix}$	$\begin{pmatrix} 0.0065 & 0.0052 \\ 0.0052 & 0.0050 \end{pmatrix}$	$\begin{pmatrix} 0.0058 & 0.0041 \\ 0.0041 & 0.0033 \end{pmatrix}$	$\begin{pmatrix} 0.0062 & 0.0047 \\ 0.0047 & 0.0042 \end{pmatrix}$
Best Fit (U floating)	(S, T) Correlation	0.908	0.908	0.908	0.909
	(S, U) Correlation	-0.629	-0.586	-0.639	-0.613
	(T, U) Correlation	-0.859	-0.801	-0.873	-0.836
	(S, T, U) Covariance Matrix	$\begin{pmatrix} 0.0100 & 0.0108 & -0.0058 \\ 0.0108 & 0.0141 & -0.0094 \\ -0.0058 & -0.0094 & 0.0084 \end{pmatrix}$	$\begin{pmatrix} 0.0100 & 0.0107 & -0.0058 \\ 0.0107 & 0.0141 & -0.0093 \\ -0.0058 & -0.0093 & 0.0097 \end{pmatrix}$	$\begin{pmatrix} 0.0100, & 0.0108 & -0.0058 \\ 0.0108 & 0.0141 & -0.0094 \\ -0.0058 & -0.0094 & 0.0082 \end{pmatrix}$	$\begin{pmatrix} 0.0100 & 0.0108 & -0.0058 \\ 0.0108 & 0.0142 & -0.0094 \\ -0.0058 & -0.0094 & 0.0090 \end{pmatrix}$

Table D.1: Correlation matrices and coefficients from the fits of the oblique parameters described in Section 6.3.1 and Section 6.3.2. Different columns correspond to different input M_W measurement scenarios around Eq. (6.2). Results of the fit including (S, T) and excluding (including) U in the list of floating parameters are included in the top (bottom) row.

E

Further Discussion of Flavorful Physics

E.1 Full Set of Consistency Conditions

Here we list the full set of consistency conditions that arise for the Yukawa couplings of the S_1 leptoquark model embedded in an FN setup.

$$
\begin{aligned}
&\left|\Delta^{ij}_{\bar{u}\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{\bar{u}\bar{e}}\cdot\Delta^{\dagger}_{\bar{u}\bar{e}}\cdot\Delta_{\bar{u}\bar{e}}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{\bar{u}\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{\bar{u}\bar{e}}\cdot Y^{\dagger}_{L\bar{e}}\cdot Y_{L\bar{e}}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{\bar{u}\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y^{T}_{Q\bar{u}}\cdot Y^{*}_{Q\bar{u}}\cdot\Delta_{\bar{u}\bar{e}}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{\bar{u}\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y^{T}_{Q\bar{u}}\cdot\Delta^{*}_{QL}\cdot Y_{L\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{d}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{d}}\cdot Y^{\dagger}_{Q\bar{d}}\cdot Y_{Q\bar{d}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{d}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{u}}\cdot Y^{\dagger}_{Q\bar{u}}\cdot Y_{Q\bar{d}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{d}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{QL}\cdot\Delta^{\dagger}_{QL}\cdot Y_{Q\bar{d}}\right)^{ij}\right|.\\
&\left|Y^{ij}_{L\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{L\bar{e}}\cdot Y^{\dagger}_{L\bar{e}}\cdot Y_{L\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{L\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta^{T}_{QL}\cdot\Delta^{*}_{QL}\cdot Y_{L\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{L\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{L\bar{e}}\cdot\Delta^{\dagger}_{\bar{u}\bar{e}}\cdot\Delta_{\bar{u}\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{L\bar{e}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta^{T}_{QL}\cdot Y^{*}_{Q\bar{u}}\cdot\Delta_{\bar{u}\bar{e}}\right)^{ij}\right|,
\end{aligned}
\qquad
\begin{aligned}
&\left|\Delta^{ij}_{QL}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{QL}\cdot\Delta^{\dagger}_{QL}\cdot\Delta_{QL}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{QL}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{d}}\cdot Y^{\dagger}_{Q\bar{d}}\cdot\Delta_{QL}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{QL}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{u}}\cdot Y^{\dagger}_{Q\bar{u}}\cdot\Delta_{QL}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{QL}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{QL}\cdot Y^{*}_{L\bar{e}}\cdot Y^{T}_{L\bar{e}}\right)^{ij}\right|,\\
&\left|\Delta^{ij}_{QL}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{u}}\cdot\Delta^{*}_{\bar{u}\bar{e}}\cdot Y^{T}_{L\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{u}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{u}}\cdot Y^{\dagger}_{Q\bar{u}}\cdot Y_{Q\bar{u}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{u}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{QL}\cdot\Delta^{\dagger}_{QL}\cdot Y_{Q\bar{u}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{u}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{u}}\cdot\Delta^{*}_{\bar{u}\bar{e}}\cdot\Delta^{T}_{\bar{u}\bar{e}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{u}}\right| \geq \frac{1}{16\pi^2}\left|\left(Y_{Q\bar{d}}\cdot Y^{\dagger}_{Q\bar{d}}\cdot Y_{Q\bar{u}}\right)^{ij}\right|,\\
&\left|Y^{ij}_{Q\bar{u}}\right| \geq \frac{1}{16\pi^2}\left|\left(\Delta_{QL}\cdot Y^{*}_{L\bar{e}}\cdot\Delta^{T}_{\bar{u}\bar{e}}\right)^{ij}\right|,
\end{aligned}
\tag{E.1}
$$

They arise from considering the representation of the Yukawas under the SM flavor symmetry group, G_{flavor} (see (7.6)), and constructing the other combinations of Yukawas that transform in the same way. Each combination produces a one-loop contribution via a diagram analogous to Fig. 7.3. The representations of $Y_{Q\bar{u}}$, $Y_{Q\bar{d}}$, $Y_{L\bar{e}}$, Δ_{QL}, and $\Delta_{\bar{u}\bar{e}}$ are listed in Eqs. (7.7) and (7.9). We could also consider additional consistency conditions with more Yukawa couplings on the right-hand side, but those will be sub-dominant to those listed above. The consistency conditions listed above are specific to the spurions we have considered, but this procedure generalizes to arbitrary new spurions under G_{flavor}.

E.2 Calculation of Other Observables

In this appendix we review the contributions of the S_1 leptoquark to various flavor observables. The emphasis is on the dependence on the flavor spurions, Δ_{QL} and $\Delta_{\bar{u}\bar{e}}$, with many details left to the references. In what follows, V is the CKM matrix and v is the SM Higgs vev. We use the CKM parameters as determined in [1173] while the remainder of our inputs are taken from the PDG [1131]. Furthermore, we work with a set of operators where the neutrinos are left in the flavor basis as the processes we consider have either a final state neutrino of a specific flavor, or a sum over all possible final state neutrinos, which can be done in any basis. Therefore, we do not include explicit factors of the PMNS matrix in the expressions for the Wilson coefficients. We assume the leptoquark Yukawas are given in the IR and neglect the running effects. These calculations are used in §7.4 to identify the most relevant constraints and the wrinkles which are useful for evading them. We also employ mostly four-component spinor notation in this appendix for consistency with the majority of the references.

E.2.1 Dipole Moments

First we calculate the contribution of S_1 to the electric and magnetic dipole moments of SM particles. After integrating out the leptoquark, the one loop diagrams of Fig. E.1 can give rise to the effective operators

$$\mathcal{L} \supset c^{R}_{ij}\bar{f}_i\sigma^{\mu\nu}P_R f_j F_{\mu\nu} + \text{h.c.}, \tag{E.2}$$

where $f_{i,j}$ are SM fermions, $F_{\mu\nu}$ is the electromagnetic field strength, and c^{R}_{ij} is the corresponding Wilson coefficient. By matching the diagrams in Fig. E.1 to this operator, we can calculate c^{R}_{ij} values in our setup. See [976, 1081, 1167, 1270–1273] for details of the calculation. Following the notation of [1081], we have

$$\begin{aligned} c^{R}_{ij} = \sum_{\bar{q}} \frac{e}{64\pi^2 m^2_{S_1}} \Big[& m_{\bar{q}} (V^* \Delta_{QL})^{\bar{q}i\,*} \Delta^{\bar{q}j\,*}_{\bar{u}\bar{e}} \Big(Q_{S_1} A(r) - Q_{\bar{q}} B(r) \Big) \\ & + \Big(m_i \Delta^{\bar{q}i}_{\bar{u}\bar{e}} \Delta^{\bar{q}j\,*}_{\bar{u}\bar{e}} + m_j (V^* \Delta_{QL})^{\bar{q}i\,*} (V^* \Delta_{QL})^{\bar{q}j} \Big) \Big(Q_{S_1} \bar{A}(r) - Q_{\bar{q}} \bar{B}(r) \Big) \Big], \end{aligned} \tag{E.3}$$

where the sum is over all possible up-type anti-quarks $\bar{q}$ that can go in the loop, Q is the electric charge,

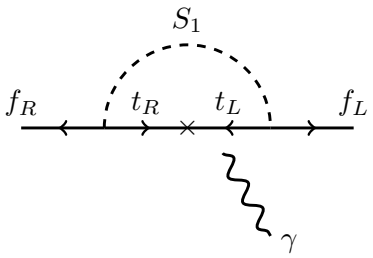


Figure E.1: Feynman diagram (in two-component notation) for the S_1 leptoquark contribution to the dipole operators of charged fermions, including $(g-2)_\mu$. The largest contribution arises from the top quark in the loop. The photon can attach to either internal line in the loop.

$m_{i,j}$ are the masses of the external leptons, $r = m_q^2/m_{S_1}^2$, and the loop functions are defined in the appendix of [1081].

In terms of these Wilson coefficients, the electric and magnetic dipole moments can be written as

$$d_f = 2\ \mathrm{Im}\, c_{ff}^R \qquad a_f = \frac{4m_f}{e}\ \mathrm{Re}\, c_{ff}^R. \tag{E.4}$$

Note that because the two fermions in the operator (E.2) have opposite chirality, all the contributions in (E.3) are proportional to the external fermion or internal quark mass. As a result, unless the Yukawas are very suppressed, the S_1 contribution to EDMs and MDMs are dominated by diagrams with the top quark in the loop, which are proportional to m_t.

E.2.2 Lepton Flavor Violating Observables

The Lagrangian from (E.2) also contributes to LFV decays as [1081, 1167, 1270]

$$\mathrm{BR}\left(\ell \to \ell'\gamma\right) = \frac{48\pi^2}{G_F^2 m_\ell^2}\left(|c_{\ell\ell'}^R|^2 + |c_{\ell'\ell}^R|^2\right). \tag{E.5}$$

Similar to the previous section, dominant contributions to $c_{\ell'\ell}^R$ come from diagrams with the heaviest quarks in the loop. More concretely, we find that in the limit $m_\ell, m_{\ell'} \ll m_{S_1}$

$$c_{\ell\ell'}^R \approx \frac{e m_q}{16\pi^2 m_{S_1}^2}\left[\ln\left(\frac{m_{S_1}^2}{m_q^2}\right) - \frac{7}{4}\right](V^*\Delta_{QL})^{q\ell\,*}\Delta_{\bar{u}\bar{e}}^{q\ell'\,*}. \tag{E.6}$$

These dipole operators also contribute to the well-constrained LFV processes $\mu \to 3e$ and $\mu \to e$ conversion in nuclei. In our leptoquark model, the dipole operator is the only contribution to $\mu \to 3e$, so these branching ratios are directly correlated:

$$\mathrm{BR}(\mu \to 3e) = \frac{\alpha}{3\pi}\left(\log\frac{m_\mu^2}{m_e^2} - \frac{11}{4}\right)\mathrm{BR}(\mu \to e\gamma) \simeq \frac{1}{162}\mathrm{BR}(\mu \to e\gamma). \tag{E.7}$$

The $\mu - e$ conversion process in nuclei, however, also receives contributions from four-fermion operators

coupling the muon and electron to quarks. The effective Hamiltonian for this process can be written [1274, 1275]:

$$\begin{aligned}\mathcal{H} \supset \frac{G_F}{\sqrt{2}} \sum_{q=u,d,s} \Big[&\big(c_{LS}^{(q)} \bar{e} P_R \mu + c_{RS}^{(q)} \bar{e} P_L \mu\big) \bar{q} q + \big(c_{LP}^{(q)} \bar{e} P_R \mu + c_{RP}^{(q)} \bar{e} P_L \mu\big) \bar{q} \gamma_5 q \\ &+ \big(c_{LV}^{(q)} \bar{e} \gamma^\mu P_L \mu + c_{RV}^{(q)} \bar{e} \gamma^\mu P_R \mu\big) \bar{q} \gamma_\mu q + \big(c_{LA}^{(q)} \bar{e} \gamma^\mu P_L \mu + c_{RA}^{(q)} \bar{e} \gamma^\mu P_R \mu\big) \bar{q} \gamma_\mu \gamma_5 q \\ &+ \frac{1}{2} \big(c_{LT}^{(q)} \bar{e} \sigma^{\mu\nu} P_R \mu + c_{RT}^{(q)} \bar{e} \sigma^{\mu\nu} P_L \mu\big) \bar{q} \sigma_{\mu\nu} q + \text{h.c.} \Big]\end{aligned} \tag{E.8}$$

where for the S_1 leptoquark,

$$\begin{aligned} c_{LS}^{(u)} = +c_{LP}^{(u)} = -c_{LT}^{(u)} &= -\frac{1}{2} \frac{v^2}{m_{S_1}^2} (V^* \Delta_{QL})^{11\,*} \Delta_{\bar{u}\bar{e}}^{12\,*} \\ c_{RS}^{(u)} = -c_{RP}^{(u)} = -c_{RT}^{(u)} &= -\frac{1}{2} \frac{v^2}{m_{S_1}^2} (V^* \Delta_{QL})^{12} \Delta_{\bar{u}\bar{e}}^{11} \\ c_{LV}^{(u)} = -c_{LA}^{(u)} &= -\frac{1}{2} \frac{v^2}{m_{S_1}^2} (V^* \Delta_{QL})^{12} (V^* \Delta_{QL})^{11\,*} \\ c_{RV}^{(u)} = c_{RA}^{(u)} &= -\frac{1}{2} \frac{v^2}{m_{S_1}^2} \Delta_{\bar{u}\bar{e}}^{11} \Delta_{\bar{u}\bar{e}}^{12\,*} \end{aligned} \tag{E.9}$$

The conversion rate is then computed by evaluating the overlap integrals of the fermion wave-function and nucleon densities. This has been performed in [1276], assuming the coherent conversion process (where the initial and final state nucleus are the same) dominates. We use the average values of their overlap integrals for the different nuclei (Al and Au).

E.2.3 Leptonic Meson Decays

$P \to \ell\nu$

The EFT for a generic meson decaying to a neutrino and a charged lepton is [1071, 1114, 1277]

$$\begin{aligned}\mathcal{H}_{\text{eff}} = \frac{4 G_F V_{ud}}{\sqrt{2}} \Big[&C_{L,ud\ell\nu}^{V} \left(\bar{u}_L \gamma^\mu d_L\right) \left(\bar{\ell}_L \gamma_\mu \nu_L\right) + C_{R,ud\ell\nu}^{V} \left(\bar{u}_R \gamma^\mu d_R\right) \left(\bar{\ell}_L \gamma_\mu \nu_L\right) \\ &+ C_{L,ud\ell\nu}^{S} \left(\bar{u}_R d_L\right) \left(\bar{\ell}_R \nu_L\right) + C_{R,ud\ell\nu}^{S} \left(\bar{u}_L d_R\right) \left(\bar{\ell}_R \nu_L\right) \Big] + \text{h.c.},\end{aligned} \tag{E.10}$$

where u (d) labels the involved up-type (down-type) quark. In the SM, these decays are mediated by a W exchange and the overall normalization is chosen such that $C_L^V = 1$, with other Wilson coefficients set to zero.

For the S_1 leptoquark, we can show that at the leptoquark mass scale

$$\begin{aligned} C_{L,ud\ell\nu}^{V} &= \frac{\Delta_{QL}^{d\nu} (V^* \Delta_{QL})^{u\ell\,*}}{V_{ud}} \frac{v^2}{4 m_{S_1}^2}, \\ C_{L,ud\ell\nu}^{S} &= \frac{\Delta_{QL}^{d\nu} \Delta_{\bar{u}\bar{e}}^{u\ell}}{V_{ud}} \frac{v^2}{4 m_{S_1}^2}, \end{aligned}$$

In our model there are no couplings to RH down-type quarks, so $C^{S}_{R,ud\ell\nu} = C^{V}_{R,ud\ell\nu} = 0$.

The meson branching ratio to $\ell\nu$ is given by

$$\begin{aligned}\mathrm{BR}\left(P^{-}_{ud} \to \ell\nu\right) = \tau_P \frac{m_P f_P^2 G_F^2 |V_{ud}|^2}{8\pi} m_\ell^2 \left(1 - \frac{m_\ell^2}{m_P^2}\right)^2 \\ \times \left|(C^{V}_{L,ud\ell\nu} - C^{V}_{R,ud\ell\nu}) + \frac{m_P^2}{m_\ell(m_u + m_d)}(C^{S}_{R,ud\ell\nu} - C^{S}_{L,ud\ell\nu})\right|^2,\end{aligned} \tag{E.11}$$

where τ_P is the meson lifetime, m_P is the meson mass, f_P is the meson decay constant, m_ℓ is the final state lepton's mass, and m_u (m_d) is the mass of the up-type (down-type) valence quark of the meson. This equation has been used to calculate the contribution of our model to various leptonic meson decays in the main text.

$P \to \ell\ell'$ and $P \to \nu\nu'$

The Hamiltonian describing a meson P decaying to charged leptons l and l′ is [1072, 1278]

$$\mathcal{H}_{\text{eff}} \supset \frac{4G_F}{\sqrt{2}} \lambda_{CKM} \left[\sum_{X=S,P,9,10} C_X^{qq';\ell\ell'} \mathcal{O}_X^{qq';\ell\ell'} + C_{X'}^{qq';\ell\ell'} \mathcal{O}_{X'}^{qq';\ell\ell'} + \text{h.c.}\right]. \tag{E.12}$$

Here, λ_{CKM} is a combination of two CKM entries involving the valence quarks of the meson, C_X are Wilson coefficients, and their associated operators are

$$\begin{aligned}\mathcal{O}_S^{qq';\ell\ell'} &= \frac{\alpha_{em}}{4\pi}(\bar{q}P_R q')(\bar{\ell}\ell') \qquad & \mathcal{O}_P^{qq';\ell\ell'} &= \frac{\alpha_{em}}{4\pi}(\bar{q}P_R q')(\bar{\ell}\gamma^5\ell') \\ \mathcal{O}_9^{qq';\ell\ell'} &= \frac{\alpha_{em}}{4\pi}(\bar{q}\gamma^\mu P_L q')(\bar{\ell}\gamma_\mu\ell') & \mathcal{O}_{10}^{qq';\ell\ell'} &= \frac{\alpha_{em}}{4\pi}(\bar{q}\gamma^\mu P_L q')(\bar{\ell}\gamma_\mu\gamma^5\ell'),\end{aligned}$$

The operators with a prime on the subscript are obtained by the replacement $P_{L/R} \to P_{R/L}$.

At tree level, our leptoquark only gives rise to decays of D and π via t-channel diagrams, while decays of K, B, and B_s take place at one-loop level and are suppressed. For the tree-level decays, the Wilson coefficients above can be calculated as a function of the leptoquark Yukawa couplings [1072]

$$\begin{aligned}C_9^{qq';\ell\ell'} &= -C_{10}^{qq';\ell\ell'} = -\frac{v^2\pi}{2\alpha_{em}\lambda_{CKM}m_{S_1}^2}(V^*\Delta_{QL})_{q'\ell'}(V\Delta_{QL})^*_{q\ell} \\ C_{9'}^{qq';\ell\ell'} &= C_{10'}^{qq';\ell\ell'} = -\frac{v^2\pi}{2\alpha_{em}\lambda_{CKM}m_{S_1}^2}(\Delta_{\bar{u}\bar{e}})^*_{q'\ell'}(\Delta_{\bar{u}\bar{e}})_{q\ell} \\ C_S^{qq';\ell\ell'} &= C_P^{qq';\ell\ell'} = -\frac{v^2\pi}{2\alpha_{em}\lambda_{CKM}m_{S_1}^2}(\Delta_{\bar{u}\bar{e}})^*_{q'\ell'}(V^*\Delta_{QL})^*_{q\ell} \\ C_{S'}^{qq';\ell\ell'} &= -C_{P'}^{qq';\ell\ell'} = -\frac{v^2\pi}{2\alpha_{em}\lambda_{CKM}m_{S_1}^2}(V^*\Delta_{QL})_{q'\ell'}(\Delta_{\bar{u}\bar{e}})_{q\ell}.\end{aligned} \tag{E.13}$$

For D and π mesons decays we set $\lambda_{CKM} = V^*_{q'b}V_{qb}$ with q, q′ referring to the valence quarks of the meson.

In terms of the Wilson coefficients above, the BR of the meson to ℓ^- and ℓ'^+ is given by [1072, 1278]

$$\begin{aligned}\mathrm{BR}(\mathrm{P}\to\ell^-\ell'^+) = \tau_\mathrm{P}\mathrm{f}_\mathrm{P}^2\mathrm{m}_\mathrm{P}^3\frac{\alpha_\mathrm{em}^2\mathrm{G}_\mathrm{F}^2}{64\pi^3}\lambda_\mathrm{CKM}^2\sqrt{\left(1-\frac{(\mathrm{m}_1-\mathrm{m}_2)^2}{\mathrm{m}_\mathrm{P}^2}\right)\left(1-\frac{(\mathrm{m}_1+\mathrm{m}_2)^2}{\mathrm{m}_\mathrm{P}^2}\right)}\\ \times\left[\left(1-\frac{(\mathrm{m}_1+\mathrm{m}_2)^2}{\mathrm{m}_\mathrm{P}^2}\right)\left|(\mathrm{C}_9-\mathrm{C}_{9'})\frac{\mathrm{m}_1-\mathrm{m}_2}{\mathrm{m}_\mathrm{P}}+\frac{\mathrm{m}_\mathrm{P}}{\mathrm{m}_{\mathrm{q}'}+\mathrm{m}_\mathrm{q}}(\mathrm{C}_\mathrm{S}-\mathrm{C}_{\mathrm{S}'})\right|^2\right.\\ \left.+\left(1-\frac{(\mathrm{m}_1-\mathrm{m}_2)^2}{\mathrm{m}_\mathrm{P}^2}\right)\left|(\mathrm{C}_{10}-\mathrm{C}_{10'})\frac{\mathrm{m}_1+\mathrm{m}_2}{\mathrm{m}_\mathrm{P}}+\frac{\mathrm{m}_\mathrm{P}}{\mathrm{m}_{\mathrm{q}'}+\mathrm{m}_\mathrm{q}}(\mathrm{C}_\mathrm{P}-\mathrm{C}_{\mathrm{P}'})\right|^2\right],\end{aligned} \tag{E.14}$$

where τ_P is the meson lifetime, m_P is the meson mass, and m_1 (m_2) is the mass of the ℓ (ℓ') lepton.

We can use (E.14) to calculate meson decay to a pair of neutrinos too. For that, we should set $\mathrm{m}_1 = \mathrm{m}_2 = 0$ and only keep couplings to LH fermions in the SM. Doing that, we find zero contribution for the S_1 leptoquark.

E.2.4 Semi-leptonic Meson Decays

Next we compute the leptoquark contribution to semi-leptonic meson decays. We ignore constraints from $\mathrm{B}\to\mathrm{K}^{(*)}\ell\ell$, since the S_1 leptoquark only contributes at loop-level, which is subdominant for leptoquark masses above a few TeV [1108]. Instead, we study the more sensitive observables $\mathrm{B}\to\mathrm{D}^{(*)}\mathrm{l}\nu$ and $\mathrm{K}\to\pi\nu\bar{\nu}$, which receive contributions at tree-level.

$\mathrm{R}_{\mathrm{D}^{(*)}}$

$\mathrm{B}\to\mathrm{D}^{(*)}\mathrm{l}\nu$ proceeds at tree-level via the exchange of the W and the leptoquark [1072, 1114, 1277, 1279–1281]. This and other leptoquark models have generated significant interest in the context of $\mathrm{B}\to\mathrm{D}^{(*)}\mathrm{l}\nu$ because some evidence of a lepton flavor non-universal BSM contribution in this channel, captured by the ratio

$$\mathrm{R}_{\mathrm{D}^{(*)}} \equiv \frac{\mathrm{BR}\left(\mathrm{B}\to\mathrm{D}^{(*)}\tau\nu\right)}{\mathrm{BR}\left(\mathrm{B}\to\mathrm{D}^{(*)}\ell\nu\right)}, \tag{E.15}$$

has been detected in various experiments [12–18] ($\ell = \mathrm{e}, \mu$).

When computing the decay rate, integrating the heavy mediators out allows us to work with a set of dimension-6 operators given by

$$\mathcal{H}_\mathrm{eff} = \frac{4\mathrm{G}_\mathrm{F}\mathrm{V}_\mathrm{cb}}{\sqrt{2}}\left(\mathcal{O}_\mathrm{LL}^\mathrm{V} + \sum_{\substack{\mathrm{X=S,V,T}\\ \mathrm{M=L,R}}}\mathrm{C}_\mathrm{ML}^\mathrm{X}\mathcal{O}_\mathrm{ML}^\mathrm{X}\right) \tag{E.16}$$

where

$$\mathcal{O}_\mathrm{ML}^\mathrm{S} \equiv (\bar{\mathrm{c}}\mathrm{P}_\mathrm{M}\mathrm{b})(\bar{\tau}\mathrm{P}_\mathrm{L}\nu) \quad \mathcal{O}_\mathrm{ML}^\mathrm{V} \equiv (\bar{\mathrm{c}}\gamma^\mu\mathrm{P}_\mathrm{M}\mathrm{b})(\bar{\tau}\gamma_\mu\mathrm{P}_\mathrm{L}\nu) \quad \mathcal{O}_\mathrm{ML}^\mathrm{T} \equiv (\bar{\mathrm{c}}\sigma^{\mu\nu}\mathrm{P}_\mathrm{M}\mathrm{b})(\bar{\tau}\sigma_{\mu\nu}\mathrm{P}_\mathrm{L}\nu) \tag{E.17}$$

Note that we have split apart the contributions to the vector operator such that the Wilson coefficients only

capture leptoquark contributions.

For the process of interest, the helicity amplitude we wish to compute is

$$-i\mathcal{M} = \langle \ell(p_\ell, \lambda_\ell), \bar{\nu}_\ell(p_\nu), D^{(*)}(p_\mu, \varepsilon(\lambda_M)) | \mathcal{H}_{\text{eff}} | B(p_B) \rangle. \tag{E.18}$$

Each of these operators can be split apart into the constituent quark and lepton bilinears, which allows us to split apart the total amplitude into a product of hadronic and leptonic amplitudes. Details of the calculation can be found in [1282, 1283]. The leptonic amplitudes, which are generically functions of various angles, are identical for both D and D^*, while the hadronic amplitudes, which are functions of q^2, vary and are determined by the specific helicity of the $D^{(*)}$ meson. The leptonic amplitudes can be found in multiple references, including [1282]; the expressions for the relevant hadronic functions are taken from [1279, 1284].[1]

To compute the differential decay rate, we use

$$\frac{d\Gamma}{dq^2\, d\cos\theta} = \frac{1}{2m_B} \sum_\ell \left| \mathcal{M}(q^2, \cos\theta) \right|^2 \frac{\sqrt{(m_B + m_D)^2 - q^2}\sqrt{(m_B - m_D)^2 - q^2}}{256\pi^3 m_B^2} \left(1 - \frac{m_\tau^2}{q^2}\right) \tag{E.19}$$

where we sum over neutrinos in the final state. Performing the angular integral over the leptonic functions first, we recover Eqs. (B.6) and (B.8) in [1283] for the differential decay rates of $B \to D\tau\nu$ and $B \to D^*\tau\nu$ respectively. This is the result for a τ in the final state, but making the replacement $m_\tau \to m_\ell$ gives us the expression for decays involving any of the SM leptons. The total decay rate can then be obtained by performing the q^2 integral over the interval $[m_\ell^2, (m_B - m_D)^2]$.

The expressions from (B.6) and (B.8) in [1283] are given in terms of the Wilson coefficients defined in (E.16), therefore, the last ingredient required to complete this computation is the set of pertinent Wilson coefficients for the leptoquark model. They are given by

$$\begin{aligned} C_{LL}^S &= -\frac{v^2}{4m_{LQ}^2} \frac{\Delta_{QL}^{3j}\, \Delta_{\bar{u}\bar{e}}^{23}}{V_{cb}} \\ C_{LL}^V &= \frac{v^2}{4m_{LQ}^2} \frac{\Delta_{QL}^{3j} (V^* \Delta_{QL})^{23\,*}}{V_{cb}} \\ C_{LL}^T &= \frac{v^2}{16m_{LQ}^2} \frac{\Delta_{QL}^{3j}\, \Delta_{\bar{u}\bar{e}}^{23}}{V_{cb}} \end{aligned} \tag{E.20}$$

$K \to \pi\nu\bar{\nu}$

The decays $K^+ \to \pi^+\nu\bar{\nu}$ and $K_L \to \pi^0\nu\bar{\nu}$ can be described with an effective Hamiltonian very similar to (7.19) [1107, 1285]:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \left[\mathcal{H}_{\text{eff}}^{(c)} + V_{td}^* V_{ts} (C_L^{K\nu} \mathcal{O}_L^{K\nu} + C_R^{K\nu} \mathcal{O}_R^{K\nu}) + \text{h.c.} \right] \tag{E.21}$$

[1] The correct sign of $h_{T_3}(w)$ is in [1279].

where

$$\mathcal{O}^{K\nu}_{L(R)} = \frac{\alpha_{em}}{4\pi}(\bar{d}\gamma^{\mu}P_{L(R)}s)(\bar{\nu}\gamma_{\mu}(1-\gamma_5)\nu), \tag{E.22}$$

and $\mathcal{H}^{(c)}_{eff}$ includes operators that encode physics below the weak scale. The branching ratios for $K^+ \to \pi^+\nu\bar{\nu}$ and $K_L \to \pi^0\nu\bar{\nu}$ are then written as

$$\begin{aligned} BR(K^+ \to \pi^+\nu\bar{\nu}) &= \kappa_+ \left[\left(\frac{Im(\lambda_t X^{K\nu})}{\lambda^5}\right)^2 + \left(-P_{(u,c)} + \frac{Re(\lambda_t X^{K\nu})}{\lambda^5}\right)^2\right] \\ BR(K_L \to \pi^0\nu\bar{\nu}) &= \kappa_L \left(\frac{Im(\lambda_t X^{K})}{\lambda^5}\right)^2 \end{aligned} \tag{E.23}$$

where $X^{K\nu} = -\sin^2\theta_W(C_L^{K\nu} + C_R^{K\nu})$, $\lambda_t = V_{td}^* V_{ts}$ and $\lambda = 0.2255$ is the Wolfenstein parameter of the CKM matrix. The κ-factors encode input from hadronic matrix elements. Following [1107], we take $\kappa_+ = (5.27 \pm 0.03) \times 10^{-11}$ and $\kappa_L = (2.27 \pm 0.01) \times 10^{-10}$. The quantity $P_{(u,c)} = 0.41 \pm 0.05$ encodes contributions from charm and light-quark loops. These two decays are related via the Grossman-Nir bound [1286].

The SM Wilson coefficient $C_L^{K\nu\,SM}$ is the same as (7.21), while the leptoquark contribution is

$$C_L^{K\nu} = \frac{v^2}{m_{S_1}^2}\frac{\pi}{2\alpha_{em}}\frac{\Delta_{QL}^{2k}\Delta_{QL}^{1k\,*}}{\lambda_t} \tag{E.24}$$

We set a constraint on the leptoquark mass by demanding that the total predicted branching ratio (including the SM contribution) be less than the 2σ upper limit of the measured branching ratio in [1127]: $BR(K^+ \to \pi^+\nu\bar{\nu}) < 1.88 \times 10^{-10}$. The analogous limit for K_L decays set by the KOTO experiment, $BR(K_L \to \pi^0\nu\bar{\nu}) < 4.9 \times 10^{-9}$ [1287] is not yet competitive in the context of this model.

E.2.5 $Z \to \ell\ell'$

Virtual corrections involving SM fermions and the S_1 leptoquark can also contribute to lepton flavor universality violating decays of the SM gauge bosons. The strongest bound on the leptoquark comes from measurements of the $Z \to \ell\ell'$ decays, which are constrained by ATLAS [1288]. Constraints on Z decays can be cast as bounds on anomalous couplings of the Z boson, δg, where

$$\mathcal{L} \supset \frac{g}{\cos\theta_W}\sum_{f,i,j}\bar{f}_i\gamma^{\mu}\left[(\delta_{ij}g_{SM}^{f_L} + \delta g_{ij}^{f_L})P_L + (\delta_{ij}g_{SM}^{f_R} + \delta g_{ij}^{f_R})P_R\right]f_j Z_{\mu}, \tag{E.25}$$

with $g_{SM}^{f_L} = T_3^f - Q_f \sin^2\theta_W$ and $g_{SM}^{f_R} = -Q_f \sin^2\theta_W$ being the left- and right-handed fermion couplings to the Z boson in the SM.

The S_1 leptoquark contributions to these anomalous couplings have been worked out in [1289, 1290]. In particular [1290] includes additional finite terms that are numerically important. The S_1 leptoquark contributions to the charged lepton couplings of the Z is

$$\delta g_{ij}^{\ell\, L(R)} = \frac{N_c}{16\pi^2} w_{L(R)}^{tj} (w_{L(R)}^{ti})^* \left[(g_{SM}^{u_{L(R)}} - g_{SM}^{u_{R(L)}}) \frac{x_t(x_t - 1 - \log x_t)}{(x_t-1)^2} + \frac{x_Z}{12} F_{L(R)}(x_t) \right] + \frac{N_c}{48\pi^2} x_Z \sum_{k=u,c} w_{L(R)}^{kj} (w_{L(R)}^{ki})^* \left[g_{SM}^{u_L(R)} \left(\log x_Z - i\pi - \frac{1}{6} \right) + \frac{1}{6} g_{SM}^{\ell_{L(R)}} \right] \tag{E.26}$$

where $x_Z = m_Z^2/m_{S_1}^2$, $x_t = m_t^2/m_{S_1}^2$, $w_L^{ij} = (V^* \Delta_{QL})^{ij}$, $w_R^{ij} = \Delta_{\bar{u}\bar{e}}^{ij}$, and $F_{L(R)}(x)$ are loop functions, which can be found in [1290].

[1291] sets bounds on combinations of these anomalous couplings with a variety of flavor ansätze, by combining the LFV decay bounds with LEP data at the Z-pole [1003]. To extract a constraint on the S_1 leptoquark, we simply demand that the anomalous couplings computed above satisfy their bounds assuming generic LFV coupling, which limits

$$\sqrt{|\delta g_{12}^{\ell_L}|^2 + |\delta g_{12}^{\ell_R}|^2} < 1.2 \times 10^{-3}, \qquad \sqrt{|\delta g_{23}^{\ell_L}|^2 + |\delta g_{23}^{\ell_R}|^2} < 4.8 \times 10^{-3}. \tag{E.27}$$

The eμ bound is most constraining for the anarchic and vanilla FN flavor ansätze, while the μτ bound is strongest with the additional wrinkles from (7.32).

E.2.6 Meson Mixing

The leptoquark S_1 also contributes at the one-loop level to operators in the SM that are responsible for meson mixing. In particular for the down type quarks, the important operator for meson mixing is the dimension-six, four-quark bilinear

$$\mathcal{H}_{mix} \supset C_{mix}^{ij} (\bar{d}_L^i \gamma^\mu d_L^j)(\bar{d}_L^i \gamma^\mu d_L^j). \tag{E.28}$$

The associated Wilson coefficient for this operator generated by the S_1 leptoquark is [1281]

$$C_{mix}^{ij} = \frac{1}{128\pi^2 m_{S_1}^2} \sum_{k=1}^{3} \left[(\Delta_{QL}^{ik\,*}) \Delta_{QL}^{jk} \right]^2, \tag{E.29}$$

where the sum above is over all neutrino flavors. Several experimental quantities of interest can then be derived from this; for instance (in the limit of negligible CP violating phases) the mass difference Δm between the mass eigenstates of the oscillating meson is given by

$$\Delta m = \frac{\langle P | \mathcal{H}_{mix} | \bar{P} \rangle}{m_P} = \frac{C_{mix}^{ij}}{m_P} \langle P | (\bar{d}_L^i \gamma^\mu d_L^j)\ (\bar{d}_L^i \gamma^\mu d_L^j) | \bar{P} \rangle\ . \tag{E.30}$$

Here, P denotes the meson whose constituent down-type quarks are in the i,j generation. The non-perturbative hadronic matrix element above is

$$\langle P | (\bar{d}_L^i \gamma^\mu d_L^j)\ (\bar{d}_L^i \gamma^\mu d_L^j) | \bar{P} \rangle = \frac{2}{3} f_P^2 m_P^2 B_P, \tag{E.31}$$

where f_P is the meson decay constant and B_P is the meson bag factor, which can be extracted from lattice computations [1292–1294].

In order to reduce uncertainties from the hadronic matrix elements, we find it advantageous to compare ratios of the matrix elements of the mixing operator (as given in (7.33)). We define

$$C_{B_q} e^{2i\varphi_{B_q}} = \frac{\langle B_q | \mathcal{H}_{\text{mix}}^{\text{SM+NP}} | \bar{B}_q \rangle}{\langle B_q | \mathcal{H}_{\text{mix}}^{\text{SM}} | \bar{B}_q \rangle} \,, \tag{E.32}$$

where $q = d, s$ and by definition in the SM, $C_{B_q} = 1$ and $\varphi_{B_q} = 0$. By definition, the C_{B_q} are free from the non-perturbative matrix elements and depend only on perturbative, short-distance Wilson coefficients. The aforementioned ratio is experimentally determined by the UTFit collaboration [1129, 1172, 1173], and can be understood as a short-distance proxy for the mass difference Δm. In principle, there can be intricate interplay between the phases of leptoquark couplings, leading to interference with the SM contributions in this ratio. In this work, we avoid making any assumptions on the underlying complex phases of the leptoquark couplings in C_{B_q}, and simply compute the absolute value of C_{B_q}.

Additional CP violation from BSM physics is also strongly constrained by other meson mixing measurements, especially in the Kaon system. The quantity of interest is ε_K, which, following standard assumptions (see e.g. [1295]), is given by

$$\varepsilon_K = \frac{1}{4} \frac{\langle K_0 | \mathcal{H}_{\text{mix}} | \bar{K}_0 \rangle}{\langle \bar{K}_0 | \mathcal{H}_{\text{mix}} | K_0 \rangle} - \frac{1}{4} \,. \tag{E.33}$$

To account for ε_K, which is much more constraining than the Kaon mass difference, we define

$$C_{\varepsilon_K} = \frac{\text{Im}\langle K^0 | \mathcal{H}_{\text{mix}}^{\text{SM+NP}} | \bar{K}^0 \rangle}{\text{Im}\langle K^0 | \mathcal{H}_{\text{mix}}^{\text{SM}} | \bar{K}^0 \rangle}, \tag{E.34}$$

where again $C_{\varepsilon_K} = 1$ in the SM.

For all of these quantities, we compute the leptoquark contributions using (E.29). We compare to the SM matrix elements, which are computed following [1295–1297], including the scale-independent, short-distance QCD corrections. Then we set constraints using the latest results from UTFit [1173].

We do not consider effects of the S_1 leptoquark on mixing in mesons with up-type quarks such as the D^0, primarily due large hadronic undertainties [1298, 1299] in current SM predictions that make it difficult to glean any information from new physics contributions.

F

More Machine Learning Details

F.1 DCTR on a toy model

The overall idea of the DCTR method to extract the top mass is inspired by finding the value of the Monte Carlo mass (and other tune parameters) which are most likely to have produced the data. In order to assess this, one needs the likelihood function covering the range of data and parameters. However, this is extremely difficult to obtain. Instead, DCTR uses the fact that, given two data sets, the likelihood ratio between the data sets is given by an ideal classifier. While we cannot access the likelihood function itself, it is still possible to find the parameters which maximize the likelihood using the ratio. We gain access to an approximation of the likelihood ratio using a well-trained, flexible, neural network, which is close to an ideal classifier. In this appendix, we review the two of the main components necessary for DCTR to work, (a) training a parameterized neural network to find the likelihood ratio and (b) using the likelihood ratio to maximize the likelihood and infer the most probable Monte Carlo parameter.

In a parameterized neural network, unobserved properties are included as input to the network. This can be useful when needing to scan over a property. For instance, when looking for BSM physics, the mass of a new resonance is unknown, and a classifier trained at one mass will be sub-optimal if the mass is substantially different. Rather than training many classifiers for different masses, one can train a single classifier where the mass is included as an input. This helps the classifier to interpolate between masses and reduces the

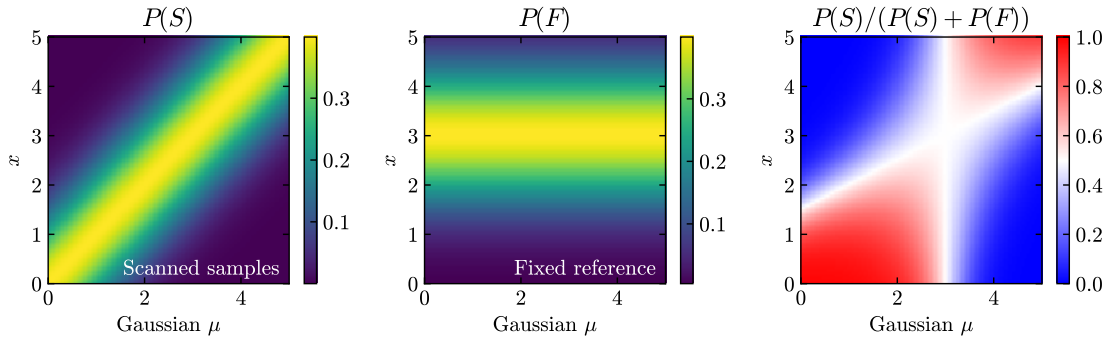


Figure F.1: DCTR works by using a classifier to approximate the likelihood ratio. In this example, we use Gaussian distributions to show the procedure using exact likelihood ratios. The left panel shows the probability density for a sample of data in which the mean of the Gaussian can take any value between 0 and 5; this is referred to as the scanned sample. The middle panel shows the probability density for the fixed sample, where the mean is always set to 3 and μ has no meaning. The right panel shows what the probability is for a point in the μ, x parameter space to come from the scanned sample as opposed to the fixed sample, which is the result of an ideal classifier.

amount of training data needed, because features are shared across the feature space.

The first step is to use a parameterized neural network as a classifier to derive an estimate for the likelihood ratio between samples. In our full set up, we include the Monte Carlo mass of the top quark as an input parameter. For this appendix, we will start by considering a simpler example. Consider the case of Gaussian distributions with different values for the mean μ. Draw samples x from this Gaussian, where the mean μ is changed for each draw. This produces a two dimensional array $\mathrm{S} = (\mu, \mathrm{x})$. We can train a network $\mathrm{f}(\mu, \mathrm{x})$ to distinguish between S and a uniform 2D distribution, U. This network will yield the ratio of their probability densities at any given point,

$$\mathrm{f}(\mu, \mathrm{x}) = \frac{\mathrm{S}(\mu, \mathrm{x})}{\mathrm{S}(\mu, \mathrm{x}) + \mathrm{U}(\mu, \mathrm{x})} \ . \tag{F.1}$$

The probability density for the data distribution can easily be solved for in terms of the output of the network,

$$\mathrm{S}(\mu, \mathrm{x}) = \frac{\mathrm{f}(\mu, \mathrm{x})}{1 - \mathrm{f}(\mu, \mathrm{x})} \ \mathrm{U}(\mu, \mathrm{x}) \ . \tag{F.2}$$

From this, we can use f to obtain the probability density of S for any μ and x within the range of the training data. This is possible because we know the probability density of the uniform distribution which we were using as a reference. In fact, we did not need to use a uniform distribution at all; the processes generalizes to choosing a different fixed reference sample. For the rest of this simple example, we will use a Gaussian with fixed mean $\mu = 3$ as the reference sample. In the main text, we use a sample of events with a fixed Monte Carlo top-quark mass and fixed tune as a reference, since a uniform distribution does not make sense in the context of jets with differing numbers of particles.

This is shown in Fig. F.1. Specifically, we show this initial setup and the idealized network output. The

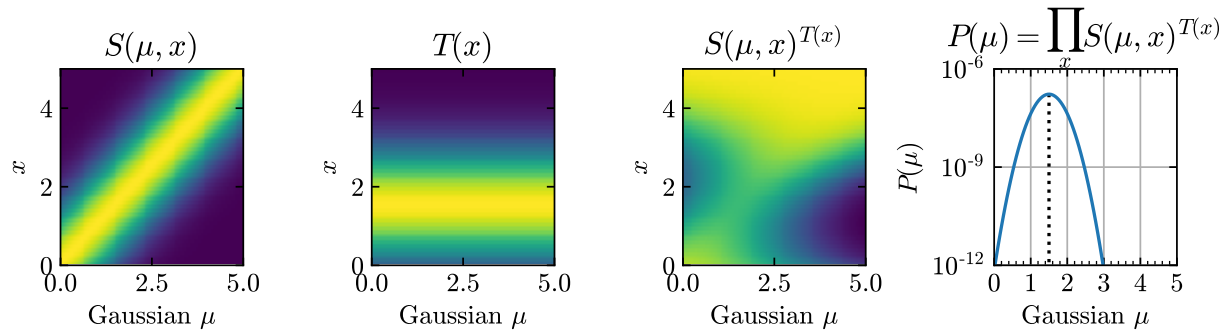


Figure F.2: If the full probability density is known (shown in the left panel), the parameter of an unknown sample (second panel) can be obtained by maximizing the likelihood. The third panel shows the probability of observing each element in the unknown set, T. The final panel shown the total probability, obtained by taking the product of the individual probabilities, and is maximized at the true underlying parameter.

first panel shows the probability density for our scanned sample P(S), where the x-axis denotes the Monte Carlo parameter and the y axis denotes the observed value. The middle panel displays the same probability density for the fixed reference sample, P(F). We can see that P(F) is uniform for all values along the x-axis (the fake input parameters) but has a non-uniform y-axis, since the sample is drawn with a specific fixed parameter. The final panel shows the ratio of these probability densities,

$$\mathrm{R}(\mu, \mathrm{x}) = \frac{\mathrm{P}\big(\mathrm{S}(\mu, \mathrm{x})\big)}{\mathrm{P}\big(\mathrm{S}(\mu, \mathrm{x})\big) + \mathrm{P}\big(\mathrm{F}(\mathrm{x})\big)}, \tag{F.3}$$

which would be the output of an ideal network trained to classify events as coming from the scanned or reference sets.

The next step is to use the classifier to infer the most probable Monte Carlo parameter for a new dataset. In this example, we use $\mu = 1.5$ as the new dataset, and denote this set by T(x). If we had access to the full likelihood $\mathrm{S}(\mu, \mathrm{x})$, we could infer the value of μ by multiplying the probabilities from each event in T to maximize the likelihood,

$$\hat{\mu} = \underset{\mu}{\mathrm{argmax}} \prod_{\mathrm{x_i} \in \mathrm{T}} \mathrm{S}(\mu, \mathrm{x_i}) = \underset{\mu}{\mathrm{argmax}} \prod_{\mathrm{x}} \mathrm{S}(\mu, \mathrm{x})^{\mathrm{T(x)}}. \tag{F.4}$$

In going from the first expression to the second expression, we transition from discrete to continuous distributions. An example of this is shown in Fig. F.2. The first panel again shows the true probability density $\mathrm{S}(\mu, \mathrm{x})$ and the second panel shows the distribution of x for the unknown set. The third panel shows $\mathrm{S}(\mu, \mathrm{x})^{\mathrm{T(x)}}$ which is the probability of observing each element in T given S. In the final panel, we take the product of the probabilities to obtain the total probability of obtaining the data T as a function of μ. The true value is the most probable one.

When using DCTR, we do not actually have access to $\mathrm{S}(\mu, \mathrm{x})$ but only the ratio $\mathrm{R}(\mu, \mathrm{x})$. However, a similar procedure still works. We want to maximize the probability, but now we must also include the

reference set. This is done by maximizing the likelihood that the new set will be classified as part of the scanned set while the reference set will be classified as the reference set. We define the total probability of events from T(x) to get classified as coming from the scanned sample as

$$P_T(\mu) = \prod_x R(\mu, x)^{T(x)}. \tag{F.5}$$

Similarly, let $P_F(\mu)$ define the probability of events from the fixed sample F(x) getting classified correctly,

$$P_F(\mu) = \prod_x \big(1 - R(\mu, x)\big)^{F(x)} \ . \tag{F.6}$$

Combining these two expressions yields total probability of classification,

$$C(\mu) = \prod_x R(\mu, x)^{T(x)} \big(1 - R(\mu, x)\big)^{F(x)}. \tag{F.7}$$

The value of μ which maximizes $C(\mu)$ then corresponds to the most probable value to have produced the test dataset, T.

We make this more explicit in Fig. F.3. The panels on the left show the output of the ideal classifier (the ratio of probability densities from the scanned and fixed sample) for the scanned (top) and reference (bottom) datasets. The second column shows the new test PDF which we are trying to infer (top) and the reference PDF (bottom). The third column displays the classifier output (R or 1-R) convolved with the probability distributions. The top panel in the last column displays the product of these. In the bottom right panel, we show the total probabilities. The new set $P_T(\mu)$ is shown in blue, the fixed reference set $P_F(\mu)$ is shown in orange, and $C(\mu)$ is the green line. Note that the blue and orange lines have quite different shapes, however, when they are combined to make the green line, it is maximized at $\mu = 1.5$, which is the value of the test set. The works for all values of μ; a video showing a scan can be found at https://bostdiek.github.io/Videos/DCTR_Gaussian_Example.mp4.

While DCTR is overly complicated for a single dimension, it can prove useful when the datasets have many dimensions. We now generalize the method by taking the single observable x to be set of observations, $x_i \to \mathbf{X}_i$, where the subscript represents a given event. Similarly, the underlying parameter μ is generalized to many model parameters $\mu \to \theta$. In many dimension, an explicit likelihood ratio can be challenging to obtain, thus a neural network will be used as an approximation. The network is trained to classify events from a scanned set $x_i \in x_{\theta_S}$ from events in a fixed reference set $x_i \in x_{\theta_0}$. We now represent the network output by $f(x, \theta)$ and train it to maximize the probability of correctly assigning the training events. Namely,

$$\begin{aligned} f(x, \theta) &= \underset{f'}{\text{argmax}} \left(\prod_{x_i \in x_{\theta_S}} f'(x_i, \theta) \times \prod_{x_i \in x_{\theta_0}} \big(1 - f'(x_i, \theta)\big) \right) \\ &= \underset{f'}{\text{argmin}} \left(- \sum_{x_i \in x_{\theta_S}} \log f'(x_i, \theta) - \sum_{x_i \in x_{\theta_0}} \log \big(1 - f'(x_i, \theta)\big) \right). \end{aligned} \tag{F.8}$$

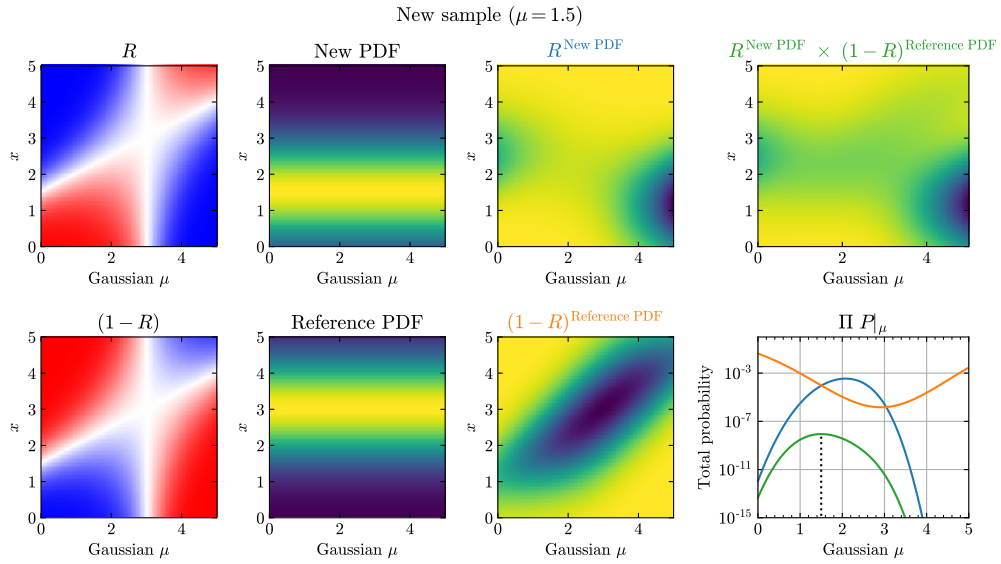


Figure F.3: The classifier is now applied to a new unknown sample and the reference sample. The upper left panel shows the classifier output R, which is the probability for a point in the space to belong to the scanned sample. Similarly, $(1 - \mathrm{R})$ is the probability to have come from the fixed sample, and is shown in the lower left panel. The probability density for the new set and the reference set are displayed in the second column, where neither depends on the model parameter μ. The third column displays the classifier output (R or 1-R) convolved with the probability distributions. The total probability is displayed in the top right panel; it is the product of the individual probabilities for each sample in the datasets. In the bottom right panel, the product of the individual probabilities along the y axis is shown as a function of the model parameter. The blue, orange, and green lines denote the probabilities for the new sample, the reference sample, and the combination, respectively. The green line is maximized at the value of the unknown parameter used to generate the new PDF.

The second line is just the usual binary cross-entropy loss function which is used to train binary classification neural networks.

Once the network is trained, we can infer the parameters of a new data, $\mathrm{x_i} \in \mathrm{x}_{\theta_\mathrm{T}}$ set by minimizing the loss of classifying the new data versus the reference set. Thus,

$$\hat{\theta} = \underset{\theta'}{\operatorname{argmin}} \left(- \sum_{\mathrm{x_i} \in \mathrm{x}_{\theta_\mathrm{T}}} \log \mathrm{f}'(\mathrm{x_i}, \theta') - \sum_{\mathrm{x_i} \in \mathrm{x}_{\theta_0}} \log \left(1 - \mathrm{f}(\mathrm{x_i}, \theta')\right) \right). \tag{F.9}$$

This is equivalent to maximizing the probability as we did in the Gaussian example.

F.2 Top Mass Training Curves

For completeness, we show training loss curves for DCTR and the linear network. As can be seen, training is stable and early stopping only becomes relevant once a plateau has been reached.

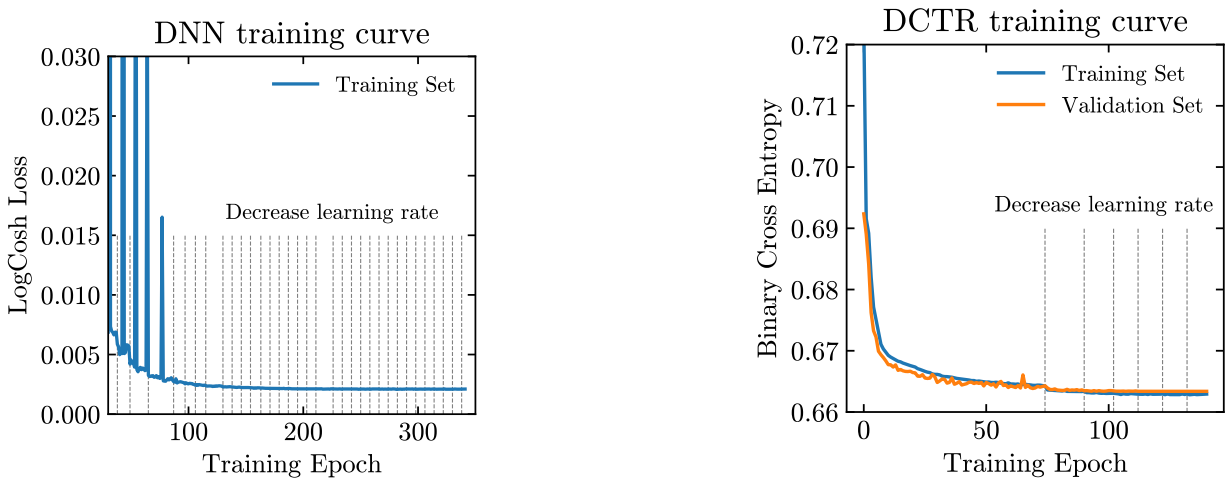


Figure F.4: (a): Training loss curve for the linear network. (b): Training loss curve for the DCTR classifier.

F.3 Variational Inference for Autoencoders

The idea behind variational inference for anomaly detection is to estimate the true probability distribution of the background, $\mathrm{p}(\mathbf{x})$. Assuming we have an underlying latent space of elements z, we can write $\mathrm{p}(\mathbf{x})$ as

$$\mathrm{p}\left(\mathbf{x}\right) = \mathbb{E}_{\mathrm{p}(\mathrm{z})}\left[\mathrm{p}\left(\mathbf{x}|\mathrm{z}\right)\right] \equiv \int \mathrm{p}\left(\mathbf{x}|\mathrm{z}\right)\ \mathrm{p}(\mathrm{z})\,\mathrm{dz}, \tag{F.10}$$

where $\mathbb{E}$ denotes the expectation value, $\mathrm{p}\left(\mathbf{x}|\mathrm{z}\right)$ is the probability of $\mathbf{x}$ given z, and $\mathrm{p}(\mathrm{z})$ is the prior likelihood of the latent data. We can take the latent space prior to be a set of independent Gaussians with zero mean and unit standard deviation, $\mathrm{z_i} \sim \mathcal{N}(0,1)$, where i runs over the dimension of the latent space. At this point $\mathrm{p}(\mathbf{x}|\mathrm{z})$ is an unknown and intractable distribution.

To make progress, we introduce a new *tractable* distribution $\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})$, where φ are some parameters to be optimized over. In an autoencoder architecture, this is the encoder. We can then write (F.10) in a more useful form:

$$\mathrm{p}(\mathbf{x}) = \int \mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})\,\frac{\mathrm{p}(\mathbf{x}|\mathrm{z})}{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\,\mathrm{p}(\mathrm{z})\mathrm{dz} = \mathbb{E}_{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\left[\frac{\mathrm{p}(\mathbf{x}|\mathrm{z})\mathrm{p}(\mathrm{z})}{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\right]. \tag{F.11}$$

The log likelihood, $\log \mathrm{p}(\mathbf{x})$, is then given as

$$\begin{aligned}
\log \mathrm{p}(\mathbf{x}) &= \log \mathbb{E}_{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\left[\frac{\mathrm{p}(\mathbf{x}|\mathrm{z})\mathrm{p}(\mathrm{z})}{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\right] \\
&\geq \mathbb{E}_{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\left[\log\left(\frac{\mathrm{p}(\mathbf{x}|\mathrm{z})\mathrm{p}(\mathrm{z})}{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\right)\right] \\
&= \mathbb{E}_{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}\left[\log \mathrm{p}(\mathbf{x}|\mathrm{z}) - \log\left(\frac{\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})}{\mathrm{p}(\mathrm{z})}\right)\right],
\end{aligned} \tag{F.12}$$

where we have used Jensen's inequality in the second line above. Let's first consider the first term in the last line of (F.12). It is the expectation value of $\mathbf{x}$ given z when z is sampled from $\mathrm{q}_\varphi(\mathrm{z}|\mathbf{x})$ (which is a distribution

in z given $\mathbf{x}$). This term can be interpreted as a (negative) reconstruction error term. If we approximate $p(\mathbf{x}|z)$ by a decoder part of the architecture $p_\theta(x|z)$ (where θ is to be optimized over), $\mathbb{E}_{q_\varphi(z|\mathbf{x})}(p_\theta(\mathbf{x}|z))$ is the usual (negative) reconstruction error term in the loss function for an autoencoder with decoder $p_\theta(\mathbf{x}|z)$ and encoder $q_\varphi(z|\mathbf{x})$.

The second term is by definition the Kullback-Leibler divergence (KLD) between the distributions $q_\varphi(z|\mathbf{x})$ and $p(z)$. Recall that $p(z) \sim \mathcal{N}(0,1)$. We take $q_\varphi(z|\mathbf{x})$ to also be a Gaussian distribution, but with a unknown mean and standard deviation (to be fixed by the optimization), i.e. $q_\varphi(z|x) = \mathcal{N}(\mu(x), \sigma^2(x))$. The KLD between these two distributions is then given exactly by (9.4). Using the reparameterization trick [1234,1252], we can write $q_\varphi(z|x)$ in terms of a standard normal:

$$z \sim q_\varphi(z|x),\ z = \mu(x) + \sigma(x)\varepsilon\ , \varepsilon \sim \mathcal{N}(0,1)\ . \tag{F.13}$$

Using the reparameterization trick allows for more efficient training of the network, as the back propagation of the gradients extends to the parameters of the distribution (μ and θ) even though a random draw from the distribution is passed to the decoder.

It's now clear that the last line in (F.12) is the negative loss for a VAE architecture. By training the VAE, we are minimizing the loss. By the inequality in (F.12), the last line is also a lower limit for the log likelihood. The optimized VAE therefore gives a *maximized* lower bound to the log likelihood, the so called Evidence LOwer Bound (ELBO). Notice that in this discussion it is imperative to use the full VAE loss in order for it to have the variational inference interpretation.

F.4 Comparing Anomaly Detection to a Supervised Method

It is well known that anomaly detection is sub-optimal for looking for any particular model; if the signal is known before-hand, supervised classification will yield the best results. We use a similar setup for our supervised classification as we did for the VAEs. The network consists of 1, 2, or 3 convolution blocks. Each block is made of two successive convolutional layers with 5 filters with a kernel size of 3 pixels, followed by an ELU activation function. After the convolutions, the data is down sampled with a 2×2 average pool operation. Following the convolution blocks, the data is flattened to a vector and a fully connected layer reduces the output to a single number with a sigmoid activation.

The networks are trained using 50000 events from the QCD sample and 50000 events from either the top or W samples. Similarly, 5000 events from each dataset are used for validation and to stop the network training when the validation loss has stopped improving. The training minimizes the binary cross entropy. After training, the network is applied to the test data of 5000 events in each class. We find that the network with three down sample layers achieves the best AUCs, with a score of 0.94 for top tagging and 0.96 for W tagging.

References

[1] K. Fraser and M. Reece, *Axion Periodicity and Coupling Quantization in the Presence of Mixing*, *JHEP* **05** (2020) 066, [arXiv:1910.11349].

[2] J. Fan, K. Fraser, M. Reece, and J. Stout, *Axion Mass from Magnetic Monopole Loops*, *Phys. Rev. Lett.* **127** (2021), no. 13 131602, [arXiv:2105.09950].

[3] H. Bagherian, K. Fraser, S. Homiller, and J. Stout, *Zero Modes of Massive Fermions Delocalize from Axion Strings*, arXiv:2310.01476.

[4] K. Fraser, A. Parikh, and W. L. Xu, *A Closer Look at CP-Violating Higgs Portal Dark Matter as a Candidate for the GCE*, *JHEP* **03** (2021) 123, [arXiv:2010.15129].

[5] P. Asadi, C. Cesarotti, K. Fraser, S. Homiller, and A. Parikh, *Oblique lessons from the W-mass measurement at CDF II*, *Phys. Rev. D* **108** (2023), no. 5 055026, [arXiv:2204.05283].

[6] P. Asadi, A. Bhattacharya, K. Fraser, S. Homiller, and A. Parikh, *Wrinkles in the Froggatt-Nielsen mechanism and flavorful new physics*, *JHEP* **10** (2023) 069, [arXiv:2308.01340].

[7] F. Flesher, K. Fraser, C. Hutchison, B. Ostdiek, and M. D. Schwartz, *Parameter inference from event ensembles and the top-quark mass*, *JHEP* **09** (2021) 058, [arXiv:2011.04666].

[8] K. Fraser, S. Homiller, R. K. Mishra, B. Ostdiek, and M. D. Schwartz, *Challenges for unsupervised anomaly detection in particle physics*, *JHEP* **03** (2022) 066, [arXiv:2110.06948].

[9] D. Tong, *Line Operators in the Standard Model*, *JHEP* **07** (2017) 104, [arXiv:1705.01853].

[10] B. W. Lee, C. Quigg, and H. B. Thacker, *Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass*, *Phys. Rev. D* **16** (1977) 1519.

[11] **CDF** Collaboration, T. Aaltonen et al., *High-precision measurement of the* W *boson mass with the CDF II detector*, *Science* **376** (2022), no. 6589 170–176.

[12] **BaBar** Collaboration, B. Aubert et al., *Observation of the semileptonic decays* $B \to D^* \tau^- \bar{\nu}(\tau)$ *and evidence for* $B \to D \tau^- \bar{\nu}(\tau)$, *Phys. Rev. Lett.* **100** (2008) 021801, [arXiv:0709.1698].

[13] **Belle** Collaboration, A. Bozek et al., *Observation of* $B^+ \to \bar{D}^* 0 \tau^+ \nu_\tau$ *and Evidence for* $B^+ \to \bar{D}^0 \tau^+ \nu_\tau$ *at Belle*, *Phys. Rev. D* **82** (2010) 072005, [arXiv:1005.2302].

[14] **BaBar** Collaboration, J. P. Lees et al., *Evidence for an excess of* $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ *decays*, *Phys. Rev. Lett.* **109** (2012) 101802, [arXiv:1205.5442].

[15] **BaBar** Collaboration, J. P. Lees et al., *Measurement of an Excess of* $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ *Decays and Implications for Charged Higgs Bosons*, *Phys. Rev. D* **88** (2013), no. 7 072012, [arXiv:1303.0571].

[16] **LHCb** Collaboration, R. Aaij et al., *Measurement of the ratio of branching fractions* $\mathcal{B}(\bar{B}^0 \to D^{*+} \tau^- \bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+} \mu^- \bar{\nu}_\mu)$, *Phys. Rev. Lett.* **115** (2015), no. 11 111803, [arXiv:1506.08614]. [Erratum: Phys.Rev.Lett. 115, 159901 (2015)].

[17] **Belle** Collaboration, M. Huschle et al., *Measurement of the branching ratio of* $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ *relative to* $\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$ *decays with hadronic tagging at Belle*, *Phys. Rev. D* **92** (2015), no. 7 072014, [arXiv:1507.03233].

[18] A. Abdesselam et al., *Measurement of the* τ *lepton polarization in the decay* $\bar{B} \to D^* \tau^- \bar{\nu}_\tau$, arXiv:1608.06391.

[19] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, *In the realm of the Hubble tension—a review of solutions*, *Class. Quant. Grav.* **38** (2021), no. 15 153001, [arXiv:2103.01183].

[20] **Planck** Collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [arXiv:1807.06209]. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[21] A. G. Riess et al., *A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^1 Mpc^1 Uncertainty from the Hubble Space Telescope and the SH0ES Team*, *Astrophys. J. Lett.* **934** (2022), no. 1 L7, [arXiv:2112.04510].

[22] A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn, and D. Scolnic, *Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM*, *Astrophys. J. Lett.* **908** (2021), no. 1 L6, [arXiv:2012.08534].

[23] A. G. Riess, *The Expansion of the Universe is Faster than Expected*, *Nature Rev. Phys.* **2** (2019), no. 1 10–12, [arXiv:2001.03624].

[24] A. H. Guth, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, *Phys. Rev. D* **23** (1981) 347–356.

[25] A. A. Starobinsky, *Spectrum of relict gravitational radiation and the early state of the universe*, *JETP Lett.* **30** (1979) 682–685.

[26] A. A. Starobinsky, *A New Type of Isotropic Cosmological Models Without Singularity*, *Phys. Lett. B* **91** (1980) 99–102.

[27] A. D. Linde, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, *Phys. Lett. B* **108** (1982) 389–393.

[28] A. Albrecht and P. J. Steinhardt, *Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking*, *Phys. Rev. Lett.* **48** (1982) 1220–1223.

[29] **TMVA** Collaboration, A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, physics/0703039.

[30] "Hep ml living review." https://iml-wg.github.io/HEPML-LivingReview/.

[31] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069, [arXiv:1511.05190].

[32] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, *Phys. Rev. D* **95** (2017), no. 1 014018, [arXiv:1609.00607].

[33] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev. D* **94** (2016), no. 11 112002, [arXiv:1607.08633].

[34] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110, [arXiv:1612.01551].

[35] G. Louppe, K. Cho, C. Becot, and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, *JHEP* **01** (2019) 057, [arXiv:1702.00748].

[36] C. Shimmin, P. Sadowski, P. Baldi, E. Weik, D. Whiteson, E. Goul, and A. Søgaard, *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, *Phys. Rev. D* **96** (2017), no. 7 074034, [arXiv:1703.03507].

[37] K. Fraser and M. D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093, [arXiv:1803.08066].

[38] **CMS** Collaboration, A. M. Sirunyan et al., *A deep neural network to search for new long-lived particles decaying to jets*, *Mach. Learn. Sci. Tech.* **1** (2020) 035012, [arXiv:1912.12238].

[39] A. Butter et al., *The Machine Learning landscape of top taggers*, *SciPost Phys.* **7** (2019) 014, [arXiv:1902.09914].

[40] T. Cai, J. Cheng, N. Craig, and K. Craig, *Linearized optimal transport for collider events*, *Phys. Rev. D* **102** (2020), no. 11 116019, [arXiv:2008.08604].

[41] **CMS** Collaboration, A. M. Sirunyan et al., *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, *JINST* **15** (2020), no. 06 P06005, [arXiv:2004.08262].

[42] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, *Binary JUNIPR: an interpretable probabilistic model for discrimination*, *Phys. Rev. Lett.* **123** (2019), no. 18 182001, [arXiv:1906.10137].

[43] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, *An efficient Lorentz equivariant graph neural network for jet tagging*, *JHEP* **07** (2022) 030, [arXiv:2201.08187].

[44] C. Li, H. Qu, S. Qian, Q. Meng, S. Gong, J. Zhang, T.-Y. Liu, and Q. Li, *Does Lorentz-symmetric design boost network performance in jet physics?*, *Phys. Rev. D* **109** (2024), no. 5 056003, [arXiv:2208.07814].

[45] H. Qu, C. Li, and S. Qian, *Particle Transformer for Jet Tagging*, arXiv:2202.03772.

[46] A. Bogatskiy, T. Hoffman, and J. T. Offermann, *19 Parameters Is All You Need: Tiny Neural Networks for Particle Physics*, in *37th Conference on Neural Information Processing Systems*, 10, 2023. arXiv:2310.16121.

[47] A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, and X. Liu, *Explainable equivariant neural networks for particle physics: PELICAN*, *JHEP* **03** (2024) 113, [arXiv:2307.16506].

[48] P. Baldi, P. Sadowski, and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, *Nature Commun.* **5** (2014) 4308, [arXiv:1402.4735].

[49] Y.-C. J. Chen, C.-W. Chiang, G. Cottin, and D. Shih, *Boosted* W *and* Z *tagging with jet charge and deep learning*, *Phys. Rev. D* **101** (2020), no. 5 053001, [arXiv:1908.08256].

[50] X. Ju and B. Nachman, *Supervised Jet Clustering with Graph Neural Networks for Lorentz Boosted Bosons*, *Phys. Rev. D* **102** (2020), no. 7 075014, [arXiv:2008.06064].

[51] F. A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052, [arXiv:2012.08526].

[52] T. Kim and A. Martin, *A* $\mathrm{W}^{\pm}$ *polarization analyzer from Deep Neural Networks*, arXiv:2102.05124.

[53] J. A. Aguilar-Saavedra, E. Arganda, F. R. Joaquim, R. M. Sandá Seoane, and J. F. Seabra, *Gradient Boosting MUST taggers for highly-boosted jets*, arXiv:2305.04957.

[54] D. Athanasakos, A. J. Larkoski, J. Mulligan,
M. Pℓosko, andF. Ringer, *Isinfrared* − *collinearsafeinformationallyouneedforjetclassification*?, arXiv : 2305.08979.

[55] K. Datta, A. Larkoski, and B. Nachman, *Automating the Construction of Jet Observables with Machine Learning*, *Phys. Rev. D* **100** (2019), no. 9 095016, [arXiv:1902.07180].

[56] E. A. Moreno, T. Q. Nguyen, J.-R. Vlimant, O. Cerri, H. B. Newman, A. Periwal, M. Spiropulu, J. M. Duarte, and M. Pierini, *Interaction networks for the identification of boosted* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decays*, *Phys. Rev. D* **102** (2020), no. 1 012010, [arXiv:1909.12285].

[57] A. Chakraborty, S. H. Lim, and M. M. Nojiri, *Interpretable deep learning for two-prong jet classification with jet spectra*, *JHEP* **07** (2019) 135, [arXiv:1904.02092].

[58] J. Guo, J. Li, T. Li, and R. Zhang, *Boosted Higgs boson jet reconstruction via a graph neural network*, *Phys. Rev. D* **103** (2021), no. 11 116025, [arXiv:2010.05464].

[59] C. K. Khosa and S. Marzani, *Higgs boson tagging with the Lund jet plane*, *Phys. Rev. D* **104** (2021), no. 5 055043, [arXiv:2105.03989].

[60] **ATLAS** Collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, tech. rep., CERN, Geneva, 2017. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-017.

[61] T. Cheng, *Recursive Neural Networks in Quark/Gluon Tagging*, *Comput. Softw. Big Sci.* **2** (2018), no. 1 3, [arXiv:1711.02633].

[62] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Periwal, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, *JEDI-net: a jet identification algorithm based on interaction networks*, *Eur. Phys. J. C* **80** (2020), no. 1 58, [arXiv:1908.05318].

[63] G. Kasieczka, N. Kiefer, T. Plehn, and J. M. Thompson, *Quark-Gluon Tagging: Machine Learning vs Detector*, *SciPost Phys.* **6** (2019), no. 6 069, [arXiv:1812.09223].

[64] G. Kasieczka, S. Marzani, G. Soyez, and G. Stagnitto, *Towards Machine Learning Analytics for Jet Substructure*, *JHEP* **09** (2020) 195, [arXiv:2007.04319].

[65] J. S. H. Lee, S. M. Lee, Y. Lee, I. Park, I. J. Watson, and S. Yang, *Quark Gluon Jet Discrimination with Weakly Supervised Learning*, *J. Korean Phys. Soc.* **75** (2019), no. 9 652–659, [arXiv:2012.02540].

[66] A. Romero, D. Whiteson, M. Fenton, J. Collado, and P. Baldi, *Safety of Quark/Gluon Jet Classification*, arXiv:2103.09103.

[67] F. A. Dreyer, G. Soyez, and A. Takacs, *Quarks and gluons in the Lund plane*, *JHEP* **08** (2022) 177, [arXiv:2112.09140].

[68] S. Bright-Thonney, I. Moult, B. Nachman, and S. Prestel, *Systematic quark/gluon identification with ratios of likelihoods*, *JHEP* **12** (2022) 021, [arXiv:2207.12411].

[69] S. Diefenbacher, H. Frost, G. Kasieczka, T. Plehn, and J. M. Thompson, *CapsNets Continuing the Convolutional Quest*, *SciPost Phys.* **8** (2020) 023, [arXiv:1906.11265].

[70] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006, [arXiv:1701.08784].

[71] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, *Deep-learned Top Tagging with a Lorentz Layer*, *SciPost Phys.* **5** (2018), no. 3 028, [arXiv:1707.08966].

[72] S. Macaluso and D. Shih, *Pulling Out All the Tops with Computer Vision and Deep Learning*, *JHEP* **10** (2018) 121, [arXiv:1803.00107].

[73] S. Bhattacharya, M. Guchait, and A. H. Vijay, *Boosted top quark tagging and polarization measurement using machine learning*, *Phys. Rev. D* **105** (2022), no. 4 042005, [arXiv:2010.11778].

[74] S. H. Lim and M. M. Nojiri, *Morphology for jet classification*, *Phys. Rev. D* **105** (2022), no. 1 014004, [arXiv:2010.13469].

[75] J. A. Aguilar-Saavedra, *Pulling the Higgs and top needles from the jet stack with feature extended supervised tagging*, *Eur. Phys. J. C* **81** (2021), no. 8 734, [arXiv:2102.01667].

[76] M. Andrews et al., *End-to-end jet classification of boosted top quarks with the CMS open data*, *EPJ Web Conf.* **251** (2021) 04030, [arXiv:2104.14659].

[77] S. K. Choi, J. Li, C. Zhang, and R. Zhang, *Automatic detection of boosted Higgs boson and top quark jets in an event image*, *Phys. Rev. D* **108** (2023), no. 11 116002, [arXiv:2302.13460].

[78] M. He and D. Wang, *Quark/gluon discrimination and top tagging with dual attention transformer*, *Eur. Phys. J. C* **83** (2023), no. 12 1116, [arXiv:2307.04723].

[79] B. Bhattacherjee, C. Bose, A. Chakraborty, and R. Sengupta, *Boosted top tagging and its interpretation using Shapley values*, arXiv:2212.11606.

[80] J. M. Munoz, I. Batatia, and C. Ortner, *Boost invariant polynomials for efficient jet tagging*, *Mach. Learn. Sci. Tech.* **3** (2022), no. 4 04LT05, [arXiv:2207.08272].

[81] I. Ahmed, A. Zada, M. Waqas, and M. U. Ashraf, *Application of deep learning in top pair and single top quark production at the LHC*, *Eur. Phys. J. Plus* **138** (2023), no. 9 795, [arXiv:2203.12871].

[82] W. Shen, D. Wang, and J. M. Yang, *Hierarchical high-point Energy Flow Network for jet tagging*, *JHEP* **09** (2023) 135, [arXiv:2308.08300].

[83] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086, [arXiv:1501.05968].

[84] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, *Neural Network-based Top Tagger with Two-Point Energy Correlations and Geometry of Soft Emissions*, *JHEP* **07** (2020) 111, [arXiv:2003.11787].

[85] F. A. Dreyer, R. Grabarczyk, and P. F. Monni, *Leveraging universality of jet taggers through transfer learning*, *Eur. Phys. J. C* **82** (2022), no. 6 564, [arXiv:2203.06210].

[86] R. Sahu and K. Ghosh, *ML-Based Top Taggers: Performance, Uncertainty and Impact of Tower & Tracker Data Integration*, arXiv:2309.01568.

[87] Y. Nakai, D. Shih, and S. Thomas, *Strange Jet Tagging*, arXiv:2003.09517.

[88] J. Erdmann, *A tagger for strange jets based on tracking information using long short-term memory*, *JINST* **15** (2020), no. 01 P01021, [arXiv:1907.07505].

[89] **ATLAS** Collaboration, *Identification of Jets Containing* b*-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, tech. rep., CERN, Geneva, 2017. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003.

[90] **ATLAS** Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, tech. rep., CERN, Geneva, 2020. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-014.

[91] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020), no. 5 056019, [arXiv:1902.08570].

[92] **ATLAS** Collaboration, G. Aad et al., *Fast b-tagging at the high-level trigger of the ATLAS experiment in LHC Run 3*, *JINST* **18** (2023), no. 11 P11006, [arXiv:2306.09738].

[93] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, *Jet Flavour Classification Using DeepJet*, *JINST* **15** (2020), no. 12 P12012, [arXiv:2008.10519].

[94] **CMS** Collaboration, A. M. Sirunyan et al., *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018), no. 05 P05011, [arXiv:1712.07158].

[95] J. Erdmann, O. Nackenhorst, and S. V. Zeißner, *Maximum performance of strange-jet tagging at hadron colliders*, *JINST* **16** (2021), no. 08 P08039, [arXiv:2011.10736].

[96] A. Furuichi, S. H. Lim, and M. M. Nojiri, *Jet Classification Using High-Level Features from Anatomy of Top Jets*, arXiv:2312.11760.

[97] R. Liu, A. Gandrakota, J. Ngadiuba, M. Spiropulu, and J.-R. Vlimant, *Efficient and Robust Jet Tagging at the LHC with Knowledge Distillation*, in *37th Conference on Neural Information Processing Systems*, 11, 2023. arXiv:2311.14160.

[98] E. M. Metodiev, B. Nachman, and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174, [arXiv:1708.02949].

[99] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, *Learning to classify from impure samples with high-dimensional data*, *Phys. Rev. D* **98** (2018), no. 1 011502, [arXiv:1801.10158].

[100] J. A. Aguilar-Saavedra, F. R. Joaquim, and J. F. Seabra, *Mass Unspecific Supervised Tagging (MUST) for boosted jets*, *JHEP* **03** (2021) 012, [arXiv:2008.12792]. [Erratum: JHEP 04, 133 (2021)].

[101] J. M. Clavijo, P. Glaysher, J. Jitsev, and J. M. Katzy, *Adversarial domain adaptation to reduce sample bias of a high energy physics event classifier **, *Mach. Learn. Sci. Tech.* **3** (2022), no. 1 015014, [arXiv:2005.00568].

[102] M. Abbas, A. Khan, A. S. Qureshi, and M. W. Khan, *Extracting Signals of Higgs Boson From Background Noise Using Deep Neural Networks*, arXiv:2010.08201.

[103] V. S. Ngairangbam and M. Spannowsky, *Interpretable deep learning models for the inference and classification of LHC data*, arXiv:2312.12330.

[104] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, *Casting a graph net to catch dark showers*, *SciPost Phys.* **10** (2021), no. 2 046, [arXiv:2006.08639].

[105] E. Arganda, A. D. Medina, A. D. Perez, and A. Szynkman, *Towards a method to anticipate dark matter signals with deep learning at the LHC*, *SciPost Phys.* **12** (2022), no. 2 063, [arXiv:2105.12018].

[106] X. C. Vidal, L. D. Maroñas, and A. D. Suárez, *How to Use Machine Learning to Improve the Discrimination between Signal and Background at Particle Colliders*, *Appl. Sciences* **11** (2021), no. 22 11076, [arXiv:2110.15099].

[107] B. Bhattacherjee, P. Konar, V. S. Ngairangbam, and P. Solanki, *LLPNet: Graph Autoencoder for Triggering Light Long-Lived Particles at HL-LHC*, arXiv:2308.13611.

[108] O. Amram and C. M. Suarez, *Tag N' Train: a technique to train improved classifiers on unlabeled data*, *JHEP* **01** (2021) 153, [arXiv:2002.12376].

[109] J. H. Collins, K. Howe, and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev. D* **99** (2019), no. 1 014038, [arXiv:1902.02634].

[110] J. H. Collins, K. Howe, and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, *Phys. Rev. Lett.* **121** (2018), no. 24 241803, [arXiv:1805.02664].

[111] R. T. D'Agnolo and A. Wulzer, *Learning New Physics from a Machine*, *Phys. Rev. D* **99** (2019), no. 1 015014, [arXiv:1806.02350].

[112] A. De Simone and T. Jacques, *Guiding New Physics Searches with Unsupervised Learning*, *Eur. Phys. J. C* **79** (2019), no. 4 289, [arXiv:1807.06038].

[113] A. Casa and G. Menardi, *Nonparametric semisupervised classification for signal detection in high energy physics*, arXiv:1809.02977.

[114] B. M. Dillon, D. A. Faroughy, and J. F. Kamenik, *Uncovering latent jet substructure*, *Phys. Rev. D* **100** (2019), no. 5 056002, [arXiv:1904.04200].

[115] A. Mullin, S. Nicholls, H. Pacey, M. Parker, M. White, and S. Williams, *Does SUSY have friends? A new approach for LHC event analysis*, *JHEP* **02** (2021) 160, [arXiv:1912.10625].

[116] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer, and M. Zanetti, *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021), no. 1 89, [arXiv:1912.12155].

[117] B. Nachman and D. Shih, *Anomaly Detection with Density Estimation*, *Phys. Rev. D* **101** (2020) 075042, [arXiv:2001.04990].

[118] A. Andreassen, B. Nachman, and D. Shih, *Simulation Assisted Likelihood-free Anomaly Detection*, *Phys. Rev. D* **101** (2020), no. 9 095004, [arXiv:2001.05001].

[119] B. M. Dillon, D. A. Faroughy, J. F. Kamenik, and M. Szewc, *Learning the latent structure of collider events*, *JHEP* **10** (2020) 206, [arXiv:2005.12319].

[120] K. Benkendorfer, L. L. Pottier, and B. Nachman, *Simulation-assisted decorrelation for resonant anomaly detection*, *Phys. Rev. D* **104** (2021), no. 3 035003, [arXiv:2009.02205].

[121] V. Mikuni and F. Canelli, *Unsupervised clustering for collider physics*, *Phys. Rev. D* **103** (2021), no. 9 092007, [arXiv:2010.07106].

[122] G. Stein, U. Seljak, and B. Dai, *Unsupervised in-distribution anomaly detection of new physics through conditional density estimation*, in *34th Conference on Neural Information Processing Systems*, 12, 2020. arXiv:2012.11638.

[123] G. Kasieczka et al., *The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics*, arXiv:2101.08320.

[124] J. Batson, C. G. Haaf, Y. Kahn, and D. A. Roberts, *Topological Obstructions to Autoencoding*, *JHEP* **04** (2021) 280, [arXiv:2102.08380].

[125] B. Bortolato, B. M. Dillon, J. F. Kamenik, and A. Smolkovi**v**c, *Bump Hunting in Latent Space*, arXiv:2103.06595.

[126] J. H. Collins, P. Martín-Ramiro, B. Nachman, and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, *Eur. Phys. J. C* **81** (2021), no. 7 617, [arXiv:2104.02092].

[127] S. Volkovich, F. De Vito Halevy, and S. Bressler, *The Data-Directed Paradigm for BSM searches*, arXiv:2107.11573.

[128] A. Hallin, J. Isaacson, G. Kasieczka, C. Krause, B. Nachman, T. Quadfasel, M. Schlaffer, D. Shih, and M. Sommerhalder, *Classifying Anomalies THrough Outer Density Estimation (CATHODE)*, arXiv:2109.00546.

[129] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163, [arXiv:1709.01087].

[130] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, *Novelty Detection Meets Collider Physics*, *Phys. Rev. D* **101** (2020), no. 7 076015, [arXiv:1807.10261].

[131] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Variational Autoencoders for New Physics Mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036, [arXiv:1811.10276].

[132] M. Farina, Y. Nakai, and D. Shih, *Searching for New Physics with Deep Autoencoders*, *Phys. Rev. D* **101** (2020), no. 7 075021, [arXiv:1808.08992].

[133] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019), no. 3 030, [arXiv:1808.08979].

[134] A. Blance, M. Spannowsky, and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047, [arXiv:1905.10384].

[135] M. Romão Crispim, N. F. Castro, R. Pedro, and T. Vale, *Transferability of Deep Learning Models in Searches for New Physics at Colliders*, *Phys. Rev. D* **101** (2020), no. 3 035042, [arXiv:1912.04220].

[136] T. S. Roy and A. H. Vijay, *A robust anomaly finder based on autoencoders*, arXiv:1903.02032.

[137] M. Crispim Romão, N. F. Castro, J. G. Milhano, R. Pedro, and T. Vale, *Use of a generalized energy Mover's distance in the search for rare phenomena at colliders*, *Eur. Phys. J. C* **81** (2021), no. 2 192, [arXiv:2004.09360].

[138] M. Crispim Romão, N. F. Castro, and R. Pedro, *Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders*, *Eur. Phys. J. C* **81** (2021), no. 1 27, [arXiv:2006.05432].

[139] D. A. Faroughy, *Uncovering hidden new physics patterns in collider events using Bayesian probabilistic models*, *PoS* **ICHEP2020** (2021) 238, [arXiv:2012.08579].

[140] C. K. Khosa and V. Sanz, *Anomaly Awareness*, arXiv:2007.14462.

[141] O. Knapp, O. Cerri, G. Dissertori, T. Q. Nguyen, M. Pierini, and J.-R. Vlimant, *Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark*, *Eur. Phys. J. Plus* **136** (2021), no. 2 236, [arXiv:2005.01598].

[142] S. E. Park, D. Rankin, S.-M. Udrescu, M. Yunus, and P. Harris, *Quasi Anomalous Knowledge: Searching for new physics with embedded knowledge*, *JHEP* **21** (2020) 030, [arXiv:2011.03550].

[143] A. A. Pol, V. Berger, G. Cerminara, C. Germain, and M. Pierini, *Anomaly Detection With Conditional Variational Autoencoders*, in *Eighteenth International Conference on Machine Learning and Applications*, 10, 2020. arXiv:2010.05531.

[144] M. van Beekveld, S. Caron, L. Hendriks, P. Jackson, A. Leinweber, S. Otten, R. Patrick, R. Ruiz De Austri, M. Santoni, and M. White, *Combining outlier analysis algorithms to identify new physics at the LHC*, *JHEP* **09** (2021) 024, [arXiv:2010.07940].

[145] E. Govorkova et al., *Autoencoders on FPGAs for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider*, arXiv:2108.03986.

[146] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette, and T. Golling, *Variational Autoencoders for Anomalous Jet Tagging*, arXiv:2007.01850.

[147] B. Ostdiek, *Deep Set Auto Encoders for Anomaly Detection in Particle Physics*, arXiv:2109.01695.

[148] B. M. Dillon, T. Plehn, C. Sauer, and P. Sorrenson, *Better Latent Spaces for Better Autoencoders*, arXiv:2104.08291.

[149] P. Thaprasop, K. Zhou, J. Steinheimer, and C. Herold, *Unsupervised Outlier Detection in Heavy-Ion Collisions*, *Phys. Scripta* **96** (2021), no. 6 064003, [arXiv:2007.15830].

[150] T. Aarrestad et al., *The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider*, arXiv:2105.14027.

[151] O. Atkinson, A. Bhardwaj, C. Englert, V. S. Ngairangbam, and M. Spannowsky, *Anomaly detection with Convolutional Graph Neural Networks*, arXiv:2105.07988.

[152] S. Caron, L. Hendriks, and R. Verheyen, *Rare and Different: Anomaly Scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC*, arXiv:2106.10164.

[153] P. Chakravarti, M. Kuusela, J. Lei, and L. Wasserman, *Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests*, arXiv:2102.07679.

[154] T. Finke, M. Krämer, A. Morandini, A. Mück, and I. Oleksiyuk, *Autoencoders for unsupervised anomaly detection in high energy physics*, *JHEP* **06** (2021) 161, [arXiv:2104.09051].

[155] J. Gonski, J. Lai, B. Nachman, and I. Ochoa, *High-dimensional Anomaly Detection with Radiative Return in* e^+e^- *Collisions*, arXiv:2108.13451.

[156] E. Govorkova, E. Puljak, T. Aarrestad, M. Pierini, K. A. Woźniak, and J. Ngadiuba, *LHC physics dataset for unsupervised New Physics detection at 40 MHz*, *Sci. Data* **9** (2022) 118, [arXiv:2107.02157].

[157] A. Kahn, J. Gonski, I. Ochoa, D. Williams, and G. Brooijmans, *Anomalous jet identification via sequence modeling*, *JINST* **16** (2021), no. 08 P08012, [arXiv:2105.09274].

[158] M. Birman, B. Nachman, R. Sebbah, G. Sela, O. Turetz, and S. Bressler, *Data-directed search for new physics based on symmetries of the SM*, *Eur. Phys. J. C* **82** (2022), no. 6 508, [arXiv:2203.07529].

[159] T. Buss, B. M. Dillon, T. Finke, M. Krämer, A. Morandini, A. Mück, I. Oleksiyuk, and T. Plehn, *What's anomalous in LHC jets?*, *SciPost Phys.* **15** (2023), no. 4 168, [arXiv:2202.00686].

[160] L. Bradshaw, S. Chang, and B. Ostdiek, *Creating simple, interpretable anomaly detectors for new physics in jet substructure*, *Phys. Rev. D* **106** (2022), no. 3 035014, [arXiv:2203.01343].

[161] S. Caron, R. R. de Austri, and Z. Zhang, *Mixture-of-Theories training: can we find new physics and anomalies better by mixing physical theories?*, *JHEP* **03** (2023) 004, [arXiv:2207.07631].

[162] B. M. Dillon, L. Favaro, T. Plehn, P. Sorrenson, and M. Krämer, *A normalized autoencoder for LHC triggers*, *SciPost Phys. Core* **6** (2023) 074, [arXiv:2206.14225].

[163] B. M. Dillon, R. Mastandrea, and B. Nachman, *Self-supervised anomaly detection for new physics*, *Phys. Rev. D* **106** (2022), no. 5 056005, [arXiv:2205.10380].

[164] C. Fanelli, J. Giroux, and Z. Papandreou, *'Flux+Mutability': a conditional generative approach to one-class classification and anomaly detection*, *Mach. Learn. Sci. Tech.* **3** (2022), no. 4 045012, [arXiv:2204.08609].

[165] T. Finke, M. Krämer, M. Lipp, and A. Mück, *Boosting mono-jet searches with model-agnostic machine learning*, *JHEP* **08** (2022) 015, [arXiv:2204.11889].

[166] T. Golling, S. Klein, R. Mastandrea, and B. Nachman, *Flow-enhanced transportation for anomaly detection*, *Phys. Rev. D* **107** (2023), no. 9 096025, [arXiv:2212.11285].

[167] A. Hallin, G. Kasieczka, T. Quadfasel, D. Shih, and M. Sommerhalder, *Resonant anomaly detection without background sculpting*, *Phys. Rev. D* **107** (2023), no. 11 114012, [arXiv:2210.14924].

[168] J. A. Aguilar-Saavedra, *Anomaly detection from mass unspecific jet tagging*, *Eur. Phys. J. C* **82** (2022), no. 2 130, [arXiv:2111.02647].

[169] F. Canelli, A. de Cosa, L. L. Pottier, J. Niedziela, K. Pedro, and M. Pierini, *Autoencoders for semivisible jet detection*, *JHEP* **02** (2022) 074, [arXiv:2112.02864].

[170] J. Herrero-Garcia, R. Patrick, and A. Scaffidi, *A semi-supervised approach to dark matter searches in direct detection data with machine learning*, *JCAP* **02** (2022), no. 02 039, [arXiv:2110.12248].

[171] B. Nachman, *Anomaly Detection for Physics Analysis and Less than Supervised Learning*, arXiv:2010.14554.

[172] G. Kasieczka, B. Nachman, and D. Shih, *New Methods and Datasets for Group Anomaly Detection From Fundamental Physics*, in *Conference on Knowledge Discovery and Data Mining*, 7, 2021. arXiv:2107.02821.

[173] V. Mikuni, B. Nachman, and D. Shih, *Online-compatible unsupervised nonresonant anomaly detection*, *Phys. Rev. D* **105** (2022), no. 5 055006, [arXiv:2111.06417].

[174] S. Chekanov and W. Hopkins, *Event-Based Anomaly Detection for Searches for New Physics*, *Universe* **8** (2022), no. 10 494, [arXiv:2111.12119].

[175] R. T. d'Agnolo, G. Grosso, M. Pierini, A. Wulzer, and M. Zanetti, *Learning new physics from an imperfect machine*, *Eur. Phys. J. C* **82** (2022), no. 3 275, [arXiv:2111.13633].

[176] J. F. Kamenik and M. Szewc, *Null hypothesis test for anomaly detection*, *Phys. Lett. B* **840** (2023) 137836, [arXiv:2210.02226].

[177] G. Kasieczka, R. Mastandrea, V. Mikuni, B. Nachman, M. Pettee, and D. Shih, *Anomaly detection under coordinate transformations*, *Phys. Rev. D* **107** (2023), no. 1 015009, [arXiv:2209.06225].

[178] M. Letizia, G. Losapio, M. Rando, G. Grosso, A. Wulzer, M. Pierini, M. Zanetti, and L. Rosasco, *Learning new physics efficiently with nonparametric methods*, *Eur. Phys. J. C* **82** (2022), no. 10 879, [arXiv:2204.02317].

[179] S. E. Park, P. Harris, and B. Ostdiek, *Neural embedding: learning the embedding of the manifold of physics data*, *JHEP* **07** (2023) 108, [arXiv:2208.05484].

[180] P. Jawahar, T. Aarrestad, N. Chernyavskaya, M. Pierini, K. A. Wozniak, J. Ngadiuba, J. Duarte, and S. Tsan, *Improving Variational Autoencoders for New Physics Detection at the LHC With Normalizing Flows*, *Front. Big Data* **5** (2022) 803685, [arXiv:2110.08508].

[181] J. A. Aguilar-Saavedra, *Taming modeling uncertainties with mass unspecific supervised tagging*, *Eur. Phys. J. C* **82** (2022), no. 3 270, [arXiv:2201.11143].

[182] J. A. Raine, S. Klein, D. Sengupta, and T. Golling, *CURTAINs for your sliding window: Constructing unobserved regions by transforming adjacent intervals*, *Front. Big Data* **6** (2023) 899345, [arXiv:2203.09470].

[183] R. Verheyen, *Event Generation and Density Estimation with Surjective Normalizing Flows*, *SciPost Phys.* **13** (2022), no. 3 047, [arXiv:2205.01697].

[184] K. Bai, R. Mastandrea, and B. Nachman, *Non-resonant anomaly detection with background extrapolation*, *JHEP* **04** (2024) 059, [arXiv:2311.12924].

[185] V. Belis, P. Odagiu, and T. K. Aarrestad, *Machine learning for anomaly detection in particle physics*, *Rev. Phys.* **12** (2024) 100091, [arXiv:2312.14190].

[186] G. Bickendorf, M. Drees, G. Kasieczka, C. Krause, and D. Shih, *Combining Resonant and Tail-based Anomaly Detection*, arXiv:2309.12918.

[187] E. Buhmann, C. Ewen, G. Kasieczka, V. Mikuni, B. Nachman, and D. Shih, *Full phase space resonant anomaly detection*, *Phys. Rev. D* **109** (2024), no. 5 055015, [arXiv:2310.06897].

[188] S. V. Chekanov and R. Zhang, *Enhancing the hunt for new phenomena in dijet final states using anomaly detection filters at the high-luminosity large Hadron Collider*, *Eur. Phys. J. Plus* **139** (2024), no. 3 237, [arXiv:2308.02671].

[189] R. Das, G. Kasieczka, and D. Shih, *Residual ANODE*, arXiv:2312.11629.

[190] B. M. Dillon, L. Favaro, F. Feiden, T. Modak, and T. Plehn, *Anomalies, Representations, and Self-Supervision*, arXiv:2301.04660.

[191] T. Finke, M. Hein, G. Kasieczka, M. Krämer, A. Mück, P. Prangchaikul, T. Quadfasel, D. Shih, and M. Sommerhalder, *Tree-based algorithms for weakly supervised anomaly detection*, *Phys. Rev. D* **109** (2024), no. 3 034033, [arXiv:2309.13111].

[192] M. Freytsis, M. Perelstein, and Y. C. San, *Anomaly detection in the presence of irrelevant features*, *JHEP* **02** (2024) 220, [arXiv:2310.13057].

[193] T. Golling et al., *The Mass-ive Issue: Anomaly Detection in Jet Physics*, in *34th Conference on Neural Information Processing Systems*, 3, 2023. arXiv:2303.14134.

[194] T. Golling, G. Kasieczka, C. Krause, R. Mastandrea, B. Nachman, J. A. Raine, D. Sengupta, D. Shih, and M. Sommerhalder, *The interplay of machine learning-based resonant anomaly detection methods*, *Eur. Phys. J. C* **84** (2024), no. 3 241, [arXiv:2307.11157].

[195] G. Grosso, N. Lai, M. Migliorini, J. Pazzini, A. Triossi, M. Zanetti, and A. Zucchetta, *Triggerless data acquisition pipeline for Machine Learning based statistical anomaly detection*, arXiv:2311.02038.

[196] C. Krause, B. Nachman, I. Pang, D. Shih, and Y. Zhu, *Anomaly detection with flow-based fast calorimeter simulators*, arXiv:2312.11618.

[197] R. Liu, A. Gandrakota, J. Ngadiuba, M. Spiropulu, and J.-R. Vlimant, *Fast Particle-based Anomaly Detection Algorithm with Variational Autoencoder*, in *37th Conference on Neural Information Processing Systems*, 11, 2023. arXiv:2311.17162.

[198] E. M. Metodiev, J. Thaler, and R. Wynne, *Anomaly Detection in Collider Physics via Factorized Observables*, arXiv:2312.00119.

[199] V. Mikuni and B. Nachman, *High-dimensional and Permutation Invariant Anomaly Detection*, *SciPost Phys.* **16** (2024) 062, [arXiv:2306.03933].

[200] S. Roche, Q. Bayer, B. Carlson, W. Ouligian, P. Serhiayenka, J. Stelzer, and T. M. Hong, *Nanosecond anomaly detection with decision trees and real-time application to exotic Higgs decays*, arXiv:2304.03836.

[201] D. Sengupta, M. Leigh, J. A. Raine, S. Klein, and T. Golling, *Improving new physics searches with diffusion models for event observables and jet constituents*, *JHEP* **04** (2024) 109, [arXiv:2312.10130].

[202] D. Sengupta, S. Klein, J. A. Raine, and T. Golling, *CURTAINs Flows For Flows: Constructing Unobserved Regions with Maximum Likelihood Estimation*, arXiv:2305.04646.

[203] L. Vaslin, V. Barra, and J. Donini, *GAN-AE: an anomaly detection algorithm for New Physics search in LHC data*, *Eur. Phys. J. C* **83** (2023), no. 11 1008, [arXiv:2305.15179].

[204] N. Craig, J. N. Howard, and H. Li, *Exploring Optimal Transport for Event-Level Anomaly Detection at the Large Hadron Collider*, arXiv:2401.15542.

[205] I. Oleksiyuk, J. A. Raine, M. Krämer, S. Voloshynovskiy, and T. Golling, *Cluster Scanning: a novel approach to resonance searches*, arXiv:2402.17714.

[206] **CMS ECAL** Collaboration, D. Abadjiev et al., *Autoencoder-based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter*, arXiv:2309.10157.

[207] **CMS** Collaboration, N. Zipper, *Testing a Neural Network for Anomaly Detection in the CMS Global Trigger Test Crate during Run 3*, *JINST* **19** (2024), no. 03 C03029, [arXiv:2312.10009].

[208] **ATLAS** Collaboration, *Simultaneous Jet Energy and Mass Calibrations with Neural Networks*, tech. rep., CERN, Geneva, 2020. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-001.

[209] **ATLAS** Collaboration, *Generalized Numerical Inversion: A Neural Network Approach to Jet Calibration*, tech. rep., CERN, Geneva, 2018. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-013.

[210] S. Cheong, A. Cukierman, B. Nachman, M. Safdari, and A. Schwartzman, *Parametrizing the Detector Response with Neural Networks*, *JINST* **15** (2020), no. 01 P01030, [arXiv:1910.03773].

[211] **CMS** Collaboration, A. M. Sirunyan et al., *A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution*, *Comput. Softw. Big Sci.* **4** (2020), no. 1 10, [arXiv:1912.06046].

[212] Y.-L. Du, D. Pablos, and K. Tywoniuk, *Deep learning jet modifications in heavy-ion collisions*, *JHEP* **21** (2020) 206, [arXiv:2012.07797].

[213] G. Kasieczka, M. Luchmann, F. Otterpohl, and T. Plehn, *Per-Object Systematics using Deep-Learned Calibration*, *SciPost Phys.* **9** (2020) 089, [arXiv:2003.11099].

[214] N. Akchurin, C. Cowden, J. Damgov, A. Hussain, and S. Kunori, *On the use of neural networks for energy reconstruction in high-granularity calorimeters*, *JINST* **16** (2021), no. 12 P12036, [arXiv:2107.10207].

[215] N. Akchurin, C. Cowden, J. Damgov, A. Hussain, and S. Kunori, *Perspectives on the Calibration of CNN Energy Reconstruction in Highly Granular Calorimeters*, arXiv:2108.10963.

[216] C. Pollard and P. Windischhofer, *Transport away your problems: Calibrating stochastic simulations with optimal transport*, *Nucl. Instrum. Meth. A* **1027** (2022) 166119, [arXiv:2107.08648].

[217] P. Ge, X. Huang, M. Saur, and L. Sun, *Improvement of q2 Resolution in Semileptonic Decays Based on Machine Learning*, *Adv. High Energy Phys.* **2023** (2023) 8127604, [arXiv:2208.02145].

[218] **CMS** Collaboration, V. Guglielmi, *Machine learning approaches for parameter reweighting for MC samples of top quark production in CMS*, *PoS* **ICHEP2022** (11, 2022) 1045, [arXiv:2211.07355].

[219] R. Mastandrea and B. Nachman, *Efficiently Moving Instead of Reweighting Collider Events with Machine Learning*, in *36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences*, 12, 2022. arXiv:2212.06155.

[220] **CMS** Collaboration, D. Valsecchi, *Deep learning techniques for energy clustering in the CMS ECAL*, *J. Phys. Conf. Ser.* **2438** (2023), no. 1 012077, [arXiv:2204.10277].

[221] G. Aad, T. Calvet, N. Chiedde, R. Faure, E. M. Fortin, L. Laatu, E. Monnier, and N. Sur, *Firmware implementation of a recurrent neural network for the computation of the energy deposited in the liquid argon calorimeter of the ATLAS experiment*, *JINST* **18** (2023), no. 05 P05017, [arXiv:2302.07555].

[222] **IceCube** Collaboration, J. Micallef, *Using convolutional neural networks to reconstruct energy of GeV scale IceCube neutrinos*, *JINST* **16** (2021), no. 09 C09019, [arXiv:2109.08152].

[223] M. Leigh, J. A. Raine, K. Zoch, and T. Golling, *ν-flows: Conditional neutrino regression*, *SciPost Phys.* **14** (2023), no. 6 159, [arXiv:2207.00664].

[224] M. Arratia, D. Britzger, O. Long, and B. Nachman, *Reconstructing the kinematics of deep inelastic scattering with deep learning*, *Nucl. Instrum. Meth. A* **1025** (2022) 166164, [arXiv:2110.05505].

[225] B. Kronheim, M. P. Kuchera, H. B. Prosper, and R. Ramanujan, *Implicit Quantile Neural Networks for Jet Simulation and Correction*, arXiv:2111.11415.

[226] D. F. Rentería-Estrada, R. J. Hernández-Pinto, G. F. R. Sborlini, and P. Zurita, *Reconstructing partonic kinematics at colliders with machine learning*, *SciPost Phys. Core* **5** (2022) 049, [arXiv:2112.05043].

[227] **CMS** Collaboration, J. Pata, J. Duarte, F. Mokhtar, E. Wulff, J. Yoo, J.-R. Vlimant, M. Pierini, and M. Girone, *Machine Learning for Particle Flow Reconstruction at CMS*, *J. Phys. Conf. Ser.* **2438** (2023), no. 1 012100, [arXiv:2203.00330].

[228] M. Chadeeva and S. Korpachev, *Machine-learning-based prediction of parameters of secondaries in hadronic showers using calorimetric observables*, *JINST* **17** (2022), no. 10 P10031, [arXiv:2205.12534].

[229] R. Gambhir, B. Nachman, and J. Thaler, *Learning Uncertainties the Frequentist Way: Calibration and Correlation in High Energy Physics*, *Phys. Rev. Lett.* **129** (2022), no. 8 082001, [arXiv:2205.03413].

[230] A. Alves and C. H. Yamaguchi, *Reconstruction of missing resonances combining nearest neighbors regressors and neural network classifiers*, *Eur. Phys. J. C* **82** (2022), no. 8 746, [arXiv:2203.03662].

[231] T. Dorigo, S. Guglielmini, J. Kieseler, L. Layer, and G. C. Strong, *Deep Regression of Muon Energy with a K-Nearest Neighbor Algorithm*, arXiv:2203.02841.

[232] G. Grosso, N. Lai, M. Letizia, J. Pazzini, M. Rando, L. Rosasco, A. Wulzer, and M. Zanetti, *Fast kernel methods for data quality monitoring as a goodness-of-fit test*, *Mach. Learn. Sci. Tech.* **4** (2023), no. 3 035029, [arXiv:2303.05413].

[233] R. Gambhir, B. Nachman, and J. Thaler, *Bias and priors in machine learning calibrations for high energy physics*, *Phys. Rev. D* **106** (2022), no. 3 036011, [arXiv:2205.05084].

[234] M. Kocot, K. Misan, V. Avati, E. Bossini, L. Grzanka, and N. Minafra, *Using deep neural networks to improve the precision of fast-sampled particle timing detectors*, arXiv:2312.05883.

[235] D. Holmberg, D. Golubovic, and H. Kirschenmann, *Jet Energy Calibration with Deep Learning as a Kubeflow Pipeline*, *Comput. Softw. Big Sci.* **7** (2023), no. 1 9, [arXiv:2308.12724].

[236] **ATLAS** Collaboration, G. Aad et al., *New techniques for jet calibration with the ATLAS detector*, *Eur. Phys. J. C* **83** (2023), no. 8 761, [arXiv:2303.17312].

[237] J. A. Raine, M. Leigh, K. Zoch, and T. Golling, *Fast and improved neutrino reconstruction in multineutrino final states with conditional normalizing flows*, *Phys. Rev. D* **109** (2024), no. 1 012005, [arXiv:2307.02405].

[238] J. Dolen, P. Harris, S. Marzani, S. Rappoccio, and N. Tran, *Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure*, *JHEP* **05** (2016) 156, [arXiv:1603.00027].

[239] A. Rogozhnikov, A. Bukva, V. V. Gligorov, A. Ustyuzhanin, and M. Williams, *New approaches for boosting to uniformity*, *JINST* **10** (2015), no. 03 T03002, [arXiv:1410.4140].

[240] S. Klein and T. Golling, *Decorrelation with conditional normalizing flows*, arXiv:2211.02486.

[241] M. Algren, J. A. Raine, and T. Golling, *Decorrelation using Optimal Transport*, arXiv:2307.05187.

[242] G. Kasieczka, B. Nachman, M. D. Schwartz, and D. Shih, *Automating the ABCD method with machine learning*, *Phys. Rev. D* **103** (2021), no. 3 035021, [arXiv:2007.14400].

[243] O. Kitouni, B. Nachman, C. Weisser, and M. Williams, *Enhancing searches for resonances with machine learning and moment decomposition*, *JHEP* **21** (2020) 070, [arXiv:2010.09745].

[244] A. Ghosh and B. Nachman, *A cautionary tale of decorrelating theory uncertainties*, *Eur. Phys. J. C* **82** (2022), no. 1 46, [arXiv:2109.08159].

[245] J. Stevens and M. Williams, *uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers*, *JINST* **8** (2013) P12013, [arXiv:1305.7248].

[246] L. Bradshaw, R. K. Mishra, A. Mitridate, and B. Ostdiek, *Mass Agnostic Jet Taggers*, *SciPost Phys.* **8** (2020), no. 1 011, [arXiv:1908.08959].

[247] **ATLAS** Collaboration, *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*, tech. rep., CERN, Geneva, 2018. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-014.

[248] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, *Phys. Rev. Lett.* **125** (2020), no. 12 122001, [arXiv:2001.05310].

[249] C. Englert, P. Galler, P. Harris, and M. Spannowsky, *Machine Learning Uncertainties with Adversarial Neural Networks*, *Eur. Phys. J. C* **79** (2019), no. 1 4, [arXiv:1807.08763].

[250] M. J. Dolan and A. Ore, *Metalearning and data augmentation for mass-generalized jet taggers*, *Phys. Rev. D* **105** (2022), no. 9 094030, [arXiv:2111.06047].

[251] A. Butter and T. Plehn, *Generative Networks for LHC events*, arXiv:2008.08558.

[252] Y. Alanazi, N. Sato, P. Ambrozewicz, A. N. H. Blin, W. Melnitchouk, M. Battaglieri, T. Liu, and Y. Li, *A survey of machine learning-based physics event generation*, arXiv:2106.00643.

[253] S. Badger et al., *Machine learning and LHC event generation*, *SciPost Phys.* **14** (2023), no. 4 079, [arXiv:2203.07460].

[254] A. Adelmann et al., *New directions for surrogate models and differentiable programming for High Energy Physics detector simulation*, in *Snowmass 2021*, 3, 2022. arXiv:2203.08806.

[255] D. Darulis, R. Tyson, D. G. Ireland, D. I. Glazier, B. McKinnon, and P. Pauli, *Machine Learned Particle Detector Simulations*, arXiv:2207.11254.

[256] **CMS** Collaboration, S. Bein, P. Connor, K. Pedro, P. Schleper, and M. Wolf, *Refining fast simulation using machine learning*, in *26th International Conference on Computing in High Energy & Nuclear Physics*, 9, 2023. arXiv:2309.12919.

[257] B. Hashemi and C. Krause, *Deep Generative Models for Detector Signature Simulation: An Analytical Taxonomy*, arXiv:2312.09597.

[258] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, *JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics*, *Eur. Phys. J. C* **79** (2019), no. 2 102, [arXiv:1804.09720].

[259] L. de Oliveira, M. Paganini, and B. Nachman, *Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis*, *Comput. Softw. Big Sci.* **1** (2017), no. 1 4, [arXiv:1701.05927].

[260] M. Paganini, L. de Oliveira, and B. Nachman, *Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters*, *Phys. Rev. Lett.* **120** (2018), no. 4 042003, [arXiv:1705.02355].

[261] M. Paganini, L. de Oliveira, and B. Nachman, *CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, *Phys. Rev. D* **97** (2018), no. 1 014021, [arXiv:1712.10321].

[262] S. Alonso-Monsalve and L. H. Whitehead, *Image-based model parameter optimization using Model-Assisted Generative Adversarial Networks*, *IEEE Trans. Neural Networks Learning Syst.* **31** (2020), no. 12 5645–5650, [arXiv:1812.00879].

[263] J. Arjona Martínez, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Particle Generative Adversarial Networks for full-event simulation at the LHC and their application to pileup description*, *J. Phys. Conf. Ser.* **1525** (2020), no. 1 012081, [arXiv:1912.02748].

[264] **SHiP** Collaboration, C. Ahdida et al., *Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks*, *JINST* **14** (2019) P11028, [arXiv:1909.04451].

[265] S. Carrazza and F. A. Dreyer, *Lund jet images from generative and cycle-consistent adversarial networks*, *Eur. Phys. J. C* **79** (2019), no. 11 979, [arXiv:1909.01359].

[266] A. Butter, T. Plehn, and R. Winterhalder, *How to GAN LHC Events*, *SciPost Phys.* **7** (2019), no. 6 075, [arXiv:1907.03764].

[267] E. Buhmann, C. Ewen, D. A. Faroughy, T. Golling, G. Kasieczka, M. Leigh, G. Quétant, J. A. Raine, D. Sengupta, and D. Shih, *EPiC-ly Fast Particle Cloud Generation with Flow-Matching and Diffusion*, arXiv:2310.00049.

[268] E. Buhmann, F. Gaede, G. Kasieczka, A. Korol, W. Korcari, K. Krüger, and P. McKeown, *CaloClouds II: ultra-fast geometry-independent highly-granular calorimeter simulation*, *JINST* **19** (2024), no. 04 P04020, [arXiv:2309.05704].

[269] V. Mikuni and B. Nachman, *CaloScore v2: single-shot calorimeter shower simulation with diffusion models*, *JINST* **19** (2024), no. 02 P02001, [arXiv:2308.03847].

[270] S. Diefenbacher, V. Mikuni, and B. Nachman, *Refining Fast Calorimeter Simulations with a Schrödinger Bridge*, arXiv:2308.12339.

[271] O. Amram and K. Pedro, *Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation*, *Phys. Rev. D* **108** (2023), no. 7 072014, [arXiv:2308.03876].

[272] Z. Imani, T. Wongjirad, and S. Aeron, *Score-based diffusion models for generating liquid argon time projection chamber images*, *Phys. Rev. D* **109** (2024), no. 7 072011, [arXiv:2307.13687].

[273] M. Leigh, D. Sengupta, J. A. Raine, G. Quétant, and T. Golling, *Faster diffusion model with improved quality for particle cloud generation*, *Phys. Rev. D* **109** (2024), no. 1 012010, [arXiv:2307.06836].

[274] F. T. Acosta, V. Mikuni, B. Nachman, M. Arratia, B. Karki, R. Milton, P. Karande, and A. Angerami, *Comparison of point cloud and image-based models for calorimeter fast simulation*, *JINST* **19** (2024), no. 05 P05003, [arXiv:2307.04780].

[275] A. Butter, N. Huetsch, S. Palacios Schweitzer, T. Plehn, P. Sorrenson, and J. Spinner, *Jet Diffusion versus JetGPT – Modern Networks for the LHC*, arXiv:2305.10475.

[276] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, W. Korcari, K. Krüger, and P. McKeown, *CaloClouds: fast geometry-independent highly-granular calorimeter simulation*, *JINST* **18** (2023), no. 11 P11025, [arXiv:2305.04847].

[277] M. Leigh, D. Sengupta, G. Quétant, J. A. Raine, K. Zoch, and T. Golling, *PC-JeDi: Diffusion for particle cloud generation in high energy physics*, *SciPost Phys.* **16** (2024), no. 1 018, [arXiv:2303.05376].

[278] V. Mikuni and B. Nachman, *Score-based generative models for calorimeter shower simulation*, *Phys. Rev. D* **106** (2022), no. 9 092009, [arXiv:2206.11898].

[279] A. Li, V. Krishnamohan, R. Kansal, R. Sen, S. Tsan, Z. Zhang, and J. Duarte, *Induced Generative Adversarial Particle Transformers*, in *37th Conference on Neural Information Processing Systems*, 12, 2023. arXiv:2312.04757.

[280] I. Pang, J. A. Raine, and D. Shih, *SuperCalo: Calorimeter shower super-resolution*, arXiv:2308.11700.

[281] M. R. Buckley, C. Krause, I. Pang, and D. Shih, *Inductive simulation of calorimeter showers with normalizing flows*, *Phys. Rev. D* **109** (2024), no. 3 033006, [arXiv:2305.11934].

[282] A. Xu, S. Han, X. Ju, and H. Wang, *Generative machine learning for detector response modeling with a conditional normalizing flow*, *JINST* **19** (2024), no. 02 P02003, [arXiv:2303.10148].

[283] B. Käch, D. Krücker, and I. Melzer-Pellmann, *Point Cloud Generation using Transformer Encoders and Normalising Flows*, arXiv:2211.13623.

[284] B. Käch, D. Krücker, I. Melzer-Pellmann, M. Scham, S. Schnake, and A. Verney-Provatas, *JetFlow: Generating Jets with Conditioned and Mass Constrained Normalising Flows*, arXiv:2211.13630.

[285] J. C. Cresswell, B. L. Ross, G. Loaiza-Ganem, H. Reyes-Gonzalez, M. Letizia, and A. L. Caterini, *CaloMan: Fast generation of calorimeter showers with density estimation on learned manifolds*, in *36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences*, 11, 2022. arXiv:2211.15380.

[286] C. Krause, I. Pang, and D. Shih, *CaloFlow for CaloChallenge Dataset 1*, arXiv:2210.14245.

[287] C. Krause and D. Shih, *Accelerating accurate simulations of calorimeter showers with normalizing flows and probability density distillation*, *Phys. Rev. D* **107** (2023), no. 11 113004, [arXiv:2110.11377].

[288] C. Krause and D. Shih, *Fast and accurate simulations of calorimeter showers with normalizing flows*, *Phys. Rev. D* **107** (2023), no. 11 113003, [arXiv:2106.05285].

[289] Y. Lu, J. Collado, D. Whiteson, and P. Baldi, *Sparse autoregressive models for scalable generation of sparse images in particle physics*, *Phys. Rev. D* **103** (2021), no. 3 036012, [arXiv:2009.14017].

[290] A. Abhishek, E. Drechsler, W. Fedorko, and B. Stelzer, *CaloDVAE : Discrete Variational Autoencoders for Fast Calorimeter Shower Simulation*, 10, 2022. arXiv:2210.07430.

[291] M. Touranakou, N. Chernyavskaya, J. Duarte, D. Gunopulos, R. Kansal, B. Orzari, M. Pierini, T. Tomei, and J.-R. Vlimant, *Particle-based fast jet simulation at the LHC with variational autoencoders*, *Mach. Learn. Sci. Tech.* **3** (2022), no. 3 035003, [arXiv:2203.00520].

[292] A. Hariri, D. Dyachkova, and S. Gleyzer, *Graph Generative Models for Fast Detector Simulations in High Energy Physics*, arXiv:2104.01725.

[293] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, and K. Krüger, *Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network*, *EPJ Web Conf.* **251** (2021) 03003, [arXiv:2102.12491].

[294] **ATLAS** Collaboration, *Deep generative models for fast shower simulation in ATLAS*, tech. rep., CERN, Geneva, 2018. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

[295] F. Rehm, S. Vallecorsa, K. Borras, and D. Krücker, *Physics Validation of Novel Convolutional 2D Architectures for Speeding Up High Energy Physics Simulations*, *EPJ Web Conf.* **251** (2021) 03042, [arXiv:2105.08960].

[296] F. Rehm, S. Vallecorsa, K. Borras, and D. Krücker, *Validation of Deep Convolutional Generative Adversarial Networks for High Energy Physics Calorimeter Simulations*, 3, 2021. arXiv:2103.13698.

[297] J. Chan, X. Ju, A. Kania, B. Nachman, V. Sangli, and A. Siodmok, *Fitting a deep generative hadronization model*, *JHEP* **09** (2023) 084, [arXiv:2305.17169].

[298] M. A. W. Scham, D. Krücker, and K. Borras, *DeepTreeGANv2: Iterative Pooling of Point Clouds*, arXiv:2312.00042.

[299] M. A. W. Scham, D. Krücker, B. Käch, and K. Borras, *DeepTreeGAN: Fast Generation of High Dimensional Point Clouds*, arXiv:2311.12616.

[300] E. Buhmann, G. Kasieczka, and J. Thaler, *EPiC-GAN: Equivariant point cloud generation for particle jets*, *SciPost Phys.* **15** (2023), no. 4 130, [arXiv:2301.08128].

[301] R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, *Particle Cloud Generation with Message Passing Generative Adversarial Networks*, in *35th Conference on Neural Information Processing Systems*, 6, 2021. arXiv:2106.11535.

[302] S. Choi and J. H. Lim, *A Data-driven Event Generator for Hadron Colliders using Wasserstein Generative Adversarial Network*, *J. Korean Phys. Soc.* **78** (2021), no. 6 482–489, [arXiv:2102.11524].

[303] M. Faucci Giannelli and R. Zhang, *CaloShowerGAN, a Generative Adversarial Networks model for fast calorimeter shower simulation*, arXiv:2309.06515.

[304] R. Kansal, J. Duarte, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, *Graph Generative Adversarial Networks for Sparse Data Generation in High Energy Physics*, in *34th Conference on Neural Information Processing Systems*, 11, 2020. arXiv:2012.00173.

[305] Y. Alanazi et al., *Machine learning-based event generator for electron-proton scattering*, *Phys. Rev. D* **106** (2022), no. 9 096002, [arXiv:2008.03151].

[306] E. Buhmann, S. Diefenbacher, E. Eren, F. Gaede, G. Kasieczka, A. Korol, and K. Krüger, *Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed*, *Comput. Softw. Big Sci.* **5** (2021), no. 1 13, [arXiv:2005.05334].

[307] L. de Oliveira, M. Paganini, and B. Nachman, *Controlling Physical Attributes in GAN-Accelerated Simulation of Electromagnetic Calorimeters*, *J. Phys. Conf. Ser.* **1085** (2018), no. 4 042017, [arXiv:1711.08813].

[308] M. Erdmann, L. Geiger, J. Glombitza, and D. Schmidt, *Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks*, *Comput. Softw. Big Sci.* **2** (2018), no. 1 4, [arXiv:1802.03325].

[309] P. Musella and F. Pandolfi, *Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks*, *Comput. Softw. Big Sci.* **2** (2018), no. 1 8, [arXiv:1805.00850].

[310] V. Chekalina, E. Orlova, F. Ratnikov, D. Ulyanov, A. Ustyuzhanin, and E. Zakharov, *Generative Models for Fast Calorimeter Simulation: the LHCb case*, *EPJ Web Conf.* **214** (2019) 02034, [arXiv:1812.01319].

[311] B. Hashemi, N. Amin, K. Datta, D. Olivito, and M. Pierini, *LHC analysis-specific datasets with Generative Adversarial Networks*, arXiv:1901.05282.

[312] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, and S. Palazzo, *DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC*, *JHEP* **08** (2019) 110, [arXiv:1903.02433].

[313] M. Erdmann, J. Glombitza, and T. Quast, *Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network*, *Comput. Softw. Big Sci.* **3** (2019), no. 1 4, [arXiv:1807.01954].

[314] A. Andreassen and B. Nachman, *Neural Networks for Full Phase-space Reweighting and Parameter Tuning*, *Phys. Rev. D* **101** (2020), no. 9 091901, [arXiv:1907.08209].

[315] L. Heinrich, S. Mishra-Sharma, C. Pollard, and P. Windischhofer, *Hierarchical Neural Simulation-Based Inference Over Event Ensembles*, arXiv:2306.12584.

[316] A. Morandini, T. Ferber, and F. Kahlhoefer, *Reconstructing axion-like particles from beam dumps with simulation-based inference*, *Eur. Phys. J. C* **84** (2024), no. 2 200, [arXiv:2308.01353].

[317] R. Barrué, P. Conde-Muíño, V. Dao, and R. Santos, *Simulation-based inference in the search for CP violation in leptonic WH production*, *JHEP* **04** (2024) 014, [arXiv:2308.02882].

[318] T. Heimel, N. Huetsch, R. Winterhalder, T. Plehn, and A. Butter, *Precision-Machine Learning for the Matrix Element Method*, arXiv:2310.07752.

[319] S. Chai, J. Gu, and L. Li, *From Optimal Observables to Machine Learning: an Effective-Field-Theory Analysis of* $e^+e^- \to W^+W^-$ *at Future Lepton Colliders*, arXiv:2401.02474.

[320] S. Chen, A. Glioti, G. Panico, and A. Wulzer, *Parametrized classifiers for optimal EFT sensitivity*, *JHEP* **05** (2021) 247, [arXiv:2007.10356].

[321] Y.-K. Lei, C. Liu, and Z. Chen, *Numerical analysis of neutrino physics within a high scale supersymmetry model via machine learning*, *Mod. Phys. Lett. A* **35** (2020), no. 26 2050218, [arXiv:2006.01495].

[322] M. Lazzarin, S. Alioli, and S. Carrazza, *MCNNTUNES: Tuning Shower Monte Carlo generators with machine learning*, *Comput. Phys. Commun.* **263** (2021) 107908, [arXiv:2010.02213].

[323] J. Alda, J. Guasch, and S. Penaranda, *Using Machine Learning techniques in phenomenological studies on flavour physics*, *JHEP* **07** (2022) 115, [arXiv:2109.07405].

[324] S. Qiu, S. Han, X. Ju, B. Nachman, and H. Wang, *Parton labeling without matching: unveiling emergent labelling capabilities in regression models*, *Eur. Phys. J. C* **83** (2023), no. 7 622, [arXiv:2304.09208].

[325] O. Al Hammal, M. Martini, J. Frontera-Pons, T. H. Nguyen, and R. Pérez-Ramos, *Neural network predictions of inclusive electron-nucleus cross sections*, *Phys. Rev. C* **107** (2023), no. 6 065501, [arXiv:2305.08217].

[326] K. Kong, K. T. Matchev, S. Mrenna, and P. Shyamsundar, *New Machine Learning Techniques for Simulation-Based Inference: InferoStatic Nets, Kernel Score Estimation, and Kernel Likelihood Ratio Estimation*, arXiv:2210.01680.

[327] H. Bahl and S. Brass, *Constraining $\mathcal{CP}$-violation in the Higgs-top-quark interaction using machine-learning-based inference*, *JHEP* **03** (2022) 017, [arXiv:2110.10177].

[328] R. K. Barman, D. Gonçalves, and F. Kling, *Machine learning the Higgs boson-top quark CP phase*, *Phys. Rev. D* **105** (2022), no. 3 035023, [arXiv:2110.07635].

[329] S. Chatterjee, N. Frohner, L. Lechner, R. Schöfbeck, and D. Schwarz, *Tree boosting for learning EFT parameters*, *Comput. Phys. Commun.* **277** (2022) 108385, [arXiv:2107.10859].

[330] B. Nachman and J. Thaler, *Learning from many collider events at once*, *Phys. Rev. D* **103** (2021), no. 11 116013, [arXiv:2101.07263].

[331] S. Bieringer, A. Butter, T. Heimel, S. Höche, U. Köthe, T. Plehn, and S. T. Radev, *Measuring QCD Splittings with Invertible Networks*, *SciPost Phys.* **10** (2021), no. 6 126, [arXiv:2012.09873].

[332] A. Andreassen, S.-C. Hsu, B. Nachman, N. Suaysom, and A. Suresh, *Parameter estimation using neural networks in the presence of detector effects*, *Phys. Rev. D* **103** (2021), no. 3 036001, [arXiv:2010.03569].

[333] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, arXiv:1506.02169.

[334] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *Mining gold from implicit models to improve likelihood-free inference*, *Proc. Nat. Acad. Sci.* **117** (2020), no. 10 5242–5249, [arXiv:1805.12244].

[335] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *A Guide to Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. D* **98** (2018), no. 5 052004, [arXiv:1805.00020].

[336] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. Lett.* **121** (2018), no. 11 111801, [arXiv:1805.00013].

[337] J. Hollingsworth and D. Whiteson, *Resonance Searches with Machine Learned Likelihood Ratios*, arXiv:2002.04699.

[338] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *Likelihood-free inference with an improved cross-entropy estimator*, arXiv:1808.00973.

[339] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, *Pileup Mitigation with Machine Learning (PUMML)*, *JHEP* **12** (2017) 051, [arXiv:1707.08600].

[340] **ATLAS** Collaboration, *Convolutional Neural Networks with Event Images for Pileup Mitigation with the ATLAS Detector*, tech. rep., CERN, Geneva, 2019. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-028.

[341] J. Arjona Martínez, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Pileup mitigation at the Large Hadron Collider with graph neural networks*, *Eur. Phys. J. Plus* **134** (2019), no. 7 333, [arXiv:1810.07988].

[342] B. Maier, S. M. Narayanan, G. de Castro, M. Goncharov, C. Paus, and M. Schott, *Pile-up mitigation using attention*, *Mach. Learn. Sci. Tech.* **3** (2022), no. 2 025012, [arXiv:2107.02779].

[343] T. Li, S. Liu, Y. Feng, G. Paspalaki, N. V. Tran, M. Liu, and P. Li, *Semi-supervised graph neural networks for pileup noise removal*, *Eur. Phys. J. C* **83** (2023), no. 1 99, [arXiv:2203.15823].

[344] **CRESST** Collaboration, G. Angloher et al., *Towards an automated data cleaning with deep learning in CRESST*, *Eur. Phys. J. Plus* **138** (2023), no. 1 100, [arXiv:2211.00564].

[345] K. Lieret, G. DeZoort, D. Chatterjee, J. Park, S. Miao, and P. Li, *High Pileup Particle Tracking with Object Condensation*, 12, 2023. arXiv:2312.03823.

[346] G. Louppe, M. Kagan, and K. Cranmer, *Learning to Pivot with Adversarial Networks*, arXiv:1611.01046.

[347] L.-G. Xia, *QBDT, a new boosting decision tree method with systematical uncertainties into training for High Energy Physics*, *Nucl. Instrum. Meth. A* **930** (2019) 15–26, [arXiv:1810.08387].

[348] S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn, and J. Thompson, *Deep-Learning Jets with Uncertainties and More*, *SciPost Phys.* **8** (2020), no. 1 006, [arXiv:1904.10004].

[349] B. Nachman, *A guide for deploying Deep Learning in LHC searches: How to achieve optimality and account for uncertainty*, *SciPost Phys.* **8** (2020) 090, [arXiv:1909.03081].

[350] B. Nachman and C. Shimmin, *AI Safety for High Energy Physics*, arXiv:1910.08606.

[351] S. Wunsch, S. Jörger, R. Wolf, and G. Quast, *Reducing the dependence of the neural network function to systematic uncertainties in the input space*, *Comput. Softw. Big Sci.* **4** (2020), no. 1 5, [arXiv:1907.11674].

[352] M. Bellagente, M. Haussmann, M. Luchmann, and T. Plehn, *Understanding Event-Generation Networks via Uncertainties*, *SciPost Phys.* **13** (2022), no. 1 003, [arXiv:2104.04543].

[353] T. Y. Chen, B. Dey, A. Ghosh, M. Kagan, B. Nord, and N. Ramachandra, *Interpretable Uncertainty Quantification in AI for HEP*, in *Snowmass 2021*, 8, 2022. arXiv:2208.03284.

[354] K. Cheung, Y.-L. Chung, S.-C. Hsu, and B. Nachman, *Exploring the universality of hadronic jet classification*, *Eur. Phys. J. C* **82** (2022), no. 12 1162, [arXiv:2204.03812].

[355] B. Viren, J. Huang, Y. Huang, M. Lin, Y. Ren, K. Terao, D. Torbunov, and H. Yu, *Solving Simulation Systematics in and with AI/ML*, in *Snowmass 2021*, 3, 2022. arXiv:2203.06112.

[356] A. Golutvin, A. Iniukhin, A. Mauri, P. Owen, N. Serra, and A. Ustyuzhanin, *The DL Advocate: playing the devil's advocate with hidden systematic uncertainties*, *Eur. Phys. J. C* **83** (2023), no. 9 779, [arXiv:2303.15956].

[357] A. Ghosh, B. Nachman, and D. Whiteson, *Uncertainty-aware machine learning for high energy physics*, *Phys. Rev. D* **104** (2021), no. 5 056026, [arXiv:2105.08742].

[358] D. Koh, A. Mishra, and K. Terao, *Deep neural network uncertainty quantification for LArTPC reconstruction*, *JINST* **18** (2023), no. 12 P12013, [arXiv:2302.03787].

[359] L. Layer, T. Dorigo, and G. Strong, *Application of Inferno to a Top Pair Cross Section Measurement with CMS Open Data*, arXiv:2301.10358.

[360] N. Simpson and L. Heinrich, *neos: End-to-End-Optimised Summary Statistics for High Energy Physics*, *J. Phys. Conf. Ser.* **2438** (2023), no. 1 012105, [arXiv:2203.05570].

[361] S. Wunsch, S. Jörger, R. Wolf, and G. Quast, *Optimal Statistical Inference in the Presence of Systematic Uncertainties Using Neural Network Optimization Based on Binned Poisson Likelihoods with Nuisance Parameters*, *Comput. Softw. Big Sci.* **5** (2021), no. 1 4, [arXiv:2003.07186].

[362] P. De Castro and T. Dorigo, *INFERNO: Inference-Aware Neural Optimisation*, *Comput. Phys. Commun.* **244** (2019) 170–179, [arXiv:1806.04743].

[363] A. Stein, X. Coubez, S. Mondal, A. Novak, and A. Schmidt, *Improving Robustness of Jet Tagging Algorithms with Adversarial Training*, *Comput. Softw. Big Sci.* **6** (2022), no. 1 15, [arXiv:2203.13890].

[364] J. Y. Araz and M. Spannowsky, *Combine and Conquer: Event Reconstruction with Bayesian Ensemble Neural Networks*, *JHEP* **04** (2021) 296, [arXiv:2102.01078].

[365] M.-L. Wong, A. Edmonds, and C. Wu, *Feed-forward neural network unfolding*, arXiv:2112.08180.

[366] M. Arratia, D. Britzger, O. Long, and B. Nachman, *Optimizing observables with machine learning for better unfolding*, *JINST* **17** (2022), no. 07 P07009, [arXiv:2203.16722].

[367] M. Backes, A. Butter, M. Dunford, and B. Malaescu, *An unfolding method based on conditional invertible neural networks (cINN) using iterative training*, *SciPost Phys. Core* **7** (2024), no. 1 007, [arXiv:2212.08674].

[368] J. Chan and B. Nachman, *Unbinned profiled unfolding*, *Phys. Rev. D* **108** (2023), no. 1 016002, [arXiv:2302.05390].

[369] A. Shmakov, K. Greif, M. Fenton, A. Ghosh, P. Baldi, and D. Whiteson, *End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics*, arXiv:2305.10399.

[370] P. Baron, *Comparison of Machine Learning Approach to Other Commonly Used Unfolding Methods*, *Acta Phys. Polon. B* **52** (2021), no. 8 863, [arXiv:2104.03036].

[371] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, A. Suresh, and J. Thaler, *Scaffolding Simulations with Deep Learning for High-dimensional Deconvolution*, in *9th International Conference on Learning Representations*, 5, 2021. arXiv:2105.04448.

[372] P. Komiske, W. P. McCormack, and B. Nachman, *Preserving new physics while simultaneously unfolding all observables*, *Phys. Rev. D* **104** (2021), no. 7 076027, [arXiv:2105.09923].

[373] **H1** Collaboration, V. Andreev et al., *Measurement of Lepton-Jet Correlation in Deep-Inelastic Scattering with the H1 Detector Using Machine Learning for Unfolding*, *Phys. Rev. Lett.* **128** (2022), no. 13 132002, [arXiv:2108.12376].

[374] M. Mieskolainen, *DeepEfficiency - optimal efficiency inversion in higher dimensions at the LHC*, arXiv:1809.06101.

[375] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, *OmniFold: A Method to Simultaneously Unfold All Observables*, *Phys. Rev. Lett.* **124** (2020), no. 18 182001, [arXiv:1911.09107].

[376] K. Datta, D. Kar, and D. Roy, *Unfolding with Generative Adversarial Networks*, arXiv:1806.00433.

[377] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, and R. Winterhalder, *How to GAN away Detector Effects*, *SciPost Phys.* **8** (2020), no. 4 070, [arXiv:1912.00477].

[378] N. D. Gagunashvili, *Machine learning approach to inverse problem and unfolding procedure*, arXiv:1004.2006.

[379] A. Glazov, *Machine learning as an instrument for data unfolding*, arXiv:1712.01814.

[380] M. Arratia et al., *Publishing unbinned differential cross section results*, *JINST* **17** (2022), no. 01 P01024, [arXiv:2109.13243].

[381] J. N. Howard, S. Mandt, D. Whiteson, and Y. Yang, *Learning to simulate high energy particle collisions from unlabeled data*, *Sci. Rep.* **12** (2022) 7567, [arXiv:2101.08944].

[382] M. Vandegar, M. Kagan, A. Wehenkel, and G. Louppe, *Neural Empirical Bayes: Source Distribution Estimation and its Applications to Simulation-Based Inference*, arXiv:2011.05836.

[383] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, A. Rousselot, R. Winterhalder, L. Ardizzone, and U. Köthe, *Invertible Networks or Partons to Detector and Back Again*, *SciPost Phys.* **9** (2020) 074, [arXiv:2006.06685].

[384] **ATLAS** Collaboration, G. Aad et al., *Dijet resonance search with weak supervision using* $\sqrt{s} = 13$ *TeV* pp *collisions in the ATLAS detector*, *Phys. Rev. Lett.* **125** (2020), no. 13 131801, [arXiv:2005.02983].

[385] **ATLAS** Collaboration, G. Aad et al., *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle* X *in hadronic final states using* $\sqrt{s} = 13$ *TeV* pp *collisions with the ATLAS detector*, *Phys. Rev. D* **108** (2023) 052009, [arXiv:2306.03637].

[386] **CMS** Collaboration, *Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, tech. rep., CERN, Geneva, 2024.

[387] **ATLAS** Collaboration, *Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, tech. rep., CERN, Geneva, 2023. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2023-022.

[388] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, *J. Phys. Conf. Ser.* **1085** (2018), no. 2 022008, [arXiv:1807.02876].

[389] D. Guest, K. Cranmer, and D. Whiteson, *Deep Learning and its Application to LHC Physics*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161–181, [arXiv:1806.11484].

[390] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (2018), no. 7716 41–48.

[391] J. Brehmer and K. Cranmer, *Simulation-based inference methods for particle physics*, arXiv:2010.06439.

[392] T. Dorigo and P. De Castro Manzano, *Dealing with Nuisance Parameters using Machine Learning in High Energy Physics: a Review*, arXiv:2007.09121.

[393] F. Psihas, M. Groh, C. Tunnell, and K. Warburton, *A Review on Machine Learning for Neutrino Experiments*, *Int. J. Mod. Phys. A* **35** (2020), no. 33 2043005, [arXiv:2008.01242].

[394] J. Shlomi, P. Battaglia, and J.-R. Vlimant, *Graph Neural Networks in Particle Physics*, arXiv:2007.13681.

[395] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, and D. Shih, *Machine Learning in the Search for New Fundamental Physics*, arXiv:2112.03769.

[396] A. Bogatskiy et al., *Symmetry Group Equivariant Architectures for Physics*, in *Snowmass 2021*, 3, 2022. arXiv:2203.06153.

[397] P. Shanahan et al., *Snowmass 2021 Computational Frontier CompF03 Topical Group Report: Machine Learning*, arXiv:2209.07559.

[398] S. Thais, P. Calafiura, G. Chachamis, G. DeZoort, J. Duarte, S. Ganguly, M. Kagan, D. Murnane, M. S. Neubauer, and K. Terao, *Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges*, in *Snowmass 2021*, 3, 2022. arXiv:2203.12852.

[399] G. DeZoort, P. W. Battaglia, C. Biscarat, and J.-R. Vlimant, *Graph neural networks at the Large Hadron Collider*, *Nature Rev. Phys.* **5** (2023), no. 5 281–303.

[400] T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause, and R. Winterhalder, *Modern Machine Learning for LHC Physicists*, arXiv:2211.01421.

[401] D. Baumann, *Inflation*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pp. 523–686, 2011. arXiv:0907.5424.

[402] V. Mukhanov, *Physical Foundations of Cosmology*. Cambridge University Press, Oxford, 2005.

[403] D. Green, *Cosmic Signals of Fundamental Physics*, *PoS* **TASI2022** (2024) 005, [arXiv:2212.08685].

[404] S. F. King, *Neutrino mass models*, *Rept. Prog. Phys.* **67** (2004) 107–158, [hep-ph/0310204].

[405] A. Riotto, *Theories of baryogenesis*, in *ICTP Summer School in High-Energy Physics and Cosmology*, pp. 326–436, 7, 1998. hep-ph/9807454.

[406] J. M. Cline, *Baryogenesis*, in *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime*, 9, 2006. hep-ph/0609145.

[407] **Particle Data Group** Collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev. D* **98** (2018), no. 3 030001.

[408] R. Bousso, *TASI Lectures on the Cosmological Constant*, *Gen. Rel. Grav.* **40** (2008) 607–637, [arXiv:0708.4231].

[409] N. Craig, *Naturalness: past, present, and future*, *Eur. Phys. J. C* **83** (2023), no. 9 825, [arXiv:2205.05708].

[410] J. Frieman, M. Turner, and D. Huterer, *Dark Energy and the Accelerating Universe*, *Ann. Rev. Astron. Astrophys.* **46** (2008) 385–432, [arXiv:0803.0982].

[411] S. Koren, *New Approaches to the Hierarchy Problem and their Signatures from Microscopic to Cosmic Scales*. PhD thesis, UC, Santa Barbara (main), 2020. arXiv:2009.11870.

[412] R. Barbieri and G. F. Giudice, *Upper Bounds on Supersymmetric Particle Masses*, *Nucl. Phys. B* **306** (1988) 63–76.

[413] G. W. Anderson and D. J. Castano, *Measures of fine tuning*, *Phys. Lett. B* **347** (1995) 300–308, [hep-ph/9409419].

[414] S. Fichet, *Quantified naturalness from Bayesian statistics*, *Phys. Rev. D* **86** (2012) 125029, [arXiv:1204.4940].

[415] P. Athron and D. J. Miller, *A New Measure of Fine Tuning*, *Phys. Rev. D* **76** (2007) 075010, [arXiv:0705.2241].

[416] S. P. Martin, *A Supersymmetry primer*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1–98, [hep-ph/9709356].

[417] F. Quevedo, S. Krippendorf, and O. Schlotterer, *Cambridge Lectures on Supersymmetry and Extra Dimensions*, arXiv:1011.1491.

[418] J. Terning, *Modern supersymmetry: Dynamics and duality*. 2006.

[419] D. Bertolini, J. Thaler, and Z. Thomas, *Super-Tricks for Superspace*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pp. 421–496, 2013. arXiv:1302.6229.

[420] K. A. Intriligator and N. Seiberg, *Lectures on Supersymmetry Breaking*, *Class. Quant. Grav.* **24** (2007) S741–S772, [hep-ph/0702069].

[421] M. A. Luty, *2004 TASI lectures on supersymmetry breaking*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in $D \geqq 4$*, pp. 495–582, 9, 2005. hep-th/0509029.

[422] S. R. Coleman and J. Mandula, *All Possible Symmetries of the S Matrix*, *Phys. Rev.* **159** (1967) 1251–1256.

[423] R. Haag, J. T. Lopuszanski, and M. Sohnius, *All Possible Generators of Supersymmetries of the s Matrix*, *Nucl. Phys. B* **88** (1975) 257.

[424] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, *Phys. Rev. D* **20** (1979) 2619–2625.

[425] E. Eichten, *IMPLICATIONS OF DYNAMICAL BREAKING OF WEAK INTERACTION SYMMETRIES*, in *Topical Workshop on Production of New Particles in Super High Energy Collisions*, 12, 1979.

[426] D. B. Kaplan and H. Georgi, *SU(2) x U(1) Breaking by Vacuum Misalignment*, *Phys. Lett. B* **136** (1984) 183–186.

[427] G. Panico and A. Wulzer, *The Composite Nambu-Goldstone Higgs*, vol. 913. Springer, 2016.

[428] M. Luty, "Composite higgs: Myth and reality."

[429] C. Csaki, T. Ma, and J. Shu, *Maximally Symmetric Composite Higgs Models*, *Phys. Rev. Lett.* **119** (2017), no. 13 131803, [arXiv:1702.00405].

[430] G. Durieux, M. McCullough, and E. Salvioni, *Gegenbauer Goldstones*, *JHEP* **01** (2022) 076, [arXiv:2110.06941].

[431] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263–272, [hep-ph/9803315].

[432] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, *New dimensions at a millimeter to a Fermi and superstrings at a TeV*, *Phys. Lett. B* **436** (1998) 257–263, [hep-ph/9804398].

[433] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370–3373, [hep-ph/9905221].

[434] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys. Rev. Lett.* **83** (1999) 4690–4693, [hep-th/9906064].

[435] M. Baryakhtar, *Graviton Phenomenology of Linear Dilaton Geometries*, *Phys. Rev. D* **85** (2012) 125019, [arXiv:1202.6674].

[436] G. F. Giudice and M. McCullough, *A Clockwork Theory*, *JHEP* **02** (2017) 036, [arXiv:1610.07962].

[437] G. F. Giudice, Y. Kats, M. McCullough, R. Torre, and A. Urbano, *Clockwork/linear dilaton: structure and phenomenology*, *JHEP* **06** (2018) 009, [arXiv:1711.08437].

[438] P. Agrawal, C. Cesarotti, A. Karch, R. K. Mishra, L. Randall, and R. Sundrum, *Warped Compactifications in Particle Physics, Cosmology and Quantum Gravity*, in *Snowmass 2021*, 3, 2022. arXiv:2203.07533.

[439] S. Weinberg, *The Cosmological Constant Problem*, *Rev. Mod. Phys.* **61** (1989) 1–23.

[440] V. Agrawal, S. M. Barr, J. F. Donoghue, and D. Seckel, *Viable range of the mass scale of the standard model*, *Phys. Rev. D* **57** (1998) 5480–5492, [hep-ph/9707380].

[441] R. Harnik, G. D. Kribs, and G. Perez, *A Universe without weak interactions*, *Phys. Rev. D* **74** (2006) 035006, [hep-ph/0604027].

[442] N. Craig, S. Knapen, and P. Longhi, *Neutral Naturalness from Orbifold Higgs Models*, *Phys. Rev. Lett.* **114** (2015), no. 6 061803, [arXiv:1410.6808].

[443] Z. Chacko, H.-S. Goh, and R. Harnik, *The Twin Higgs: Natural electroweak breaking from mirror symmetry*, *Phys. Rev. Lett.* **96** (2006) 231802, [hep-ph/0506256].

[444] P. W. Graham, D. E. Kaplan, and S. Rajendran, *Cosmological Relaxation of the Electroweak Scale*, *Phys. Rev. Lett.* **115** (2015), no. 22 221801, [arXiv:1504.07551].

[445] M. Geller, Y. Hochberg, and E. Kuflik, *Inflating to the Weak Scale*, *Phys. Rev. Lett.* **122** (2019), no. 19 191802, [arXiv:1809.07338].

[446] C. Cheung and P. Saraswat, *Mass Hierarchy and Vacuum Energy*, arXiv:1811.12390.

[447] R. Tito D'Agnolo and D. Teresi, *Sliding Naturalness: New Solution to the Strong-*CP *and Electroweak-Hierarchy Problems*, *Phys. Rev. Lett.* **128** (2022), no. 2 021803, [arXiv:2106.04591].

[448] C. Csáki, R. T. D'Agnolo, M. Geller, and A. Ismail, *Crunching Dilaton, Hidden Naturalness*, *Phys. Rev. Lett.* **126** (2021) 091801, [arXiv:2007.14396].

[449] N. Arkani-Hamed, T. Cohen, R. T. D'Agnolo, A. Hook, H. D. Kim, and D. Pinner, *Solving the Hierarchy Problem at Reheating with a Large Number of Degrees of Freedom*, *Phys. Rev. Lett.* **117** (2016), no. 25 251801, [arXiv:1607.06821].

[450] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The String landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[451] C. Cheung and G. N. Remmen, *Naturalness and the Weak Gravity Conjecture*, *Phys. Rev. Lett.* **113** (2014) 051601, [arXiv:1402.2287].

[452] N. Craig, I. Garcia Garcia, and S. Koren, *The Weak Scale from Weak Gravity*, *JHEP* **09** (2019) 081, [arXiv:1904.08426].

[453] L. E. Ibanez, V. Martin-Lozano, and I. Valenzuela, *Constraining the EW Hierarchy from the Weak Gravity Conjecture*, arXiv:1707.05811.

[454] J. March-Russell and R. Petrossian-Byrne, *QCD, Flavor, and the de Sitter Swampland*, arXiv:2006.01144.

[455] M. Montero, T. Van Riet, and G. Venken, *Festina Lente: EFT Constraints from Charged Black Hole Evaporation in de Sitter*, *JHEP* **01** (2020) 039, [arXiv:1910.01648].

[456] E. Palti, *The Swampland: Introduction and Review*, *Fortsch. Phys.* **67** (2019), no. 6 1900037, [arXiv:1903.06239].

[457] D. Harlow, B. Heidenreich, M. Reece, and T. Rudelius, *Weak gravity conjecture*, *Rev. Mod. Phys.* **95** (2023), no. 3 035003, [arXiv:2201.08380].

[458] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized Global Symmetries*, *JHEP* **02** (2015) 172, [arXiv:1412.5148].

[459] J. H. Oort, *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*, **6** (Aug., 1932) 249.

[460] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, *Helv. Phys. Acta* **6** (1933) 110–127.

[461] V. C. Rubin, *Dark matter in spiral galaxies*, *Scientific American* **248** (1983), no. 6 96–109.

[462] S. J. Penny, C. J. Conselice, S. De Rijcke, and E. V. Held, *Hubble Space Telescope survey of the Perseus Cluster – I. The structure and dark matter content of cluster dwarf spheroidals*, *Monthly Notices of the Royal Astronomical Society* **393** (02, 2009) 1054–1062, [https://academic.oup.com/mnras/article-pdf/393/3/1054/2907743/mnras0393-1054.pdf].

[463] L. V. E. Koopmans and T. Treu, *The structure and dynamics of luminous and dark matter in the early-type lens galaxy of 0047-281 at z=0.485*, *Astrophys. J.* **583** (2003) 606–615, [astro-ph/0205281].

[464] R. B. Metcalf, L. A. Moustakas, A. J. Bunker, and I. R. Parry, *Spectroscopic gravitational lensing and limits on the dark matter substructure in Q2237+0305*, *Astrophys. J.* **607** (2004) 43–59, [astro-ph/0309738].

[465] L. A. Moustakas and R. B. Metcalf, *Detecting dark matter substructure spectroscopically in strong gravitational lenses*, *Mon. Not. Roy. Astron. Soc.* **339** (2003) 607, [astro-ph/0206176].

[466] M. Bartelmann, *Gravitational Lensing*, *Class. Quant. Grav.* **27** (2010) 233001, [arXiv:1010.3829].

[467] H. Hoekstra, H. Yee, and M. Gladders, *Current status of weak gravitational lensing*, *New Astron. Rev.* **46** (2002) 767–781, [astro-ph/0205205].

[468] M. Brada**v**c, S. W. Allen, T. Treu, H. Ebeling, R. Massey, R. G. Morris, A. von der Linden, and D. Applegate, *Revealing the Properties of Dark Matter in the Merging Cluster MACS J0025.4-1222*, **687** (Nov., 2008) 959–967, [arXiv:0806.2320].

[469] M. J. Jee et al., *Discovery of a Ringlike Dark Matter Structure in the Core of the Galaxy Cluster Cl 0024+17*, *Astrophys. J.* **661** (2007) 728–749, [arXiv:0705.2171].

[470]

[471] A. G. Bergmann, V. Petrosian, and R. Lynds, *Gravitational Lens Models of Arcs in Clusters*, **350** (Feb., 1990) 23.

[472] J. A. Tyson, G. P. Kochanski, and I. P. Dell'Antonio, *Detailed mass map of cl 0024+1654 from strong lensing*, *The Astrophysical Journal* **498** (apr, 1998) L107.

[473] S. Chabanier, M. Millea, and N. Palanque-Delabrouille, *Matter power spectrum: from Lyα forest to CMB scales*, *Mon. Not. Roy. Astron. Soc.* **489** (2019), no. 2 2247–2253, [arXiv:1905.08103].

[474] K. Garrett and G. Duda, *Dark Matter: A Primer*, *Adv. Astron.* **2011** (2011) 968283, [arXiv:1006.2483].

[475] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [hep-ph/0404175].

[476] K. K. Boddy et al., *Snowmass2021 theory frontier white paper: Astrophysical and cosmological probes of dark matter*, *JHEAp* **35** (2022) 112–138, [arXiv:2203.06380].

[477] W. Hu, R. Barkana, and A. Gruzinov, *Cold and fuzzy dark matter*, *Phys. Rev. Lett.* **85** (2000) 1158–1161, [astro-ph/0003365].

[478] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, *Ultralight scalars as cosmological dark matter*, *Phys. Rev. D* **95** (2017), no. 4 043541, [arXiv:1610.08297].

[479] N. Dalal and A. Kravtsov, *Excluding fuzzy dark matter with sizes and stellar kinematics of ultrafaint dwarf galaxies*, *Phys. Rev. D* **106** (2022), no. 6 063517, [arXiv:2203.05750].

[480] K. K. Rogers and H. V. Peiris, *Strong Bound on Canonical Ultralight Axion Dark Matter from the Lyman-Alpha Forest*, *Phys. Rev. Lett.* **126** (2021), no. 7 071302, [arXiv:2007.12705].

[481] P. Meszaros, *Primeval black holes and galaxy formation*, *Astron. Astrophys.* **38** (1975) 5–13.

[482] Y. B. Zel'dovich and I. D. Novikov, *The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model*, *Sov. Astron.* **10** (1967) 602.

[483] P. Villanueva-Domingo, O. Mena, and S. Palomares-Ruiz, *A brief review on primordial black holes as dark matter*, *Front. Astron. Space Sci.* **8** (2021) 87, [arXiv:2103.12087].

[484] A. Escrivà, F. Kuhnel, and Y. Tada, *Primordial Black Holes*, arXiv:2211.05767.

[485] A. M. Green and B. J. Kavanagh, *Primordial Black Holes as a dark matter candidate*, *J. Phys. G* **48** (2021), no. 4 043001, [arXiv:2007.10722].

[486] H. Vogel and J. Redondo, *Dark Radiation constraints on minicharged particles in models with a hidden photon*, *JCAP* **02** (2014) 029, [arXiv:1311.2600].

[487] C. Dvorkin, T. Lin, and K. Schutz, *Making dark matter out of light: freeze-in from plasma effects*, *Phys. Rev. D* **99** (2019), no. 11 115009, [arXiv:1902.08623]. [Erratum: Phys.Rev.D 105, 119901 (2022)].

[488] T. Lin, *Dark matter models and direct detection*, *PoS* **333** (2019) 009, [arXiv:1904.07915].

[489] S. L. Dubovsky, D. S. Gorbunov, and G. I. Rubtsov, *Narrowing the window for millicharged particles by CMB anisotropy*, *JETP Lett.* **79** (2004) 1–5, [hep-ph/0311189].

[490] S. D. McDermott, H.-B. Yu, and K. M. Zurek, *Turning off the Lights: How Dark is Dark Matter?*, *Phys. Rev. D* **83** (2011) 063509, [arXiv:1011.2907].

[491] C. Dvorkin, K. Blum, and M. Kamionkowski, *Constraining Dark Matter-Baryon Scattering with Linear Cosmology*, *Phys. Rev. D* **89** (2014), no. 2 023519, [arXiv:1311.2937].

[492] W. L. Xu, C. Dvorkin, and A. Chael, *Probing sub-GeV Dark Matter-Baryon Scattering with Cosmological Observables*, *Phys. Rev. D* **97** (2018), no. 10 103530, [arXiv:1802.06788].

[493] T. R. Slatyer and C.-L. Wu, *Early-Universe constraints on dark matter-baryon scattering and their implications for a global 21 cm signal*, *Phys. Rev. D* **98** (2018), no. 2 023013, [arXiv:1803.09734].

[494] K. K. Boddy, V. Gluscevic, V. Poulin, E. D. Kovetz, M. Kamionkowski, and R. Barkana, *Critical assessment of CMB limits on dark matter-baryon scattering: New treatment of the relative bulk velocity*, *Phys. Rev. D* **98** (2018), no. 12 123506, [arXiv:1808.00001].

[495] A. Fung, S. Heeba, Q. Liu, V. Muralidharan, K. Schutz, and A. C. Vincent, *New bounds on light millicharged particles from the tip of the red-giant branch*, *Phys. Rev. D* **109** (2024), no. 8 083011, [arXiv:2309.06465].

[496] B. R. Safdi, *TASI Lectures on the Particle Physics and Astrophysics of Dark Matter*, *PoS* **TASI2022** (2024) 009, [arXiv:2303.02169].

[497] K. M. Zurek, *Dark Matter Candidates of a Very Low Mass*, arXiv:2401.03025.

[498] J. Silk et al., *Particle Dark Matter: Observations, Models and Searches*. Cambridge Univ. Press, Cambridge, 2010.

[499] M. Bauer and T. Plehn, *Yet Another Introduction to Dark Matter: The Particle Physics Approach*, vol. 959 of *Lecture Notes in Physics*. Springer, 2019.

[500] T. Marrodán Undagoitia and L. Rauch, *Dark matter direct-detection experiments*, *J. Phys. G* **43** (2016), no. 1 013001, [arXiv:1509.08767].

[501] M. Lisanti, *Lectures on Dark Matter Physics*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings*, pp. 399–446, 2017. arXiv:1603.03797.

[502] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre, S. Profumo, and F. S. Queiroz, *The waning of the WIMP? A review of models, searches, and constraints*, *Eur. Phys. J. C* **78** (2018), no. 3 203, [arXiv:1703.07364].

[503] A. Merle, *keV sterile neutrino Dark Matter*, *PoS* **NOW2016** (2017) 082, [arXiv:1702.08430].

[504] S. Hawking, *Gravitationally collapsed objects of very low mass*, **152** (Jan., 1971) 75.

[505] G. F. Chapline, *Cosmological effects of primordial black holes*, *Nature* **253** (1975), no. 5489 251–252.

[506] E. W. Kolb, D. J. H. Chung, and A. Riotto, *WIMPzillas!*, *AIP Conf. Proc.* **484** (1999), no. 1 91–105, [hep-ph/9810361].

[507] K. Griest and M. Kamionkowski, *Unitarity Limits on the Mass and Radius of Dark Matter Particles*, *Phys. Rev. Lett.* **64** (1990) 615.

[508] D. J. H. Chung, P. Crotty, E. W. Kolb, and A. Riotto, *On the Gravitational Production of Superheavy Dark Matter*, *Phys. Rev. D* **64** (2001) 043503, [hep-ph/0104100].

[509] Y. Bai, A. J. Long, and S. Lu, *Dark Quark Nuggets*, *Phys. Rev. D* **99** (2019), no. 5 055047, [arXiv:1810.04360].

[510] A. Kusenko and M. E. Shaposhnikov, *Supersymmetric Q balls as dark matter*, *Phys. Lett. B* **418** (1998) 46–54, [hep-ph/9709492].

[511] K. Petraki and R. R. Volkas, *Review of asymmetric dark matter*, *Int. J. Mod. Phys. A* **28** (2013) 1330028, [arXiv:1305.4939].

[512] D. E. Kaplan, M. A. Luty, and K. M. Zurek, *Asymmetric Dark Matter*, *Phys. Rev. D* **79** (2009) 115016, [arXiv:0901.4117].

[513] K. M. Zurek, *Asymmetric Dark Matter: Theories, Signatures, and Constraints*, *Phys. Rept.* **537** (2014) 91–121, [arXiv:1308.0338].

[514] M. L. Graesser, I. M. Shoemaker, and L. Vecchi, *Asymmetric WIMP dark matter*, *JHEP* **10** (2011) 110, [arXiv:1103.2771].

[515] T. Lin, H.-B. Yu, and K. M. Zurek, *On Symmetric and Asymmetric Light Dark Matter*, *Phys. Rev. D* **85** (2012) 063503, [arXiv:1111.0293].

[516] R. T. D'Agnolo and J. T. Ruderman, *Light Dark Matter from Forbidden Channels*, *Phys. Rev. Lett.* **115** (2015), no. 6 061301, [arXiv:1505.07107].

[517] K. Griest and D. Seckel, *Three exceptions in the calculation of relic abundances*, *Phys. Rev. D* **43** (1991) 3191–3203.

[518] R. T. D'Agnolo, D. Pappadopulo, J. T. Ruderman, and P.-J. Wang, *Thermal Relic Targets with Exponentially Small Couplings*, *Phys. Rev. Lett.* **124** (2020), no. 15 151801, [arXiv:1906.09269].

[519] F. D'Eramo and J. Thaler, *Semi-annihilation of Dark Matter*, *JHEP* **06** (2010) 109, [arXiv:1003.5912].

[520] R. T. D'Agnolo, D. Pappadopulo, and J. T. Ruderman, *Fourth Exception in the Calculation of Relic Abundances*, *Phys. Rev. Lett.* **119** (2017), no. 6 061102, [arXiv:1705.08450].

[521] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, *Freeze-In Production of FIMP Dark Matter*, *JHEP* **03** (2010) 080, [arXiv:0911.1120].

[522] N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen, and V. Vaskonen, *The Dawn of FIMP Dark Matter: A Review of Models and Constraints*, *Int. J. Mod. Phys. A* **32** (2017), no. 27 1730023, [arXiv:1706.07442].

[523] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, *Mechanism for Thermal Relic Dark Matter of Strongly Interacting Massive Particles*, *Phys. Rev. Lett.* **113** (2014) 171301, [arXiv:1402.5143].

[524] E. D. Carlson, M. E. Machacek, and L. J. Hall, *Self-interacting dark matter*, *Astrophys. J.* **398** (1992) 43–52.

[525] M. Farina, D. Pappadopulo, J. T. Ruderman, and G. Trevisan, *Phases of Cannibal Dark Matter*, *JHEP* **12** (2016) 039, [arXiv:1607.03108].

[526] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, *Elastically Decoupling Dark Matter*, *Phys. Rev. Lett.* **116** (2016), no. 22 221302, [arXiv:1512.04545].

[527] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, *Phenomenology of ELDER Dark Matter*, *JHEP* **08** (2017) 078, [arXiv:1706.05381].

[528] J. Alexander et al., *Dark Sectors 2016 Workshop: Community Report*, 8, 2016. arXiv:1608.08632.

[529] Y. Hochberg, "Tasi 2022 lectures on dark matter: Particle physics models." https://www.youtube.com/watch?v=Sq7JGnLcqWY, 2022.

[530] M. W. Goodman and E. Witten, *Detectability of Certain Dark Matter Candidates*, *Phys. Rev. D* **31** (1985) 3059.

[531] **XENON** Collaboration, E. Aprile et al., *First Dark Matter Search Results from the XENON1T Experiment*, *Phys. Rev. Lett.* **119** (2017), no. 18 181301, [arXiv:1705.06655].

[532] **XENON** Collaboration, E. Aprile et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, *Phys. Rev. Lett.* **121** (2018), no. 11 111302, [arXiv:1805.12562].

[533] **XENON** Collaboration, E. Aprile et al., *Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T*, *Phys. Rev. Lett.* **122** (2019), no. 14 141301, [arXiv:1902.03234].

[534] **XENON** Collaboration, E. Aprile et al., *First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment*, *Phys. Rev. Lett.* **131** (2023), no. 4 041003, [arXiv:2303.14729].

[535] **PICO Collaboration** Collaboration, C. Amole et al., *Dark matter search results from the* PICO−60 c_3f_8 *bubble chamber*, *Phys. Rev. Lett.* **118** (Jun, 2017) 251301.

[536] **PICO Collaboration** Collaboration, C. Amole et al., *Dark matter search results from the complete exposure of the pico-60* c_3f_8 *bubble chamber*, *Phys. Rev. D* **100** (Jul, 2019) 022001.

[537] **LZ** Collaboration, J. Aalbers et al., *First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment*, *Phys. Rev. Lett.* **131** (2023), no. 4 041002, [arXiv:2207.03764].

[538] Y. Hochberg, M. Pyle, Y. Zhao, and K. M. Zurek, *Detecting Superlight Dark Matter with Fermi-Degenerate Materials*, *JHEP* **08** (2016) 057, [arXiv:1512.04533].

[539] R. Essig, J. Mardon, and T. Volansky, *Direct Detection of Sub-GeV Dark Matter*, *Phys. Rev. D* **85** (2012) 076007, [arXiv:1108.5383].

[540] P. W. Graham, D. E. Kaplan, S. Rajendran, and M. T. Walters, *Semiconductor Probes of Light Dark Matter*, *Phys. Dark Univ.* **1** (2012) 32–49, [arXiv:1203.2531].

[541] R. Essig, M. Fernandez-Serra, J. Mardon, A. Soto, T. Volansky, and T.-T. Yu, *Direct Detection of sub-GeV Dark Matter with Semiconductor Targets*, *JHEP* **05** (2016) 046, [arXiv:1509.01598].

[542] Y. Hochberg, Y. Kahn, M. Lisanti, K. M. Zurek, A. G. Grushin, R. Ilan, S. M. Griffin, Z.-F. Liu, S. F. Weber, and J. B. Neaton, *Detection of sub-MeV Dark Matter with Three-Dimensional Dirac Materials*, *Phys. Rev. D* **97** (2018), no. 1 015004, [arXiv:1708.08929].

[543] R. Essig, J. Mardon, O. Slone, and T. Volansky, *Detection of sub-GeV Dark Matter and Solar Neutrinos via Chemical-Bond Breaking*, *Phys. Rev. D* **95** (2017), no. 5 056011, [arXiv:1608.02940].

[544] S. M. Griffin, K. Inzani, T. Trickle, Z. Zhang, and K. M. Zurek, *Multichannel direct detection of light dark matter: Target comparison*, *Phys. Rev. D* **101** (2020), no. 5 055004, [arXiv:1910.10716].

[545] T. Trickle, Z. Zhang, and K. M. Zurek, *Detecting Light Dark Matter with Magnons*, *Phys. Rev. Lett.* **124** (2020), no. 20 201801, [arXiv:1905.13744].

[546] C. Blanco, J. I. Collar, Y. Kahn, and B. Lillard, *Dark Matter-Electron Scattering from Aromatic Organic Targets*, *Phys. Rev. D* **101** (2020), no. 5 056001, [arXiv:1912.02822].

[547] Y. Hochberg, Y. Zhao, and K. M. Zurek, *Superconducting Detectors for Superlight Dark Matter*, *Phys. Rev. Lett.* **116** (2016), no. 1 011301, [arXiv:1504.07237].

[548] H.-Y. Chen, A. Mitridate, T. Trickle, Z. Zhang, M. Bernardi, and K. M. Zurek, *Dark matter direct detection in materials with spin-orbit coupling*, *Phys. Rev. D* **106** (2022), no. 1 015024, [arXiv:2202.11716].

[549] P. Du, D. Egaña Ugrinovic, R. Essig, and M. Sholapurkar, *Doped semiconductor devices for sub-MeV dark matter detection*, *Phys. Rev. D* **109** (2024), no. 5 055009, [arXiv:2212.04504].

[550] A. Hook, *TASI Lectures on the Strong CP Problem and Axions*, *PoS* **TASI2018** (2019) 004, [arXiv:1812.02669].

[551] M. Reece, *TASI Lectures: (No) Global Symmetries to Axion Physics*, *PoS* **TASI2022** (2024) 008, [arXiv:2304.08512].

[552] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, *Phys. Rev. Lett.* **124** (2020), no. 8 081803, [arXiv:2001.11966].

[553] J. R. Ellis and M. K. Gaillard, *Strong and Weak CP Violation*, *Nucl. Phys. B* **150** (1979) 141–162.

[554] K. Choi, C. W. Kim, and W. K. Sze, *Mass Renormalization by Instantons and the Strong CP Problem*, *Phys. Rev. Lett.* **61** (1988) 794.

[555] H. Georgi and I. N. McArthur, *INSTANTONS AND THE mu QUARK MASS*, .

[556] M. Srednicki, *Comment on "Ambiguities in the up-quark mass"*, *Phys. Rev. Lett.* **95** (2005) 059101, [hep-ph/0503051].

[557] T. Banks, Y. Nir, and N. Seiberg, *Missing (up) mass, accidental anomalous symmetries, and the strong CP problem*, in *2nd IFT Workshop on Yukawa Couplings and the Origins of Mass*, pp. 26–41, 2, 1994. hep-ph/9403203.

[558] C. Cordova, S. Hong, and S. Koren, *Non-Invertible Peccei-Quinn Symmetry and the Massless Quark Solution to the Strong CP Problem*, arXiv:2402.12453.

[559] D. B. Kaplan and A. V. Manohar, *Current Mass Ratios of the Light Quarks*, *Phys. Rev. Lett.* **56** (1986) 2004.

[560] **Flavour Lattice Averaging Group (FLAG)** Collaboration, Y. Aoki et al., *FLAG Review 2021*, *Eur. Phys. J. C* **82** (2022), no. 10 869, [arXiv:2111.09849].

[561] C. Alexandrou, J. Finkenrath, L. Funcke, K. Jansen, B. Kostrzewa, F. Pittler, and C. Urbach, *Ruling Out the Massless Up-Quark Solution to the Strong* **CP** *Problem by Computing the Topological Mass Contribution with Lattice QCD*, *Phys. Rev. Lett.* **125** (2020), no. 23 232001, [arXiv:2002.07802].

[562] D. Davies, M. Dine, and B. V. Lehmann, *Light Quarks at Large* N, arXiv:2201.05719.

[563] P. Agrawal and K. Howe, *A Flavorful Factoring of the Strong CP Problem*, *JHEP* **12** (2018) 035, [arXiv:1712.05803].

[564] A. Hook, *Anomalous solutions to the strong CP problem*, *Phys. Rev. Lett.* **114** (2015), no. 14 141801, [arXiv:1411.3325].

[565] A. E. Nelson, *Naturally Weak CP Violation*, *Phys. Lett. B* **136** (1984) 387–391.

[566] S. M. Barr, *Solving the Strong CP Problem Without the Peccei-Quinn Symmetry*, *Phys. Rev. Lett.* **53** (1984) 329.

[567] L. Bento, G. C. Branco, and P. A. Parada, *A Minimal model with natural suppression of strong CP violation*, *Phys. Lett. B* **267** (1991) 95–99.

[568] K. S. Babu and R. N. Mohapatra, *CP Violation in Seesaw Models of Quark Masses*, *Phys. Rev. Lett.* **62** (1989) 1079.

[569] K. S. Babu and R. N. Mohapatra, *A Solution to the Strong CP Problem Without an Axion*, *Phys. Rev. D* **41** (1990) 1286.

[570] S. M. Barr, D. Chang, and G. Senjanovic, *Strong CP problem and parity*, *Phys. Rev. Lett.* **67** (1991) 2765–2768.

[571] M. Dine and P. Draper, *Challenges for the Nelson-Barr Mechanism*, *JHEP* **08** (2015) 132, [arXiv:1506.05433].

[572] A. Albaid, M. Dine, and P. Draper, *Strong CP and SUZ_2*, *JHEP* **12** (2015) 046, [arXiv:1510.03392].

[573] J. de Vries, P. Draper, and H. H. Patel, *Do Minimal Parity Solutions to the Strong* CP *Problem Work?*, arXiv:2109.01630.

[574] R. N. Mohapatra and A. Rasin, *Simple supersymmetric solution to the strong CP problem*, *Phys. Rev. Lett.* **76** (1996) 3490–3493, [hep-ph/9511391].

[575] R. Kuchimanchi, *Solution to the strong CP problem: Supersymmetry with parity*, *Phys. Rev. Lett.* **76** (1996) 3486–3489, [hep-ph/9511376].

[576] N. Craig, I. Garcia Garcia, G. Koszegi, and A. McCune, *P not PQ*, *JHEP* **09** (2021) 130, [arXiv:2012.13416].

[577] G. Hiller and M. Schmaltz, *Solving the Strong CP Problem with Supersymmetry*, *Phys. Lett. B* **514** (2001) 263–268, [hep-ph/0105254].

[578] L. Vecchi, *Spontaneous CP violation and the strong CP problem*, *JHEP* **04** (2017) 149, [arXiv:1412.3805].

[579] D. Dunsky, L. J. Hall, and K. Harigaya, *Higgs Parity, Strong CP, and Dark Matter*, *JHEP* **07** (2019) 016, [arXiv:1902.07726].

[580] A. Valenti and L. Vecchi, *Super-soft CP violation*, *JHEP* **07** (2021), no. 152 152, [arXiv:2106.09108].

[581] R. N. Mohapatra and A. Rasin, *A Supersymmetric solution to CP problems*, *Phys. Rev. D* **54** (1996) 5835–5844, [hep-ph/9604445].

[582] A. Valenti and L. Vecchi, *The CKM phase and $\overline{\theta}$ in Nelson-Barr models*, *JHEP* **07** (2021), no. 203 203, [arXiv:2105.09122].

[583] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev.* **D16** (1977) 1791–1797.

[584] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[585] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223–226.

[586] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[587] J. E. Kim and G. Carosi, *Axions and the Strong CP Problem*, *Rev. Mod. Phys.* **82** (2010) 557–602, [arXiv:0807.3125].

[588] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, *Experimental Searches for the Axion and Axion-Like Particles*, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 485–514, [arXiv:1602.00039].

[589] P. Agrawal, K. V. Berghaus, J. Fan, A. Hook, G. Marques-Tavares, and T. Rudelius, *Some open questions in axion theory*, in *Snowmass 2021*, 3, 2022. arXiv:2203.08026.

[590] D. J. E. Marsh, *Axion Cosmology*, *Phys. Rept.* **643** (2016) 1–79, [arXiv:1510.07633].

[591] C. A. J. O'Hare, *Cosmology of axion dark matter*, *PoS* **COSMICWISPers** (2024) 040, [arXiv:2403.17697].

[592] J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43** (1979) 103.

[593] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys.* **B166** (1980) 493–506.

[594] A. R. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*, *Sov. J. Nucl. Phys.* **31** (1980) 260. [Yad. Fiz.31,497(1980)].

[595] M. Dine, W. Fischler, and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett.* **104B** (1981) 199–202.

[596] E. Witten, *Some Properties of O(32) Superstrings*, *Phys. Lett. B* **149** (1984) 351–356.

[597] K. Choi and J. E. Kim, *Harmful Axions in Superstring Models*, *Phys. Lett. B* **154** (1985) 393. [Erratum: Phys.Lett.B 156, 452 (1985)].

[598] S. M. Barr, *Harmless Axions in Superstring Theories*, *Phys. Lett. B* **158** (1985) 397–400.

[599] S. M. Barr and D. Seckel, *Planck scale corrections to axion models*, *Phys. Rev. D* **46** (1992) 539–549.

[600] M. Kamionkowski and J. March-Russell, *Planck scale physics and the Peccei-Quinn mechanism*, *Phys. Lett. B* **282** (1992) 137–141, [hep-th/9202003].

[601] R. Holman, S. D. H. Hsu, T. W. Kephart, E. W. Kolb, R. Watkins, and L. M. Widrow, *Solutions to the strong CP problem in a world with gravity*, *Phys. Lett. B* **282** (1992) 132–136, [hep-ph/9203206].

[602] T. Banks and N. Seiberg, *Symmetries and Strings in Field Theory and Gravity*, *Phys. Rev. D* **83** (2011) 084019, [arXiv:1011.5120].

[603] D. Harlow and H. Ooguri, *Constraints on Symmetries from Holography*, *Phys. Rev. Lett.* **122** (2019), no. 19 191601, [arXiv:1810.05337].

[604] D. Harlow and H. Ooguri, *Symmetries in quantum field theory and quantum gravity*, *Commun. Math. Phys.* **383** (2021), no. 3 1669–1804, [arXiv:1810.05338].

[605] L. Randall, *Composite axion models and Planck scale physics*, *Phys. Lett. B* **284** (1992) 77–80.

[606] K. S. Babu, I. Gogoladze, and K. Wang, *Stabilizing the axion by discrete gauge symmetries*, *Phys. Lett. B* **560** (2003) 214–222, [hep-ph/0212339].

[607] J. Preskill, M. B. Wise, and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett.* **120B** (1983) 127–132.

[608] M. Dine and W. Fischler, *The Not So Harmless Axion*, *Phys. Lett.* **120B** (1983) 137–141.

[609] L. F. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, *Phys. Lett.* **120B** (1983) 133–136.

[610] R. T. Co, L. J. Hall, and K. Harigaya, *Axion Kinetic Misalignment Mechanism*, *Phys. Rev. Lett.* **124** (2020), no. 25 251802, [arXiv:1910.14152].

[611] C.-F. Chang and Y. Cui, *New Perspectives on Axion Misalignment Mechanism*, *Phys. Rev. D* **102** (2020), no. 1 015003, [arXiv:1911.11885].

[612] J. Lee, K. Murai, F. Takahashi, and W. Yin, *Bubble Misalignment Mechanism for Axions*, arXiv:2402.09501.

[613] L. Di Luzio, B. Gavela, P. Quilez, and A. Ringwald, *Dark matter from an even lighter QCD axion: trapped misalignment*, *JCAP* **10** (2021) 001, [arXiv:2102.01082].

[614] S. Nakagawa, F. Takahashi, and M. Yamada, *Trapping Effect for QCD Axion Dark Matter*, *JCAP* **05** (2021) 062, [arXiv:2012.13592].

[615] K. S. Jeong, K. Matsukawa, S. Nakagawa, and F. Takahashi, *Cosmological effects of Peccei-Quinn symmetry breaking on QCD axion dark matter*, *JCAP* **03** (2022), no. 03 026, [arXiv:2201.00681].

[616] H.-J. Li, Y.-Q. Peng, W. Chao, and Y.-F. Zhou, *Light QCD axion dark matter from double level crossings*, *Phys. Lett. B* **849** (2024) 138444, [arXiv:2310.02126].

[617] N. Blinov, M. J. Dolan, P. Draper, and J. Kozaczuk, *Dark matter targets for axionlike particle searches*, *Phys. Rev. D* **100** (2019), no. 1 015049, [arXiv:1905.06952].

[618] A. Papageorgiou, P. Quílez, and K. Schmitz, *Axion dark matter from frictional misalignment*, *JHEP* **01** (2023) 169, [arXiv:2206.01129].

[619] R. T. Co, L. J. Hall, and K. Harigaya, *QCD Axion Dark Matter with a Small Decay Constant*, *Phys. Rev. Lett.* **120** (2018), no. 21 211602, [arXiv:1711.10486].

[620] L. D. McLerran, E. Mottola, and M. E. Shaposhnikov, *Sphalerons and Axion Dynamics in High Temperature QCD*, *Phys. Rev. D* **43** (1991) 2027–2035.

[621] R. T. Co and K. Harigaya, *Axiogenesis*, *Phys. Rev. Lett.* **124** (2020), no. 11 111602, [arXiv:1910.02080].

[622] T. W. B. Kibble, *Topology of Cosmic Domains and Strings*, *J. Phys. A* **9** (1976) 1387–1398.

[623] W. H. Zurek, *Cosmological Experiments in Superfluid Helium?*, *Nature* **317** (1985) 505–508.

[624] M. Buschmann, J. W. Foster, A. Hook, A. Peterson, D. E. Willcox, W. Zhang, and B. R. Safdi, *Dark matter from axion strings with adaptive mesh refinement*, *Nature Commun.* **13** (2022), no. 1 1049, [arXiv:2108.05368].

[625] K. Freese, J. A. Frieman, and A. V. Olinto, *Natural inflation with pseudo - Nambu-Goldstone bosons*, *Phys. Rev. Lett.* **65** (1990) 3233–3236.

[626] M. S. Turner and L. M. Widrow, *Inflation Produced, Large Scale Magnetic Fields*, *Phys. Rev.* **D37** (1988) 2743.

[627] W. D. Garretson, G. B. Field, and S. M. Carroll, *Primordial magnetic fields from pseudoGoldstone bosons*, *Phys. Rev.* **D46** (1992) 5346–5351, [hep-ph/9209238].

[628] S. H.-S. Alexander, M. E. Peskin, and M. M. Sheikh-Jabbari, *Leptogenesis from gravity waves in models of inflation*, *Phys. Rev. Lett.* **96** (2006) 081301, [hep-th/0403069].

[629] R. T. Co, N. Fernandez, A. Ghalsasi, L. J. Hall, and K. Harigaya, *Lepto-Axiogenesis*, *JHEP* **03** (2021) 017, [arXiv:2006.05687].

[630] P. Barnes, R. T. Co, K. Harigaya, and A. Pierce, *Lepto-axiogenesis with light right-handed neutrinos*, arXiv:2402.10263.

[631] G. Ballesteros, J. Redondo, A. Ringwald, and C. Tamarit, *Unifying inflation with the axion, dark matter, baryogenesis and the seesaw mechanism*, *Phys. Rev. Lett.* **118** (2017), no. 7 071802, [arXiv:1608.05414].

[632] A. H. Sopov and R. R. Volkas, *VISHν: solving five Standard Model shortcomings with a Poincaré-protected electroweak scale*, *Phys. Dark Univ.* **42** (2023) 101381, [arXiv:2206.11598].

[633] M. Dine and N. Seiberg, *String Theory and the Strong CP Problem*, *Nucl. Phys.* **B273** (1986) 109–124.

[634] P. Svrcek and E. Witten, *Axions In String Theory*, *JHEP* **06** (2006) 051, [hep-th/0605206].

[635] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, *String Axiverse*, *Phys. Rev.* **D81** (2010) 123530, [arXiv:0905.4720].

[636] M. Demirtas, C. Long, L. McAllister, and M. Stillman, *The Kreuzer-Skarke Axiverse*, *JHEP* **04** (2020) 138, [arXiv:1808.01282].

[637] M. Demirtas, N. Gendler, C. Long, L. McAllister, and J. Moritz, *PQ axiverse*, *JHEP* **06** (2023) 092, [arXiv:2112.04503].

[638] C. O'Hare, "cajohare/axionlimits: Axionlimits." https://cajohare.github.io/AxionLimits/, July, 2020.

[639] **ADMX** Collaboration, C. Bartram et al., *Dark matter axion search using a Josephson Traveling wave parametric amplifier*, *Rev. Sci. Instrum.* **94** (2023), no. 4 044703, [arXiv:2110.10262].

[640] **ADMX** Collaboration, C. Boutan et al., *Piezoelectrically Tuned Multimode Cavity Search for Axion Dark Matter*, *Phys. Rev. Lett.* **121** (2018), no. 26 261302, [arXiv:1901.00920].

[641] S. J. Asztalos, G. Carosi, C. Hagmann, D. Kinion, K. van Bibber, M. Hotz, L. J. Rosenberg, G. Rybka, J. Hoskins, J. Hwang, P. Sikivie, D. B. Tanner, R. Bradley, J. Clarke, and ADMX Collaboration, *SQUID-Based Microwave Cavity Search for Dark-Matter Axions*, **104** (Jan., 2010) 041301, [arXiv:0910.5914].

[642] **ADMX** Collaboration, N. Du et al., *A Search for Invisible Axion Dark Matter with the Axion Dark Matter Experiment*, *Phys. Rev. Lett.* **120** (2018), no. 15 151301, [arXiv:1804.05750].

[643] **ADMX** Collaboration, T. Braine et al., *Extended Search for the Invisible Axion with the Axion Dark Matter Experiment*, *Phys. Rev. Lett.* **124** (2020), no. 10 101303, [arXiv:1910.08638].

[644] **ADMX** Collaboration, C. Bartram et al., *Search for Invisible Axion Dark Matter in the 3.3–4.2 μeV Mass Range*, *Phys. Rev. Lett.* **127** (2021), no. 26 261803, [arXiv:2110.06096].

[645] Y. Kahn, B. R. Safdi, and J. Thaler, *Broadband and Resonant Approaches to Axion Dark Matter Detection*, *Phys. Rev. Lett.* **117** (2016), no. 14 141801, [arXiv:1602.01086].

[646] J. L. Ouellet et al., *Design and implementation of the ABRACADABRA-10 cm axion dark matter search*, *Phys. Rev. D* **99** (2019), no. 5 052012, [arXiv:1901.10652].

[647] J. L. Ouellet et al., *First Results from ABRACADABRA-10 cm: A Search for Sub-μeV Axion Dark Matter*, *Phys. Rev. Lett.* **122** (2019), no. 12 121802, [arXiv:1810.12257].

[648] J. W. Foster, N. L. Rodd, and B. R. Safdi, *Revealing the Dark Matter Halo with Axion Direct Detection*, *Phys. Rev. D* **97** (2018), no. 12 123006, [arXiv:1711.10489].

[649] **DMRadio** Collaboration, L. Brouwer et al., *Projected sensitivity of DMRadio-m3: A search for the QCD axion below 1 μeV*, *Phys. Rev. D* **106** (2022), no. 10 103008, [arXiv:2204.13781].

[650] K. Ehret et al., *New ALPS Results on Hidden-Sector Lightweights*, *Phys. Lett. B* **689** (2010) 149–155, [arXiv:1004.1313].

[651] M. D. Ortiz et al., *Design of the ALPS II optical system*, *Phys. Dark Univ.* **35** (2022) 100968, [arXiv:2009.14294].

[652] **OSQAR** Collaboration, R. Ballou et al., *New exclusion limits on scalar and pseudoscalar axionlike particles from light shining through a wall*, *Phys. Rev. D* **92** (2015), no. 9 092002, [arXiv:1506.08082].

[653] A. Garcon et al., *Constraints on bosonic dark matter from ultralow-field nuclear magnetic resonance*, *Sci. Adv.* **5** (2019), no. 10 eaax4539, [arXiv:1902.04644].

[654] T. Wu et al., *Search for Axionlike Dark Matter with a Liquid-State Nuclear Spin Comagnetometer*, *Phys. Rev. Lett.* **122** (2019), no. 19 191302, [arXiv:1901.10843].

[655] D. F. Jackson Kimball et al., *Overview of the Cosmic Axion Spin Precession Experiment (CASPEr)*, *Springer Proc. Phys.* **245** (2020) 105–121, [arXiv:1711.08999].

[656] **NASDUCK** Collaboration, I. M. Bloch, R. Shaham, Y. Hochberg, E. Kuflik, T. Volansky, and O. Katz, *Constraints on axion-like dark matter from a SERF comagnetometer*, *Nature Commun.* **14** (2023), no. 1 5784, [arXiv:2209.13588].

[657] **NASDUCK** Collaboration, I. M. Bloch, G. Ronen, R. Shaham, O. Katz, T. Volansky, and O. Katz, *New constraints on axion-like dark matter using a Floquet quantum detector*, *Sci. Adv.* **8** (2022), no. 5 abl8919, [arXiv:2105.04603].

[658] R. Janish and E. Pinetti, *Hunting Dark Matter Lines in the Infrared Background with the James Webb Space Telescope*, arXiv:2310.15395.

[659] W. Yin et al., *First Result for Dark Matter Search by WINERED*, arXiv:2402.07976.

[660] E. Todarello, M. Regis, J. Reynoso-Cordova, M. Taoso, D. Vaz, J. Brinchmann, M. Steinmetz, and S. L. Zoutendijke, *Robust bounds on ALP dark matter from dwarf spheroidal galaxies in the optical MUSE-Faint survey*, arXiv:2307.07403.

[661] D. Grin, G. Covone, J.-P. Kneib, M. Kamionkowski, A. Blain, and E. Jullo, *A Telescope Search for Decaying Relic Axions*, *Phys. Rev. D* **75** (2007) 105018, [astro-ph/0611502].

[662] K. Nakayama and W. Yin, *Anisotropic cosmic optical background bound for decaying dark matter in light of the LORRI anomaly*, *Phys. Rev. D* **106** (2022), no. 10 103505, [arXiv:2205.01079].

[663] P. Carenza, G. Lucente, and E. Vitagliano, *Probing the blue axion with cosmic optical background anisotropies*, *Phys. Rev. D* **107** (2023), no. 8 083032, [arXiv:2301.06560].

[664] J. W. Foster, M. Kongsore, C. Dessert, Y. Park, N. L. Rodd, K. Cranmer, and B. R. Safdi, *Deep Search for Decaying Dark Matter with XMM-Newton Blank-Sky Observations*, *Phys. Rev. Lett.* **127** (2021), no. 5 051101, [arXiv:2102.02207].

[665] F. Capozzi, R. Z. Ferreira, L. Lopez-Honorez, and O. Mena, *CMB and Lyman-α constraints on dark matter decays to photons*, *JCAP* **06** (2023) 060, [arXiv:2303.07426].

[666] C. Fong, K. C. Y. Ng, and Q. Liu, *Searching for Particle Dark Matter with eROSITA Early Data*, arXiv:2401.16747.

[667] J. L. Bernal, A. Caputo, G. Sato-Polito, J. Mirocha, and M. Kamionkowski, *Seeking dark matter with γ-ray attenuation*, *Phys. Rev. D* **107** (2023), no. 10 103046, [arXiv:2208.13794].

[668] M. Buschmann, C. Dessert, J. W. Foster, A. J. Long, and B. R. Safdi, *Upper Limit on the QCD Axion Mass from Isolated Neutron Star Cooling*, *Phys. Rev. Lett.* **128** (2022), no. 9 091102, [arXiv:2111.09892].

[669] M. Baryakhtar, M. Galanis, R. Lasenby, and O. Simon, *Black hole superradiance of self-interacting scalar fields*, *Phys. Rev. D* **103** (2021), no. 9 095019, [arXiv:2011.11646].

[670] V. M. Mehta, M. Demirtas, C. Long, D. J. E. Marsh, L. McAllister, and M. J. Stott, *Superradiance in string theory*, *JCAP* **07** (2021) 033, [arXiv:2103.06812].

[671] M. Baryakhtar et al., *Dark Matter In Extreme Astrophysical Environments*, in *Snowmass 2021*, 3, 2022. arXiv:2203.07984.

[672] M. J. Stott, *Ultralight Bosonic Field Mass Bounds from Astrophysical Black Hole Spin*, arXiv:2009.07206.

[673] K. Langhoff, N. J. Outmezguine, and N. L. Rodd, *Irreducible Axion Background*, *Phys. Rev. Lett.* **129** (2022), no. 24 241101, [arXiv:2209.06216].

[674] **CAST** Collaboration, S. Andriamonje et al., *An Improved limit on the axion-photon coupling from the CAST experiment*, *JCAP* **04** (2007) 010, [hep-ex/0702006].

[675] **CAST** Collaboration, V. Anastassopoulos et al., *New CAST Limit on the Axion-Photon Interaction*, *Nature Phys.* **13** (2017) 584–590, [arXiv:1705.02290].

[676] M. J. Dolan, F. J. Hiskens, and R. R. Volkas, *Advancing globular cluster constraints on the axion-photon coupling*, *JCAP* **10** (2022) 096, [arXiv:2207.03102].

[677] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi, and O. Straniero, *Revisiting the bound on axion-photon coupling from Globular Clusters*, *Phys. Rev. Lett.* **113** (2014), no. 19 191302, [arXiv:1406.6053].

[678] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, *Phys. Rev. Lett.* **32** (1974) 438–441.

[679] H. Georgi, *The State of the Art—Gauge Theories*, *AIP Conf. Proc.* **23** (1975) 575–582.

[680] H. Fritzsch and P. Minkowski, *Unified Interactions of Leptons and Hadrons*, *Annals Phys.* **93** (1975) 193–266.

[681] G. G. Ross, *GRAND UNIFIED THEORIES*. 1985.

[682] S. Raby, *Supersymmetric Grand Unified Theories: From Quarks to Strings via SUSY GUTs*, vol. 939. Springer, 2017.

[683] **Muon g-2** Collaboration, B. Abi et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (2021), no. 14 141801, [arXiv:2104.03281].

[684] **Muon g-2** Collaboration, G. W. Bennett et al., *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys. Rev. D* **73** (2006) 072003, [hep-ex/0602035].

[685] T. Aoyama et al., *The anomalous magnetic moment of the muon in the Standard Model*, *Phys. Rept.* **887** (2020) 1–166, [arXiv:2006.04822].

[686] **Muon g-2** Collaboration, D. P. Aguillard et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*, *Phys. Rev. Lett.* **131** (2023), no. 16 161802, [arXiv:2308.06230].

[687] S. Borsanyi et al., *Leading hadronic contribution to the muon magnetic moment from lattice QCD*, *Nature* **593** (2021), no. 7857 51–55, [arXiv:2002.12347].

[688] **CMD-3** Collaboration, F. V. Ignatov et al., *Measurement of the* $e^+e^- \to \pi^+\pi^-$ *cross section from threshold to 1.2 GeV with the CMD-3 detector*, arXiv:2302.08834.

[689] A. Czarnecki, W. J. Marciano, and A. Sirlin, *Neutron Lifetime and Axial Coupling Connection*, *Phys. Rev. Lett.* **120** (2018), no. 20 202002, [arXiv:1802.01804].

[690] J. C. Hardy and I. S. Towner, *Superallowed* $0^+ \to 0^+$ *nuclear* β *decays: 2020 critical survey, with implications for* V_{ud} *and CKM unitarity*, *Phys. Rev. C* **102** (2020), no. 4 045501.

[691] **KLOE** Collaboration, F. Ambrosino et al., *Measurement of the absolute branching ratio for the* $K^+ \to \mu^+\nu(\gamma)$ *decay with the KLOE detector*, *Phys. Lett. B* **632** (2006) 76–80, [hep-ex/0509045].

[692] **KLOE** Collaboration, F. Ambrosino et al., *Measurement of the charged kaon lifetime with the KLOE detector*, *JHEP* **01** (2008) 073, [arXiv:0712.1112].

[693] J. Byrne and P. G. Dawber, *A Revised Value for the Neutron Lifetime Measured Using a Penning Trap*, *EPL* **33** (1996) 187.

[694] A. T. Yue, M. S. Dewey, D. M. Gilliam, G. L. Greene, A. B. Laptev, J. S. Nico, W. M. Snow, and F. E. Wietfeldt, *Improved Determination of the Neutron Lifetime*, *Phys. Rev. Lett.* **111** (2013), no. 22 222501, [arXiv:1309.2623].

[695] **UCN**τ Collaboration, F. M. Gonzalez et al., *Improved neutron lifetime measurement with UCN*τ, *Phys. Rev. Lett.* **127** (2021), no. 16 162501, [arXiv:2106.10375].

[696] **HFLAV** Collaboration, Y. S. Amhis et al., *Averages of b-hadron, c-hadron, and* τ*-lepton properties as of 2021*, *Phys. Rev. D* **107** (2023), no. 5 052008, [arXiv:2206.07501].

[697] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality in* $b \to s\ell^+\ell^-$ *decays*, *Phys. Rev. Lett.* **131** (2023), no. 5 051803, [arXiv:2212.09152].

[698] **Belle-II** Collaboration, F. Abudinén et al., *Search for* $B^+ \to K^+\nu\bar{\nu}$ *Decays Using an Inclusive Tagging Method at Belle II*, *Phys. Rev. Lett.* **127** (2021), no. 18 181802, [arXiv:2104.12624].

[699] **LHCb** Collaboration, R. Aaij et al., *Measurement of CP Violation in the Decay* $B^+ \to K^+\pi^0$, *Phys. Rev. Lett.* **126** (2021), no. 9 091802, [arXiv:2012.12789].

[700] A. J. Buras, R. Fleischer, S. Recksiegel, and F. Schwab, *B —> pi pi, new physics in B —> pi K and implications for rare K and B decays*, *Phys. Rev. Lett.* **92** (2004) 101804, [hep-ph/0312259].

[701] **CMS** Collaboration, A. M. Sirunyan et al., *Search for resonances in the mass spectrum of muon pairs produced in association with b quark jets in proton-proton collisions at* $\sqrt{s} = 8$ *and 13 TeV*, *JHEP* **11** (2018) 161, [arXiv:1808.01890].

[702] **CMS** Collaboration, A. M. Sirunyan et al., *Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at* $\sqrt{s} = 8$ *and 13 TeV*, *Phys. Lett. B* **793** (2019) 320–347, [arXiv:1811.08459].

[703] O. Fischer et al., *Unveiling hidden physics at the LHC*, *Eur. Phys. J. C* **82** (2022), no. 8 665, [arXiv:2109.06065].

[704] T. Bose and S. P. Griso, "Existing anomalies in the atlas and cms physics program."

[705] **LSND** Collaboration, C. Athanassopoulos et al., *Evidence for anti-muon-neutrino —> anti-electron-neutrino oscillations from the LSND experiment at LAMPF*, *Phys. Rev. Lett.* **77** (1996) 3082–3085, [nucl-ex/9605003].

[706] **LSND** Collaboration, A. Aguilar-Arevalo et al., *Evidence for neutrino oscillations from the observation of* $\bar{\nu}_e$ *appearance in a* $\bar{\nu}_\mu$ *beam*, *Phys. Rev. D* **64** (2001) 112007, [hep-ex/0104049].

[707] **MiniBooNE** Collaboration, A. A. Aguilar-Arevalo et al., *Updated MiniBooNE neutrino oscillation results with increased data and new background studies*, *Phys. Rev. D* **103** (2021), no. 5 052002, [arXiv:2006.16883].

[708] **MicroBooNE** Collaboration, P. Abratenko et al., *First Constraints on Light Sterile Neutrino Oscillations from Combined Appearance and Disappearance Searches with the MicroBooNE Detector*, *Phys. Rev. Lett.* **130** (2023), no. 1 011801, [arXiv:2210.10216].

[709] **SAGE** Collaboration, J. N. Abdurashitov et al., *Measurement of the response of the Russian-American gallium experiment to neutrinos from a Cr-51 source*, *Phys. Rev. C* **59** (1999) 2246–2263, [hep-ph/9803418].

[710] **GALLEX** Collaboration, W. Hampel et al., *Final results of the Cr-51 neutrino source experiments in GALLEX*, *Phys. Lett. B* **420** (1998) 114–126.

[711] V. V. Barinov et al., *Results from the Baksan Experiment on Sterile Transitions (BEST)*, *Phys. Rev. Lett.* **128** (2022), no. 23 232501, [arXiv:2109.11482].

[712] V. V. Barinov et al., *Search for electron-neutrino transitions to sterile states in the BEST experiment*, *Phys. Rev. C* **105** (2022), no. 6 065502, [arXiv:2201.07364].

[713] J. M. Berryman, P. Coloma, P. Huber, T. Schwetz, and A. Zhou, *Statistical significance of the sterile-neutrino hypothesis in the context of reactor and gallium data*, *JHEP* **02** (2022) 055, [arXiv:2111.12530].

[714] **NF02 Topical Group** Collaboration, G. Karagiorgi, B. R. Littlejohn, P. Machado, and A. Sousa, *Snowmass Neutrino Frontier: NF02 Topical Group Report on Understanding Experimental Neutrino Anomalies*, in *Snowmass 2021*, 9, 2022. arXiv:2209.05352.

[715] M. A. Acero et al., *White Paper on Light Sterile Neutrino Searches and Related Phenomenology*, arXiv:2203.07323.

[716] E. Abdalla et al., *Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies*, *JHEAp* **34** (2022) 49–211, [arXiv:2203.06142].

[717] **AMS** Collaboration, M. Aguilar et al., *Towards Understanding the Origin of Cosmic-Ray Positrons*, *Phys. Rev. Lett.* **122** (2019), no. 4 041102.

[718] **PAMELA** Collaboration, O. Adriani et al., *Cosmic-Ray Positron Energy Spectrum Measured by PAMELA*, *Phys. Rev. Lett.* **111** (2013) 081102, [arXiv:1308.0133].

[719] **AMS** Collaboration, M. Aguilar et al., *Antiproton Flux, Antiproton-to-Proton Flux Ratio, and Properties of Elementary Particle Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **117** (2016), no. 9 091103.

[720] O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, L. Bonechi, M. Bongi, V. Bonvicini, S. Borisov, S. Bottai, A. Bruno, F. Cafagna, D. Campana, R. Carbone, P. Carlson, M. Casolino, G. Castellini, L. Consiglio, M. P. De Pascale, C. De Santis, N. De Simone, V. Di Felice, A. M. Galper, W. Gillard, L. Grishantseva, G. Jerse, A. V. Karelin, S. V. Koldashov, S. Y. Krutkov, A. N. Kvashnin, A. Leonov, V. Malakhov, V. Malvezzi, L. Marcelli, A. G. Mayorov, W. Menn, V. V. Mikhailov, E. Mocchiutti, A. Monaco, N. Mori, N. Nikonov, G. Osteria, F. Palma, P. Papini, M. Pearce, P. Picozza, C. Pizzolotto, M. Ricci, S. B. Ricciarini, L. Rossetto, R. Sarkar, M. Simon, R. Sparvoli, P. Spillantini, Y. I. Stozhkov, A. Vacchi, E. Vannuccini, G. Vasilyev, S. A. Voronov, Y. T. Yurkin, J. Wu, G. Zampa, N. Zampa, and V. G. Zverev, *PAMELA Measurements of Cosmic-Ray Proton and Helium Spectra*, *Science* **332** (Apr., 2011) 69, [arXiv:1103.4055].

[721] S. Ting, *The First Five Years of the Alpha Magnetic Spectrometer on the International Space Station*, *Press Conference at CERN, December 8* (2016).

[722] S. Ting, *Latest Results from the AMS Experiment on the International Space Station*, *Colloquium at CERN, May 24* (2018).

[723] A. Kounine, *AMS Experiment on the International Space Station*, *Next Generation of AstroParticle Experiments in Space (NextGAPES-2019)* http://www.sinp.msu.ru/contrib/NextGAPES/files/AMS_AK.pdf, *June 21* (2019).

[724] R. K. Leane et al., *Snowmass2021 Cosmic Frontier White Paper: Puzzling Excesses in Dark Matter Searches and How to Resolve Them*, arXiv:2203.06859.

[725] J. Heisig, M. Korsmeier, and M. W. Winkler, *Dark matter or correlated errors: Systematics of the AMS-02 antiproton excess*, *Phys. Rev. Res.* **2** (2020), no. 4 043017, [arXiv:2005.04237].

[726] M. Boudaud, Y. Génolini, L. Derome, J. Lavalle, D. Maurin, P. Salati, and P. D. Serpico, *AMS-02 antiprotons' consistency with a secondary astrophysical origin*, *Phys. Rev. Res.* **2** (2020), no. 2 023022, [arXiv:1906.07119].

[727] P. De La Torre Luque, M. W. Winkler, and T. Linden, *Cosmic-Ray Propagation Models Elucidate the Prospects for Antinuclei Detection*, arXiv:2404.13114.

[728] **DESI** Collaboration, A. G. Adame et al., *DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations*, arXiv:2404.03002.

[729] M. Cortês and A. R. Liddle, *Interpreting DESI's evidence for evolving dark energy*, arXiv:2404.08056.

[730] P. Adshead, J. T. Giblin, T. R. Scully, and E. I. Sfakianakis, *Gauge-preheating and the end of axion inflation*, *JCAP* **1512** (2015), no. 12 034, [arXiv:1502.06506].

[731] M. M. Anber and L. Sorbo, *Naturally inflating on steep potentials through electromagnetic dissipation*, *Phys. Rev.* **D81** (2010) 043534, [arXiv:0908.4089].

[732] P. Adshead and M. Wyman, *Chromo-Natural Inflation: Natural inflation on a steep potential with classical non-Abelian gauge fields*, *Phys. Rev. Lett.* **108** (2012) 261302, [arXiv:1202.2366].

[733] E. Silverstein and A. Westphal, *Monodromy in the CMB: Gravity Waves and String Inflation*, *Phys. Rev.* **D78** (2008) 106003, [arXiv:0803.3085].

[734] L. McAllister, E. Silverstein, and A. Westphal, *Gravity Waves and Linear Inflation from Axion Monodromy*, *Phys.Rev.* **D82** (2010) 046003, [arXiv:0808.0706].

[735] K. S. Babu, S. M. Barr, and D. Seckel, *Axion dissipation through the mixing of Goldstone bosons*, *Phys. Lett.* **B336** (1994) 213–220, [hep-ph/9406308].

[736] J. E. Kim, H. P. Nilles, and M. Peloso, *Completing natural inflation*, *JCAP* **0501** (2005) 005, [hep-ph/0409138].

[737] T. C. Bachlechner, M. Dias, J. Frazer, and L. McAllister, *Chaotic inflation with kinetic alignment of axion fields*, *Phys. Rev.* **D91** (2015), no. 2 023520, [arXiv:1404.7496].

[738] H.-C. Cheng and D. E. Kaplan, *Axions and a gauged Peccei-Quinn symmetry*, hep-ph/0103346.

[739] P. Anastasopoulos, M. Bianchi, E. Dudas, and E. Kiritsis, *Anomalies, anomalous U(1)'s and generalized Chern-Simons terms*, *JHEP* **11** (2006) 057, [hep-th/0605225].

[740] G. Shiu, W. Staessens, and F. Ye, *Widening the Axion Window via Kinetic and Stückelberg Mixings*, *Phys. Rev. Lett.* **115** (2015) 181601, [arXiv:1503.01015].

[741] G. Shiu, W. Staessens, and F. Ye, *Large Field Inflation from Axion Mixing*, *JHEP* **06** (2015) 026, [arXiv:1503.02965].

[742] K. Choi, C. S. Shin, and S. Yun, *Axion scales and couplings with Stückelberg mixing*, *JHEP* **12** (2019) 033, [arXiv:1909.11685].

[743] M. Berg, E. Pajer, and S. Sjors, *Dante's Inferno*, *Phys. Rev.* **D81** (2010) 103535, [arXiv:0912.1341].

[744] B. Holdom, *Two U(1)'s and Epsilon Charge Shifts*, *Phys. Lett.* **166B** (1986) 196–198.

[745] G. Shiu, P. Soler, and F. Ye, *Milli-Charged Dark Matter in Quantum Gravity and String Theory*, *Phys. Rev. Lett.* **110** (2013), no. 24 241304, [arXiv:1302.5471].

[746] K. Choi, H. Kim, and S. Yun, *Natural inflation with multiple sub-Planckian axions*, *Phys. Rev.* **D90** (2014) 023545, [arXiv:1404.6209].

[747] K. Choi and S. H. Im, *Realizing the relaxion from multiple axions and its UV completion with high scale supersymmetry*, *JHEP* **01** (2016) 149, [arXiv:1511.00132].

[748] D. E. Kaplan and R. Rattazzi, *Large field excursions and approximate discrete symmetries from a clockwork axion*, *Phys. Rev.* **D93** (2016), no. 8 085007, [arXiv:1511.01827].

[749] M. Farina, D. Pappadopulo, F. Rompineve, and A. Tesi, *The photo-philic QCD axion*, *JHEP* **01** (2017) 095, [arXiv:1611.09855].

[750] P. Agrawal, J. Fan, M. Reece, and L.-T. Wang, *Experimental Targets for Photon Couplings of the QCD Axion*, *JHEP* **02** (2018) 006, [arXiv:1709.06085].

[751] D. H. Lyth, *What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?*, *Phys. Rev. Lett.* **78** (1997) 1861–1863, [hep-ph/9606387].

[752] T. Banks, M. Dine, P. J. Fox, and E. Gorbatov, *On the possibility of large axion decay constants*, *JCAP* **0306** (2003) 001, [hep-th/0303252].

[753] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The String landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[754] H. Ooguri and C. Vafa, *On the Geometry of the String Landscape and the Swampland*, *Nucl. Phys.* **B766** (2007) 21–33, [hep-th/0605264].

[755] T. Rudelius, *On the Possibility of Large Axion Moduli Spaces*, *JCAP* **1504** (2015), no. 04 049, [arXiv:1409.5793].

[756] B. Heidenreich, M. Reece, and T. Rudelius, *The Weak Gravity Conjecture and Emergence from an Ultraviolet Cutoff*, *Eur. Phys. J.* **C78** (2018), no. 4 337, [arXiv:1712.01868].

[757] P. Agrawal, J. Fan, and M. Reece, *Clockwork Axions in Cosmology: Is Chromonatural Inflation Chrononatural?*, *JHEP* **10** (2018) 193, [arXiv:1806.09621].

[758] P. Agrawal, G. Marques-Tavares, and W. Xue, *Opening up the QCD axion window*, *JHEP* **03** (2018) 049, [arXiv:1708.05008].

[759] N. Kitajima, T. Sekiguchi, and F. Takahashi, *Cosmological abundance of the QCD axion coupled to hidden photons*, *Phys. Lett.* **B781** (2018) 684–687, [arXiv:1711.06590].

[760] P. Agrawal, N. Kitajima, M. Reece, T. Sekiguchi, and F. Takahashi, *Relic Abundance of Dark Photon Dark Matter*, *Phys. Lett.* **B801** (2020) 135136, [arXiv:1810.07188].

[761] R. T. Co, A. Pierce, Z. Zhang, and Y. Zhao, *Dark Photon Dark Matter Produced by Axion Oscillations*, *Phys. Rev.* **D99** (2019), no. 7 075002, [arXiv:1810.07196].

[762] M. Bastero-Gil, J. Santiago, L. Ubaldi, and R. Vega-Morales, *Vector dark matter production at the end of inflation*, *JCAP* **1904** (2019), no. 04 015, [arXiv:1810.07208].

[763] N. Fonseca, B. von Harling, L. de Lima, and C. S. Machado, *Super-Planckian axions from near-conformality*, *Phys. Rev.* **D100** (2019), no. 10 105019, [arXiv:1906.10193].

[764] G. Dvali, *Black Holes and Large N Species Solution to the Hierarchy Problem*, *Fortsch. Phys.* **58** (2010) 528–536, [arXiv:0706.2050].

[765] A. Hebecker, P. Mangat, F. Rompineve, and L. T. Witkowski, *Winding out of the Swamp: Evading the Weak Gravity Conjecture with F-term Winding Inflation?*, *Phys. Lett.* **B748** (2015) 455–462, [arXiv:1503.07912].

[766] D. B. Kaplan, *Opening the Axion Window*, *Nucl. Phys.* **B260** (1985) 215–226.

[767] M. Srednicki, *Axion Couplings to Matter. 1. CP Conserving Parts*, *Nucl. Phys.* **B260** (1985) 689–700.

[768] H. Georgi, D. B. Kaplan, and L. Randall, *Manifesting the Invisible Axion at Low-energies*, *Phys. Lett.* **169B** (1986) 73–78.

[769] T. Higaki, N. Kitajima, and F. Takahashi, *Hidden axion dark matter decaying through mixing with QCD axion and the 3.5 keV X-ray line*, *JCAP* **1412** (2014), no. 12 004, [arXiv:1408.3936].

[770] H. J. S. Smith, *On systems of linear indeterminate equations and congruences*, *Proceedings of the Royal Society of London* (1862), no. 11 86–89.

[771] Y. Hosotani, *Dynamical Gauge Symmetry Breaking as the Casimir Effect*, *Phys. Lett.* **129B** (1983) 193–197.

[772] N. Arkani-Hamed, H.-C. Cheng, P. Creminelli, and L. Randall, *Extra natural inflation*, *Phys.Rev.Lett.* **90** (2003) 221302, [hep-th/0301218].

[773] B. Feng, M.-z. Li, R.-J. Zhang, and X.-m. Zhang, *An inflation model with large variations in spectral index*, *Phys. Rev.* **D68** (2003) 103511, [astro-ph/0302479].

[774] A. Delgado, A. Pomarol, and M. Quiros, *Supersymmetry and electroweak breaking from extra dimensions at the TeV scale*, *Phys. Rev.* **D60** (1999) 095008, [hep-ph/9812489].

[775] H.-C. Cheng, K. T. Matchev, and M. Schmaltz, *Radiative corrections to Kaluza-Klein masses*, *Phys. Rev.* **D66** (2002) 036005, [hep-ph/0204342].

[776] H. Hatanaka, T. Inami, and C. S. Lim, *The Gauge hierarchy problem and higher dimensional gauge theories*, *Mod. Phys. Lett.* **A13** (1998) 2601–2612, [hep-th/9805067].

[777] I. Antoniadis, K. Benakli, and M. Quiros, *Finite Higgs mass without supersymmetry*, *New J. Phys.* **3** (2001) 20, [hep-th/0108005].

[778] N. Kaloper and L. Sorbo, *A Natural Framework for Chaotic Inflation*, *Phys. Rev. Lett.* **102** (2009) 121301, [arXiv:0811.1989].

[779] X. Dong, B. Horn, E. Silverstein, and A. Westphal, *Simple exercises to flatten your potential*, *Phys. Rev.* **D84** (2011) 026011, [arXiv:1011.4521].

[780] N. Kaloper, A. Lawrence, and L. Sorbo, *An Ignoble Approach to Large Field Inflation*, *JCAP* **1103** (2011) 023, [arXiv:1101.0026].

[781] F. Marchesano, G. Shiu, and A. M. Uranga, *F-term Axion Monodromy Inflation*, *JHEP* **09** (2014) 184, [arXiv:1404.3040].

[782] K. Furuuchi, *Excursions through KK modes*, *JCAP* **1607** (2016), no. 07 008, [arXiv:1512.04684].

[783] M. J. Dolan, P. Draper, J. Kozaczuk, and H. Patel, *Transplanckian Censorship and Global Cosmic Strings*, *JHEP* **04** (2017) 133, [arXiv:1701.05572].

[784] L. McAllister, E. Silverstein, A. Westphal, and T. Wrase, *The Powers of Monodromy*, *JHEP* **09** (2014) 123, [arXiv:1405.3652].

[785] R. Bousso, *A Covariant entropy conjecture*, *JHEP* **07** (1999) 004, [hep-th/9905177].

[786] C. Vafa, *The String landscape and the swampland*, hep-th/0509212.

[787] M. Dine, N. Seiberg, X. G. Wen, and E. Witten, *Nonperturbative Effects on the String World Sheet*, *Nucl. Phys. B* **278** (1986) 769–789.

[788] M. Dine, N. Seiberg, X. G. Wen, and E. Witten, *Nonperturbative Effects on the String World Sheet. 2.*, *Nucl. Phys. B* **289** (1987) 319–363.

[789] K. Becker, M. Becker, and A. Strominger, *Five-branes, membranes and nonperturbative string theory*, *Nucl. Phys. B* **456** (1995) 130–152, [hep-th/9507158].

[790] E. Witten, *Nonperturbative superpotentials in string theory*, *Nucl. Phys. B* **474** (1996) 343–360, [hep-th/9604030].

[791] H. Ooguri and C. Vafa, *Summing up D instantons*, *Phys. Rev. Lett.* **77** (1996) 3296–3298, [hep-th/9608079].

[792] E. Witten, *Dyons of Charge e theta/2 pi*, *Phys. Lett.* **86B** (1979) 283–287.

[793] B. Heidenreich, J. McNamara, M. Montero, M. Reece, T. Rudelius, and I. Valenzuela, *Chern-Weil Global Symmetries and How Quantum Gravity Avoids Them*, arXiv:2012.00009.

[794] J. Stout, *Instanton Expansions and Phase Transitions*, arXiv:2012.11605.

[795] J. McNamara, *The Charge Lattice of Maxwell Theory with an Axion*, . Unpublished notes, May 2020.

[796] J. Polchinski, *Monopoles, duality, and string theory*, *Int.J.Mod.Phys.* **A19S1** (2004) 145–156, [hep-th/0304042].

[797] A. de la Fuente, P. Saraswat, and R. Sundrum, *Natural Inflation and Quantum Gravity*, *Phys. Rev. Lett.* **114** (2015), no. 15 151303, [arXiv:1412.3457].

[798] P. Sikivie, *Experimental Tests of the Invisible Axion*, *Phys. Rev. Lett.* **51** (1983) 1415–1417. [Erratum: Phys.Rev.Lett. 52, 695 (1984)].

[799] P. Sikivie, *Detection Rates for 'Invisible' Axion Searches*, *Phys. Rev. D* **32** (1985) 2988. [Erratum: Phys.Rev.D 36, 974 (1987)].

[800] **ADMX** Collaboration, S. J. Asztalos et al., *A SQUID-based microwave cavity search for dark-matter axions*, *Phys. Rev. Lett.* **104** (2010) 041301, [arXiv:0910.5914].

[801] R. Jackiw, *Charge and Mass Spectrum of Quantum Solitons*, in *Gauge Theories and Modern Field Theory* (R. Arnowitt and P. Nath, eds.), p. 377, Jan., 1976.

[802] D. Gaiotto, A. Kapustin, Z. Komargodski, and N. Seiberg, *Theta, Time Reversal, and Temperature*, *JHEP* **05** (2017) 091, [arXiv:1703.00501].

[803] W. Fischler and J. Preskill, *Dyon–Axion Dynamics*, *Phys. Lett. B* **125** (1983) 165–170.

[804] G. 't Hooft, *Magnetic Monopoles in Unified Gauge Theories*, *Nucl.Phys.* **B79** (1974) 276–284.

[805] A. M. Polyakov, *Particle Spectrum in the Quantum Field Theory*, *JETP Lett.* **20** (1974) 194–195.

[806] S. R. Coleman and E. J. Weinberg, *Radiative Corrections as the Origin of Spontaneous Symmetry Breaking*, *Phys. Rev. D* **7** (1973) 1888–1910.

[807] M. Kawasaki, F. Takahashi, and M. Yamada, *Suppressing the QCD Axion Abundance by Hidden Monopoles*, *Phys. Lett. B* **753** (2016) 677–681, [arXiv:1511.05030].

[808] Y. Nomura, S. Rajendran, and F. Sanches, *Axion Isocurvature and Magnetic Monopoles*, *Phys. Rev. Lett.* **116** (2016), no. 14 141803, [arXiv:1511.06347].

[809] M. Kawasaki, F. Takahashi, and M. Yamada, *Adiabatic suppression of the axion abundance and isocurvature due to coupling to hidden monopoles*, *JHEP* **01** (2018) 053, [arXiv:1708.06047].

[810] E. Poppitz and M. Unsal, *Seiberg-Witten and 'Polyakov-like' magnetic bion confinements are continuously connected*, *JHEP* **07** (2011) 082, [arXiv:1105.3969].

[811] I. K. Affleck and N. S. Manton, *Monopole Pair Production in a Magnetic Field*, *Nucl. Phys. B* **194** (1982) 38–64.

[812] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 12, 2007.

[813] Y. Hosotani, *Dynamical Mass Generation by Compact Extra Dimensions*, *Phys. Lett. B* **126** (1983) 309–313.

[814] N. Arkani-Hamed, H.-C. Cheng, P. Creminelli, and L. Randall, *Extra natural inflation*, *Phys. Rev. Lett.* **90** (2003) 221302, [hep-th/0301218].

[815] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, *Pseudoparticle Solutions of the Yang-Mills Equations*, *Phys. Lett. B* **59** (1975) 85–87.

[816] G. 't Hooft, *Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle*, *Phys. Rev. D* **14** (1976) 3432–3450. [Erratum: Phys.Rev.D 18, 2199 (1978)].

[817] H. Murayama and J. Shu, *Topological Dark Matter*, *Phys. Lett. B* **686** (2010) 162–165, [arXiv:0905.1720].

[818] Y. B. Zeldovich and M. Y. Khlopov, *On the Concentration of Relic Magnetic Monopoles in the Universe*, *Phys. Lett. B* **79** (1978) 239–241.

[819] J. Preskill, *Cosmological Production of Superheavy Magnetic Monopoles*, *Phys. Rev. Lett.* **43** (1979) 1365.

[820] A. Vilenkin and A. E. Everett, *Cosmic Strings and Domain Walls in Models with Goldstone and PseudoGoldstone Bosons*, *Phys. Rev. Lett.* **48** (1982) 1867–1870.

[821] P. Sikivie, *Of Axions, Domain Walls and the Early Universe*, *Phys. Rev. Lett.* **48** (1982) 1156–1159.

[822] R. L. Davis, *Cosmic Axions from Cosmic Strings*, *Phys. Lett. B* **180** (1986) 225–230.

[823] D. Harari and P. Sikivie, *On the Evolution of Global Strings in the Early Universe*, *Phys. Lett. B* **195** (1987) 361–365.

[824] E. P. S. Shellard, *Cosmic String Interactions*, *Nucl. Phys. B* **283** (1987) 624–656.

[825] R. L. Davis and E. P. S. Shellard, *Do Axions Need Inflation?*, *Nucl. Phys. B* **324** (1989) 167–186.

[826] C. Hagmann and P. Sikivie, *Computer simulations of the motion and decay of global strings*, *Nucl. Phys. B* **363** (1991) 247–280.

[827] R. A. Battye and E. P. S. Shellard, *Global string radiation*, *Nucl. Phys. B* **423** (1994) 260–304, [astro-ph/9311017].

[828] R. A. Battye and E. P. S. Shellard, *Axion string constraints*, *Phys. Rev. Lett.* **73** (1994) 2954–2957, [astro-ph/9403018]. [Erratum: Phys.Rev.Lett. 76, 2203–2204 (1996)].

[829] M. Yamaguchi, M. Kawasaki, and J. Yokoyama, *Evolution of axionic strings and spectrum of axions radiated from them*, *Phys. Rev. Lett.* **82** (1999) 4578–4581, [hep-ph/9811311].

[830] V. B. . Klaer and G. D. Moore, *The dark-matter axion mass*, *JCAP* **11** (2017) 049, [arXiv:1708.07521].

[831] M. Gorghetto, E. Hardy, and G. Villadoro, *Axions from Strings: the Attractive Solution*, *JHEP* **07** (2018) 151, [arXiv:1806.04677].

[832] A. Vaquero, J. Redondo, and J. Stadler, *Early seeds of axion miniclusters*, *JCAP* **04** (2019) 012, [arXiv:1809.09241].

[833] M. Buschmann, J. W. Foster, and B. R. Safdi, *Early-Universe Simulations of the Cosmological Axion*, *Phys. Rev. Lett.* **124** (2020), no. 16 161103, [arXiv:1906.00967].

[834] M. Gorghetto, E. Hardy, and G. Villadoro, *More axions from strings*, *SciPost Phys.* **10** (2021), no. 2 050, [arXiv:2007.04990].

[835] M. Dine, N. Fernandez, A. Ghalsasi, and H. H. Patel, *Comments on axions, domain walls, and cosmic strings*, *JCAP* **11** (2021) 041, [arXiv:2012.13065].

[836] P. Agrawal, A. Hook, and J. Huang, *A CMB Millikan experiment with cosmic axiverse strings*, *JHEP* **07** (2020) 138, [arXiv:1912.02823].

[837] J. N. Benabou, M. Buschmann, S. Kumar, Y. Park, and B. R. Safdi, *Signatures of Primordial Energy Injection from Axion Strings*, arXiv:2308.01334.

[838] C. G. Callan, Jr. and J. A. Harvey, *Anomalies and Fermion Zero Modes on Strings and Domain Walls*, *Nucl. Phys. B* **250** (1985) 427–436.

[839] D. B. Kaplan and A. Manohar, *Anomalous Vortices and Electromagnetism*, *Nucl. Phys. B* **302** (1988) 280–290.

[840] S. G. Naculich, *Axionic Strings: Covariant Anomalies and Bosonization of Chiral Zero Modes*, *Nucl. Phys. B* **296** (1988) 837–867.

[841] A. Manohar, *Anomalous Vortices and Electromagnetism. 2.*, *Phys. Lett. B* **206** (1988) 276. [Erratum: Phys.Lett. 209, 543 (1988)].

[842] J. A. Harvey and S. G. Naculich, *Cosmic Strings From Pseudoanomalous U(1)*s, *Phys. Lett. B* **217** (1989) 231–237.

[843] D. Harari and P. Sikivie, *Effects of a Nambu-Goldstone boson on the polarization of radio galaxies and the cosmic microwave background*, *Phys. Lett. B* **289** (1992) 67–72.

[844] J. D. Blum and J. A. Harvey, *Anomaly inflow for gauge defects*, *Nucl. Phys. B* **416** (1994) 119–136, [hep-th/9310035].

[845] J. A. Harvey and O. Ruchayskiy, *The Local structure of anomaly inflow*, *JHEP* **06** (2001) 044, [hep-th/0007037].

[846] B. Heidenreich, M. Reece, and T. Rudelius, *The Weak Gravity Conjecture and axion strings*, *JHEP* **11** (2021) 004, [arXiv:2108.11383].

[847] H. Fukuda and K. Yonekura, *Witten effect, anomaly inflow, and charge teleportation*, *JHEP* **01** (2021) 119, [arXiv:2010.02221].

[848] H. Fukuda, A. V. Manohar, H. Murayama, and O. Telem, *Axion strings are superconducting*, *JHEP* **06** (2021) 052, [arXiv:2010.02763].

[849] M. Ibe, S. Kobayashi, Y. Nakayama, and S. Shirai, *On Stability of Fermionic Superconducting Current in Cosmic String*, *JHEP* **05** (2021) 217, [arXiv:2102.05412].

[850] P. Agrawal, A. Hook, J. Huang, and G. Marques-Tavares, *Axion string signatures: a cosmological plasma collider*, *JHEP* **01** (2022) 103, [arXiv:2010.15848].

[851] E. Witten, *Superconducting Strings*, *Nucl. Phys. B* **249** (1985) 557–592.

[852] Y. Abe, Y. Hamada, and K. Yoshioka, *Electroweak axion string and superconductivity*, *JHEP* **06** (2021) 172, [arXiv:2010.02834].

[853] C. T. Hill and L. M. Widrow, *Superconducting Cosmic Strings with Massive Fermions*, *Phys. Lett. B* **189** (1987) 17–22.

[854] H. A. Weldon, *Effective Fermion Masses of Order gT in High Temperature Gauge Theories with Exact Chiral Invariance*, *Phys. Rev. D* **26** (1982) 2789.

[855] A. Achucarro and T. Vachaspati, *Semilocal and electroweak strings*, *Phys. Rept.* **327** (2000) 347–426, [hep-ph/9904229].

[856] R. Jackiw and P. Rossi, *Zero Modes of the Vortex - Fermion System*, *Nucl. Phys. B* **190** (1981) 681–691.

[857] G. D. Starkman, D. Stojkovic, and T. Vachaspati, *Neutrino zero modes on electroweak strings*, *Phys. Rev. D* **63** (2001) 085011, [hep-ph/0007071].

[858] D. Stojkovic, *Neutrino zero modes and stability of electroweak strings*, *Int. J. Mod. Phys. A* **16S1C** (2001) 1034–1036, [hep-th/0103216].

[859] G. Starkman, D. Stojkovic, and T. Vachaspati, *Zero modes of fermions with a general mass matrix*, *Phys. Rev. D* **65** (2002) 065003, [hep-th/0103039].

[860] S. G. Naculich, *Fermions destabilize electroweak strings*, *Phys. Rev. Lett.* **75** (1995) 998–1001, [hep-ph/9501388].

[861] H. Liu and T. Vachaspati, *Perturbed electroweak strings and fermion zero modes*, *Nucl. Phys. B* **470** (1996) 176–194, [hep-ph/9511216].

[862] P. Agrawal and A. Platschorre, *The monodromic axion-photon coupling*, *JHEP* **01** (2024) 169, [arXiv:2309.03934].

[863] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer, 1999.

[864] D. B. Kaplan, *Chiral fermions on the lattice*, *Nucl. Phys. B Proc. Suppl.* **30** (1993) 597–600.

[865] K. Aitken, A. Baumgartner, A. Karch, and B. Robinson, *3d Abelian Dualities with Boundaries*, *JHEP* **03** (2018) 053, [arXiv:1712.02801].

[866] E. H. Fradkin, *Field Theories of Condensed Matter Physics*, vol. 82. Cambridge Univ. Press, Cambridge, UK, 2, 2013.

[867] F. D. M. Haldane, *Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the 'Parity Anomaly'*, *Phys. Rev. Lett.* **61** (1988) 2015–2018.

[868] K. Jansen and M. Schmaltz, *Critical momenta of lattice chiral fermions*, *Phys. Lett. B* **296** (1992) 374–378, [hep-lat/9209002].

[869] E. Witten and K. Yonekura, *Anomaly Inflow and the η-Invariant*, in *The Shoucheng Zhang Memorial Workshop*, 9, 2019. arXiv:1909.08775.

[870] J. Goldstone and F. Wilczek, *Fractional Quantum Numbers on Solitons*, *Phys. Rev. Lett.* **47** (1981) 986–989.

[871] A. Ross, *Multipole expansion at the level of the action*, *Phys. Rev. D* **85** (2012) 125033, [arXiv:1202.4750].

[872] C. G. Callan, Jr., *Disappearing Dyons*, *Phys. Rev. D* **25** (1982) 2141.

[873] **Fermi-LAT** Collaboration, M. Ajello et al., *Fermi-LAT Observations of High-Energy γ-Ray Emission Toward the Galactic Center*, *Astrophys. J.* **819** (2016), no. 1 44, [arXiv:1511.02938].

[874] L. Goodenough and D. Hooper, *Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope*, arXiv:0910.2998.

[875] D. Hooper and L. Goodenough, *Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope*, *Phys. Lett.* **B697** (2011) 412–428, [arXiv:1010.2752].

[876] D. Hooper and T. Linden, *On The Origin Of The Gamma Rays From The Galactic Center*, *Phys. Rev. D* **84** (2011) 123005, [arXiv:1110.0006].

[877] C. Gordon and O. Macias, *Dark Matter and Pulsar Model Constraints from Galactic Center Fermi-LAT Gamma Ray Observations*, *Phys. Rev. D* **88** (2013), no. 8 083521, [arXiv:1306.5725]. [Erratum: Phys.Rev.D 89, 049901 (2014)].

[878] K. N. Abazajian, N. Canac, S. Horiuchi, and M. Kaplinghat, *Astrophysical and Dark Matter Interpretations of Extended Gamma-Ray Emission from the Galactic Center*, *Phys. Rev. D* **90** (2014), no. 2 023526, [arXiv:1402.4090].

[879] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, *The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter*, *Phys. Dark Univ.* **12** (2016) 1–23, [arXiv:1402.6703].

[880] F. Calore, I. Cholis, and C. Weniger, *Background Model Systematics for the Fermi GeV Excess*, *JCAP* **03** (2015) 038, [arXiv:1409.0042].

[881] **AMS** Collaboration, M. Aguilar et al., *Antiproton Flux, Antiproton-to-Proton Flux Ratio, and Properties of Elementary Particle Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **117** (2016), no. 9 091103.

[882] A. Cuoco, J. Heisig, M. Korsmeier, and M. Krämer, *Probing dark matter annihilation in the Galaxy with antiprotons and gamma rays*, *JCAP* **10** (2017) 053, [arXiv:1704.08258].

[883] A. Cuoco, J. Heisig, L. Klamt, M. Korsmeier, and M. Krämer, *Scrutinizing the evidence for dark matter in cosmic-ray antiprotons*, *Phys. Rev. D* **99** (2019), no. 10 103014, [arXiv:1903.01472].

[884] I. Cholis, T. Linden, and D. Hooper, *A Robust Excess in the Cosmic-Ray Antiproton Spectrum: Implications for Annihilating Dark Matter*, *Phys. Rev. D* **99** (2019), no. 10 103026, [arXiv:1903.02549].

[885] D. Hooper, R. K. Leane, Y.-D. Tsai, S. Wegsman, and S. J. Witte, *A Systematic Study of Hidden Sector Dark Matter: Application to the Gamma-Ray and Antiproton Excesses*, arXiv:1912.08821.

[886] I. Cholis, D. Hooper, and T. Linden, *A New Determination of the Spectra and Luminosity Function of Gamma-Ray Millisecond Pulsars*, arXiv:1407.5583.

[887] I. Cholis, D. Hooper, and T. Linden, *Challenges in Explaining the Galactic Center Gamma-Ray Excess with Millisecond Pulsars*, *JCAP* **06** (2015) 043, [arXiv:1407.5625].

[888] S. K. Lee, M. Lisanti, and B. R. Safdi, *Distinguishing Dark Matter from Unresolved Point Sources in the Inner Galaxy with Photon Statistics*, *JCAP* **05** (2015) 056, [arXiv:1412.6099].

[889] R. Bartels, S. Krishnamurthy, and C. Weniger, *Strong support for the millisecond pulsar origin of the Galactic center GeV excess*, *Phys. Rev. Lett.* **116** (2016), no. 5 051102, [arXiv:1506.05104].

[890] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, *Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy*, *Phys. Rev. Lett.* **116** (2016), no. 5 051103, [arXiv:1506.05124].

[891] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi, and M. Pohl, *Galactic bulge preferred over dark matter for the Galactic centre gamma-ray excess*, *Nature Astron.* **2** (2018), no. 5 387–392, [arXiv:1611.06644].

[892] D. Haggard, C. Heinke, D. Hooper, and T. Linden, *Low Mass X-Ray Binaries in the Inner Galaxy: Implications for Millisecond Pulsars and the GeV Excess*, *JCAP* **05** (2017) 056, [arXiv:1701.02726].

[893] R. Bartels, E. Storm, C. Weniger, and F. Calore, *The Fermi-LAT GeV excess as a tracer of stellar mass in the Galactic bulge*, *Nature Astron.* **2** (2018), no. 10 819–828, [arXiv:1711.04778].

[894] O. Macias, S. Horiuchi, M. Kaplinghat, C. Gordon, R. M. Crocker, and D. M. Nataf, *Strong Evidence that the Galactic Bulge is Shining in Gamma Rays*, *JCAP* **09** (2019) 042, [arXiv:1901.03822].

[895] R. K. Leane and T. R. Slatyer, *Revival of the Dark Matter Hypothesis for the Galactic Center Gamma-Ray Excess*, *Phys. Rev. Lett.* **123** (2019), no. 24 241101, [arXiv:1904.08430].

[896] Y.-M. Zhong, S. D. McDermott, I. Cholis, and P. J. Fox, *A New Mask for An Old Suspect: Testing the Sensitivity of the Galactic Center Excess to the Point Source Mask*, *Phys. Rev. Lett.* **124** (2020), no. 23 231103, [arXiv:1911.12369].

[897] R. K. Leane and T. R. Slatyer, *Spurious Point Source Signals in the Galactic Center Excess*, arXiv:2002.12370.

[898] R. K. Leane and T. R. Slatyer, *The Enigmatic Galactic Center Excess: Spurious Point Sources and Signal Mismodeling*, arXiv:2002.12371.

[899] M. Buschmann, N. L. Rodd, B. R. Safdi, L. J. Chang, S. Mishra-Sharma, M. Lisanti, and O. Macias, *Foreground Mismodeling and the Point Source Explanation of the Fermi Galactic Center Excess*, arXiv:2002.12373.

[900] K. N. Abazajian, S. Horiuchi, M. Kaplinghat, R. E. Keeley, and O. Macias, *Strong constraints on thermal relic dark matter from Fermi-LAT observations of the Galactic Center*, arXiv:2003.10416.

[901] F. List, N. L. Rodd, G. F. Lewis, and I. Bhat, *The GCE in a New Light: Disentangling the γ-ray Sky with Bayesian Graph Convolutional Neural Networks*, arXiv:2006.12504.

[902] S. Mishra-Sharma and K. Cranmer, *Semi-parametric γ-ray modeling with Gaussian processes and variational inference*, arXiv:2010.10450.

[903] C. Karwin, S. Murgia, T. M. P. Tait, T. A. Porter, and P. Tanedo, *Dark Matter Interpretation of the Fermi-LAT Observation Toward the Galactic Center*, *Phys. Rev. D* **95** (2017), no. 10 103005, [arXiv:1612.05687].

[904] W.-C. Huang, A. Urbano, and W. Xue, *Fermi Bubbles under Dark Matter Scrutiny Part II: Particle Physics Analysis*, *JCAP* **04** (2014) 020, [arXiv:1310.7609].

[905] C. Boehm, M. J. Dolan, C. McCabe, M. Spannowsky, and C. J. Wallace, *Extended gamma-ray emission from Coy Dark Matter*, *JCAP* **05** (2014) 009, [arXiv:1401.6458].

[906] C. Cheung, M. Papucci, D. Sanford, N. R. Shah, and K. M. Zurek, *NMSSM Interpretation of the Galactic Center Excess*, *Phys. Rev.* **D90** (2014), no. 7 075011, [arXiv:1406.6372].

[907] J. Guo, J. Li, T. Li, and A. G. Williams, *NMSSM explanations of the Galactic center gamma ray excess and promising LHC searches*, *Phys. Rev. D* **91** (2015), no. 9 095003, [arXiv:1409.7864].

[908] J. Cao, L. Shang, P. Wu, J. M. Yang, and Y. Zhang, *Supersymmetry explanation of the Fermi Galactic Center excess and its test at LHC run II*, *Phys. Rev. D* **91** (2015), no. 5 055005, [arXiv:1410.3239].

[909] A. Berlin, S. Gori, T. Lin, and L.-T. Wang, *Pseudoscalar Portal Dark Matter*, *Phys. Rev. D* **92** (2015) 015005, [arXiv:1502.06000].

[910] T. Gherghetta, B. von Harling, A. D. Medina, M. A. Schmidt, and T. Trott, *SUSY implications from WIMP annihilation into scalars at the Galactic Center*, *Phys. Rev.* **D91** (2015) 105004, [arXiv:1502.07173].

[911] M. Duerr, P. Fileviez Pérez, and J. Smirnov, *Gamma-Ray Excess and the Minimal Dark Matter Model*, *JHEP* **06** (2016) 008, [arXiv:1510.07562].

[912] M. Carena, J. Osborne, N. R. Shah, and C. E. M. Wagner, *Return of the WIMP: Missing energy signals and the Galactic Center excess*, *Phys. Rev.* **D100** (2019), no. 5 055002, [arXiv:1905.03768].

[913] R. Mahbubani and L. Senatore, *The Minimal model for dark matter and unification*, *Phys. Rev. D* **73** (2006) 043510, [hep-ph/0510064].

[914] F. D'Eramo, *Dark matter and Higgs boson physics*, *Phys. Rev. D* **76** (2007) 083522, [arXiv:0705.4493].

[915] R. Enberg, P. Fox, L. Hall, A. Papaioannou, and M. Papucci, *Lhc and dark matter signals of improved naturalness*, *Journal of High Energy Physics* **2007** (Nov, 2007) 014–014.

[916] T. Cohen, J. Kearney, A. Pierce, and D. Tucker-Smith, *Singlet-Doublet Dark Matter*, *Phys. Rev. D* **85** (2012) 075003, [arXiv:1109.2604].

[917] C. Cheung and D. Sanford, *Simplified Models of Mixed Dark Matter*, *JCAP* **02** (2014) 011, [arXiv:1311.5896].

[918] T. Abe, R. Kitano, and R. Sato, *Discrimination of dark matter models in future experiments*, *Phys. Rev. D* **91** (2015), no. 9 095004, [arXiv:1411.1335]. [Erratum: Phys.Rev.D 96, 019902 (2017)].

[919] L. Calibbi, A. Mariotti, and P. Tziveloglou, *Singlet-Doublet Model: Dark matter searches and LHC constraints*, *JHEP* **10** (2015) 116, [arXiv:1505.03867].

[920] A. Freitas, S. Westhoff, and J. Zupan, *Integrating in the Higgs Portal to Fermion Dark Matter*, *JHEP* **09** (2015) 015, [arXiv:1506.04149].

[921] S. Banerjee, S. Matsumoto, K. Mukaida, and Y.-L. S. Tsai, *WIMP Dark Matter in a Well-Tempered Regime: A case study on Singlet-Doublets Fermionic WIMP*, *JHEP* **11** (2016) 070, [arXiv:1603.07387].

[922] C. Cai, Z.-H. Yu, and H.-H. Zhang, *Cepc precision of electroweak oblique parameters and weakly interacting dark matter: The fermionic case*, *Nuclear Physics B* **921** (Aug, 2017) 181–210.

[923] L. Lopez Honorez, M. H. G. Tytgat, P. Tziveloglou, and B. Zaldivar, *On Minimal Dark Matter coupled to the Higgs*, *JHEP* **04** (2018) 011, [arXiv:1711.08619].

[924] A. Dedes and D. Karamitros, *Doublet-Triplet Fermionic Dark Matter*, *Phys. Rev. D* **89** (2014), no. 11 115002, [arXiv:1403.7744].

[925] Y. Zeldovich, *Survey of Modern Cosmology*, vol. 3, pp. 241–379. 1965.

[926] H.-Y. Chiu, *Symmetry between particle and anti-particle populations in the universe*, *Phys. Rev. Lett.* **17** (1966) 712.

[927] B. W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino Masses*, *Phys. Rev. Lett.* **39** (1977) 165–168.

[928] P. Hut, *Limits on Masses and Number of Neutral Weakly Interacting Particles*, *Phys. Lett. B* **69** (1977) 85.

[929] S. Wolfram, *Abundances of Stable Particles Produced in the Early Universe*, *Phys. Lett. B* **82** (1979) 65–68.

[930] G. Steigman, *Cosmology Confronts Particle Physics*, *Ann. Rev. Nucl. Part. Sci.* **29** (1979) 313–338.

[931] R. J. Scherrer and M. S. Turner, *On the Relic, Cosmic Abundance of Stable Weakly Interacting Massive Particles*, *Phys. Rev. D* **33** (1986) 1585. [Erratum: Phys.Rev.D 34, 3263 (1986)].

[932] J. Bernstein, L. S. Brown, and G. Feinberg, *Cosmological heavy-neutrino problem*, *Phys. Rev. D* **32** (Dec, 1985) 3261–3267.

[933] M. Srednicki, R. Watkins, and K. A. Olive, *Calculations of Relic Densities in the Early Universe*, *Nucl. Phys. B* **310** (1988) 693.

[934] P. Gondolo and G. Gelmini, *Cosmic abundances of stable particles: Improved analysis*, *Nucl. Phys. B* **360** (1991) 145–179.

[935] G. Steigman, B. Dasgupta, and J. F. Beacom, *Precise Relic WIMP Abundance and its Impact on Searches for Dark Matter Annihilation*, *Phys. Rev. D* **86** (2012) 023506, [arXiv:1204.3622].

[936] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, *Early kinetic decoupling of dark matter: when the standard way of calculating the thermal relic density fails*, *Phys. Rev. D* **96** (2017), no. 11 115010, [arXiv:1706.07433]. [Erratum: Phys.Rev.D 101, 099901 (2020)].

[937] T. Abe, *The effect of the early kinetic decoupling in a fermionic dark matter model*, arXiv:2004.10041.

[938] M. Di Mauro, *The characteristics of the Galactic center excess measured with 11 years of Fermi-LAT data*, arXiv:2101.04694.

[939] M. Cirelli, E. Del Nobile, and P. Panci, *Tools for model-independent bounds in direct dark matter searches*, *JCAP* **10** (2013) 019, [arXiv:1307.5955].

[940] R. J. Hill and M. P. Solon, *Standard Model anatomy of WIMP dark matter direct detection II: QCD analysis and hadronic matrix elements*, *Phys. Rev. D* **91** (2015) 043505, [arXiv:1409.8290].

[941] F. Bishara, J. Brod, B. Grinstein, and J. Zupan, *From quarks to nucleons in dark matter direct detection*, *JHEP* **11** (2017) 059, [arXiv:1707.06998].

[942] J. Ellis, N. Nagata, and K. A. Olive, *Uncertainties in WIMP Dark Matter Scattering Revisited*, *Eur. Phys. J. C* **78** (2018), no. 7 569, [arXiv:1805.09795].

[943] **LUX-ZEPLIN** Collaboration, D. Akerib et al., *Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment*, *Phys. Rev. D* **101** (2020), no. 5 052002, [arXiv:1802.06039].

[944] **XENON** Collaboration, E. Aprile et al., *Projected WIMP Sensitivity of the XENONnT Dark Matter Experiment*, arXiv:2007.08796.

[945] **IceCube** Collaboration, M. Aartsen et al., *Search for annihilating dark matter in the Sun with 3 years of IceCube data*, *Eur. Phys. J. C* **77** (2017), no. 3 146, [arXiv:1612.05949]. [Erratum: Eur.Phys.J.C 79, 214 (2019)].

[946] **LUX** Collaboration, D. Akerib et al., *Results on the Spin-Dependent Scattering of Weakly Interacting Massive Particles on Nucleons from the Run 3 Data of the LUX Experiment*, *Phys. Rev. Lett.* **116** (2016), no. 16 161302, [arXiv:1602.03489].

[947] **LUX** Collaboration, D. Akerib et al., *Results from a search for dark matter in the complete LUX exposure*, *Phys. Rev. Lett.* **118** (2017), no. 2 021303, [arXiv:1608.07648].

[948] **PandaX-II** Collaboration, X. Cui et al., *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, *Phys. Rev. Lett.* **119** (2017), no. 18 181302, [arXiv:1708.06917].

[949] M. Carena, J. Osborne, N. R. Shah, and C. E. M. Wagner, *Supersymmetry and LHC Missing Energy Signals*, *Phys. Rev.* **D98** (2018), no. 11 115010, [arXiv:1809.11082].

[950] **LEPSUSYWG, ALEPH, DELPHI, L3 and OPAL** Collaboration, LEPSUSYWG/02-04.1, *Combined lep chargino results, up to 208 gev for large m0*, http://lepsusy.web.cern.ch/lepsusy/www/inos_moriond01/charginos_pub.html.

[951] **LEPSUSYWG, ALEPH, DELPHI, L3 and OPAL** Collaboration, LEPSUSYWG/01-03.1, *Combined lep chargino results, up to 208 gev for low dm*, https://lepsusy.web.cern.ch/lepsusy/www/inoslowdmsummer02/charginolowdm_pub.html.

[952] **ACME** Collaboration, V. Andreev et al., *Improved limit on the electric dipole moment of the electron*, *Nature* **562** (2018), no. 7727 355–360.

[953] S. M. Barr and A. Zee, *Electric Dipole Moment of the Electron and of the Neutron*, *Phys. Rev. Lett.* **65** (1990) 21–24. [Erratum: Phys. Rev. Lett.65,2920(1990)].

[954] D. Atwood, C. P. Burgess, C. Hamazaou, B. Irwin, and J. A. Robinson, *One loop P and T odd W+- electromagnetic moments*, *Phys. Rev.* **D42** (1990) 3770–3777.

[955] M. E. Peskin and T. Takeuchi, *Estimation of oblique electroweak corrections*, *Phys. Rev. D* **46** (1992) 381–409.

[956] G. Cacciapaglia, C. Csaki, G. Marandella, and A. Strumia, *The Minimal Set of Electroweak Precision Parameters*, *Phys. Rev. D* **74** (2006) 033011, [hep-ph/0604111].

[957] **Particle Data Group** Collaboration, P. Zyla et al., *Review of Particle Physics*, *Progress of Theoretical and Experimental Physics* **2020** (08, 2020) [https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf]. 083C01.

[958] R. Barbieri, A. Pomarol, R. Rattazzi, and A. Strumia, *Electroweak symmetry breaking after LEP-1 and LEP-2*, *Nucl. Phys. B* **703** (2004) 127–146, [hep-ph/0405040].

[959] C. K. Khosa, S. Kraml, A. Lessa, P. Neuhuber, and W. Waltenberger, *SModelS database update v1.2.3, to appear in LHEP.* (2020) [arXiv:2005.00555].

[960] F. Ambrogi et al., *SModelS v1.2: long-lived particles, combination of signal regions, and other novelties*, arXiv:1811.10624.

[961] J. Dutta, S. Kraml, A. Lessa, and W. Waltenberger, *SModelS extension with the CMS supersymmetry search results from Run 2*, *LHEP* **1** (2018), no. 1 5–12, [arXiv:1803.02204].

[962] F. Ambrogi, S. Kraml, S. Kulkarni, U. Laa, A. Lessa, V. Magerl, J. Sonneveld, M. Traub, and W. Waltenberger, *SModelS v1.1 user manual*, arXiv:1701.06586.

[963] S. Kraml, S. Kulkarni, U. Laa, A. Lessa, W. Magerl, D. Proschofsky, and W. Waltenberger, *SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry*, *Eur.Phys.J.* **C74** (2014) 2868, [arXiv:1312.4175].

[964] ATLAS Collaboration, *Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods*, Tech. Rep. ATL-PHYS-PUB-2019-029, CERN, Geneva, Aug, 2019. https://cds.cern.ch/record/2684863.

[965] P. Z. Skands, B. Allanach, H. Baer, C. Balazs, G. Belanger, et al., *SUSY Les Houches accord: Interfacing SUSY spectrum calculators, decay packages, and event generators*, *JHEP* **0407** (2004) 036, [hep-ph/0311123].

[966] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, et al., *A Standard format for Les Houches event files*, *Comput.Phys.Commun.* **176** (2007) 300–304, [hep-ph/0609017].

[967] A. Buckley, *PySLHA: a Pythonic interface to SUSY Les Houches Accord data*, arXiv:1305.4194.

[968] F. Staub, *SARAH*, arXiv:0806.0538.

[969] F. Staub, *SARAH 4 : A tool for (not only SUSY) model builders*, *Comput. Phys. Commun.* **185** (2014) 1773–1790, [arXiv:1309.7223].

[970] F. Staub, *Exploring new models in all detail with SARAH*, *Adv. High Energy Phys.* **2015** (2015) 840780, [arXiv:1503.04200].

[971] W. Porod, *SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e+ e- colliders*, *Comput. Phys. Commun.* **153** (2003) 275–315, [hep-ph/0301101].

[972] W. Porod and F. Staub, *SPheno 3.1: Extensions including flavour, CP-phases and models beyond the MSSM*, *Comput. Phys. Commun.* **183** (2012) 2458–2469, [arXiv:1104.1573].

[973] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301].

[974] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [arXiv:1106.0522].

[975] **CMS** Collaboration, A. M. Sirunyan et al., *Measurements of the Higgs boson width and anomalous* HVV *couplings from on-shell and off-shell production in the four-lepton final state*, *Phys. Rev. D* **99** (2019), no. 11 112003, [arXiv:1901.00174].

[976] Y. Nakai and M. Reece, *Electric Dipole Moments in Natural Supersymmetry*, *JHEP* **08** (2017) 031, [arXiv:1612.08090].

[977] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1–29, [arXiv:1207.7214].

[978] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30–61, [arXiv:1207.7235].

[979] **Muon g-2** Collaboration, T. Albahri et al., *Measurement of the anomalous precession frequency of the muon in the Fermilab Muon g-2 Experiment*, *Phys. Rev. D* **103** (2021), no. 7 072002, [arXiv:2104.03247].

[980] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, *JHEP* **08** (2017) 055, [arXiv:1705.05802].

[981] **LHCb** Collaboration, R. Aaij et al., *Search for lepton-universality violation in* $B^+ \to K^+\ell^+\ell^-$ *decays*, *Phys. Rev. Lett.* **122** (2019), no. 19 191801, [arXiv:1903.09252].

[982] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality in beauty-quark decays*, *Nature Phys.* **18** (2022), no. 3 277–282, [arXiv:2103.11769].

[983] H. Flacher, M. Goebel, J. Haller, A. Hocker, K. Monig, and J. Stelzer, *Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter*, *Eur. Phys. J. C* **60** (2009) 543–583, [arXiv:0811.0009]. [Erratum: Eur.Phys.J.C 71, 1718 (2011)].

[984] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Ludwig, K. Moenig, M. Schott, and J. Stelzer, *Updated Status of the Global Electroweak Fit and Constraints on New Physics*, *Eur. Phys. J. C* **72** (2012) 2003, [arXiv:1107.0975].

[985] **Gfitter Group** Collaboration, M. Baak, J. Cúth, J. Haller, A. Hoecker, R. Kogler, K. Mönig, M. Schott, and J. Stelzer, *The global electroweak fit at NNLO and prospects for the LHC and ILC*, *Eur. Phys. J. C* **74** (2014) 3046, [arXiv:1407.3792].

[986] M. E. Peskin and T. Takeuchi, *A New constraint on a strongly interacting Higgs sector*, *Phys. Rev. Lett.* **65** (1990) 964–967.

[987] D. C. Kennedy and B. W. Lynn, *Electroweak Radiative Corrections with an Effective Lagrangian: Four Fermion Processes*, *Nucl. Phys. B* **322** (1989) 1–54.

[988] B. Holdom and J. Terning, *Large corrections to electroweak parameters in technicolor theories*, *Phys. Lett. B* **247** (1990) 88–92.

[989] M. Golden and L. Randall, *Radiative Corrections to Electroweak Parameters in Technicolor Theories*, *Nucl. Phys. B* **361** (1991) 3–23.

[990] Z. Han and W. Skiba, *Effective theory analysis of precision electroweak data*, *Phys. Rev. D* **71** (2005) 075009, [hep-ph/0412166].

[991] Z. Han, *Effective Theories and Electroweak Precision Constraints*, *Int. J. Mod. Phys. A* **23** (2008) 2653–2685, [arXiv:0807.0490].

[992] **Particle Data Group** Collaboration, R. L. Workman and Others, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[993] M. Ciuchini, E. Franco, S. Mishima, and L. Silvestrini, *Electroweak Precision Observables, New Physics and the Nature of a 126 GeV Higgs Boson*, *JHEP* **08** (2013) 106, [arXiv:1306.4644].

[994] J. de Blas, M. Ciuchini, E. Franco, S. Mishima, M. Pierini, L. Reina, and L. Silvestrini, *Electroweak precision observables and Higgs-boson signal strengths in the Standard Model and beyond: present and future*, *JHEP* **12** (2016) 135, [arXiv:1608.01509].

[995] **CDF** Collaboration, T. A. Aaltonen et al., *Precise Measurement of the W -Boson Mass with the Collider Detector at Fermilab*, *Phys. Rev. D* **89** (2014), no. 7 072003, [arXiv:1311.0894].

[996] **CDF, D0** Collaboration, T. A. Aaltonen et al., *Combination of CDF and D0* W*-Boson Mass Measurements*, *Phys. Rev. D* **88** (2013), no. 5 052018, [arXiv:1307.7627].

[997] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the* W*-boson mass in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, *Eur. Phys. J. C* **78** (2018), no. 2 110, [arXiv:1701.07240]. [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[998] **LHCb** Collaboration, R. Aaij et al., *Measurement of the W boson mass*, *JHEP* **01** (2022) 036, [arXiv:2109.01113].

[999] J. Haller, A. Hoecker, R. Kogler, K. Mönig, T. Peiffer, and J. Stelzer, *Update of the global electroweak fit and constraints on two-Higgs-doublet models*, *Eur. Phys. J. C* **78** (2018), no. 8 675, [arXiv:1803.01853].

[1000] P. Janot and S. Jadach, *Improved Bhabha cross section at LEP and the number of light neutrino species*, *Phys. Lett. B* **803** (2020) 135319, [arXiv:1912.02067].

[1001] J. Fan, M. Reece, and L.-T. Wang, *Possible Futures of Electroweak Precision: ILC, FCC-ee, and CEPC*, *JHEP* **09** (2015) 196, [arXiv:1411.1054].

[1002] J. Fan, M. Reece, and L.-T. Wang, *Precision Natural SUSY at CEPC, FCC-ee, and ILC*, *JHEP* **08** (2015) 152, [arXiv:1412.3107].

[1003] **ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group** Collaboration, S. Schael et al., *Precision electroweak measurements on the* Z *resonance*, *Phys. Rept.* **427** (2006) 257–454, [hep-ex/0509008].

[1004] M. Steinhauser, *Leptonic contribution to the effective electromagnetic coupling constant up to three loops*, *Phys. Lett. B* **429** (1998) 158–161, [hep-ph/9803313].

[1005] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, *Precise prediction for the W boson mass in the standard model*, *Phys. Rev. D* **69** (2004) 053006, [hep-ph/0311148].

[1006] G.-C. Cho, K. Hagiwara, Y. Matsumoto, and D. Nomura, *The MSSM confronts the precision electroweak data and the muon g-2*, *JHEP* **11** (2011) 068, [arXiv:1104.1769].

[1007] I. Dubovyk, A. Freitas, J. Gluza, T. Riemann, and J. Usovitsch, *Complete electroweak two-loop corrections to Z boson production and decay*, *Phys. Lett. B* **783** (2018) 86–94, [arXiv:1804.10236].

[1008] M. Awramik, M. Czakon, and A. Freitas, *Electroweak two-loop corrections to the effective weak mixing angle*, *JHEP* **11** (2006) 048, [hep-ph/0608099].

[1009] I. Dubovyk, A. Freitas, J. Gluza, T. Riemann, and J. Usovitsch, *The two-loop electroweak bosonic corrections to* $\sin^2\theta^b_{eff}$, *Phys. Lett. B* **762** (2016) 184–189, [arXiv:1607.08375].

[1010] J. Erler and M. Schott, *Electroweak Precision Tests of the Standard Model after the Discovery of the Higgs Boson*, *Prog. Part. Nucl. Phys.* **106** (2019) 68–119, [arXiv:1902.05142].

[1011] P. L. Cho and E. H. Simmons, *Searching for G3 in* $t\bar{t}$ *production*, *Phys. Rev. D* **51** (1995) 2360–2370, [hep-ph/9408206].

[1012] I. Maksymyk, C. P. Burgess, and D. London, *Beyond S, T and U*, *Phys. Rev. D* **50** (1994) 529–535, [hep-ph/9306267].

[1013] C. P. Burgess, S. Godfrey, H. Konig, D. London, and I. Maksymyk, *A Global fit to extended oblique parameters*, *Phys. Lett. B* **326** (1994) 276–281, [hep-ph/9307337].

[1014] C. P. Burgess, S. Godfrey, H. Konig, D. London, and I. Maksymyk, *Model independent global constraints on new physics*, *Phys. Rev. D* **49** (1994) 6115–6147, [hep-ph/9312291].

[1015] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [arXiv:1008.4884].

[1016] J. D. Wells and Z. Zhang, *Effective theories of universal theories*, *JHEP* **01** (2016) 123, [arXiv:1510.08462].

[1017] B. Grinstein and M. B. Wise, *Operator analysis for precision electroweak physics*, *Phys. Lett. B* **265** (1991) 326–334.

[1018] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1–102, [arXiv:1106.0034].

[1019] T. Blank and W. Hollik, *Precision observables in SU(2) x U(1) models with an additional Higgs triplet*, *Nucl. Phys. B* **514** (1998) 113–134, [hep-ph/9703392].

[1020] P. J. Fox, A. E. Nelson, and N. Weiner, *Dirac gaugino masses and supersoft supersymmetry breaking*, *JHEP* **08** (2002) 035, [hep-ph/0206096].

[1021] G. D. Kribs, E. Poppitz, and N. Weiner, *Flavor in supersymmetry with an extended R-symmetry*, *Phys. Rev. D* **78** (2008) 055010, [arXiv:0712.2039].

[1022] P. Bandyopadhyay and A. Costantini, *Obscure Higgs boson at Colliders*, *Phys. Rev. D* **103** (2021), no. 1 015025, [arXiv:2010.02597].

[1023] Z. U. Khandker, D. Li, and W. Skiba, *Electroweak Corrections from Triplet Scalars*, *Phys. Rev. D* **86** (2012) 015006, [arXiv:1201.4383].

[1024] B. Henning, X. Lu, and H. Murayama, *How to use the Standard Model effective field theory*, *JHEP* **01** (2016) 023, [arXiv:1412.1837].

[1025] M. Passera, W. J. Marciano, and A. Sirlin, *The Muon g-2 and the bounds on the Higgs boson mass*, *Phys. Rev. D* **78** (2008) 013009, [arXiv:0804.1142].

[1026] A. Keshavarzi, W. J. Marciano, M. Passera, and A. Sirlin, *Muon* $g-2$ *and* $\Delta\alpha$ *connection*, *Phys. Rev. D* **102** (2020), no. 3 033002, [arXiv:2006.12666].

[1027] A. Crivellin, M. Hoferichter, C. A. Manzari, and M. Montull, *Hadronic Vacuum Polarization:* $(g-2)_\mu$ *versus Global Electroweak Fits*, *Phys. Rev. Lett.* **125** (2020), no. 9 091801, [arXiv:2003.04886].

[1028] P. Athron, A. Fowlie, C.-T. Lu, L. Wu, Y. Wu, and B. Zhu, *The* W *boson Mass and Muon* $g-2$*: Hadronic Uncertainties or New Physics?*, arXiv:2204.03996.

[1029] J. M. Yang and Y. Zhang, *Low energy SUSY confronted with new measurements of W-boson mass and muon g-2*, arXiv:2204.04202.

[1030] D. Choudhury, T. M. P. Tait, and C. E. M. Wagner, *Beautiful mirrors and precision electroweak data*, *Phys. Rev. D* **65** (2002) 053002, [hep-ph/0109097].

[1031] Y.-Z. Fan, T.-P. Tang, Y.-L. S. Tsai, and L. Wu, *Inert Higgs Dark Matter for New CDF W-boson Mass and Detection Prospects*, arXiv:2204.03693.

[1032] C.-T. Lu, L. Wu, Y. Wu, and B. Zhu, *Electroweak Precision Fit and New Physics in light of* W *Boson Mass*, arXiv:2204.03796.

[1033] G.-W. Yuan, L. Zu, L. Feng, and Y.-F. Cai, W*-boson mass anomaly: probing the models of axion-like particle, dark photon and Chameleon dark energy*, arXiv:2204.04183.

[1034] J. de Blas, M. Pierini, L. Reina, and L. Silvestrini, *Impact of the recent measurements of the top-quark and W-boson masses on electroweak precision fits*, arXiv:2204.04204.

[1035] C.-R. Zhu, M.-Y. Cui, Z.-Q. Xia, Z.-H. Yu, X. Huang, Q. Yuan, and Y. Z. Fan, *GeV antiproton/gamma-ray excesses and the* W*-boson mass anomaly: three faces of* $\sim 60-70$ *GeV dark matter particle?*, arXiv:2204.03767.

[1036] J. D. Bjorken and S. L. Glashow, *Elementary Particles and SU(4)*, *Phys. Lett.* **11** (1964) 255–257.

[1037] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, *Phys. Rev. D* **2** (1970) 1285–1292.

[1038] B. A. Campbell and P. J. O'Donnell, *Mass of the Top Quark and Induced Decay and Neutral Mixing of B Mesons*, *Phys. Rev. D* **25** (1982) 1989.

[1039] M. A. Shifman, *Theoretical Status of Weak Decays*, *Nucl. Phys. B Proc. Suppl.* **3** (1988) 289.

[1040] J. R. Ellis, J. S. Hagelin, S. Rudaz, and D. D. Wu, *Implications of recent measurements of* B *meson mixing and* $\varepsilon'/\varepsilon_{\mathrm{K}}$, *Nucl. Phys. B* **304** (1988) 205–235.

[1041] A. D. Martin, *Top and Bottom Physics: The K-M Matrix and CP Violation*, *J. Phys. G* **15** (1989) 1073.

[1042] **CDF** Collaboration, F. Abe et al., *Observation of top quark production in* $\bar{\mathrm{p}}\mathrm{p}$ *collisions*, *Phys. Rev. Lett.* **74** (1995) 2626–2631, [hep-ex/9503002].

[1043] **D0** Collaboration, S. Abachi et al., *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632–2637, [hep-ex/9503003].

[1044] R. K. Ellis et al., *Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020*, arXiv:1910.11775.

[1045] M. Artuso et al., *Report of the Frontier For Rare Processes and Precision Measurements*, arXiv:2210.04765.

[1046] C. D. Froggatt and H. B. Nielsen, *Hierarchy of Quark Masses, Cabibbo Angles and CP Violation*, *Nucl. Phys. B* **147** (1979) 277–298.

[1047] M. Leurer, Y. Nir, and N. Seiberg, *Mass matrix models*, *Nucl. Phys. B* **398** (1993) 319–342, [hep-ph/9212278].

[1048] M. Leurer, Y. Nir, and N. Seiberg, *Mass matrix models: The Sequel*, *Nucl. Phys. B* **420** (1994) 468–504, [hep-ph/9310320].

[1049] P. Pouliot and N. Seiberg, *(S)quark masses and nonAbelian horizontal symmetries*, *Phys. Lett. B* **318** (1993) 169–173, [hep-ph/9308363].

[1050] V. Jain and R. Shrock, *Models of fermion mass matrices based on a flavor dependent and generation dependent U(1) gauge symmetry*, *Phys. Lett. B* **352** (1995) 83–91, [hep-ph/9412367].

[1051] N. Arkani-Hamed and M. Schmaltz, *Hierarchies without symmetries from extra dimensions*, *Phys. Rev. D* **61** (2000) 033005, [hep-ph/9903417].

[1052] T. Gherghetta and A. Pomarol, *Bulk fields and supersymmetry in a slice of AdS*, *Nucl. Phys. B* **586** (2000) 141–162, [hep-ph/0003129].

[1053] S. J. Huber and Q. Shafi, *Fermion masses, mixings and proton decay in a Randall-Sundrum model*, *Phys. Lett. B* **498** (2001) 256–262, [hep-ph/0010195].

[1054] D. E. Kaplan and T. M. P. Tait, *New tools for fermion masses from extra dimensions*, *JHEP* **11** (2001) 051, [hep-ph/0110126].

[1055] A. Ahmed, A. Carmona, J. Castellano Ruiz, Y. Chung, and M. Neubert, *Dynamical origin of fermion bulk masses in a warped extra dimension*, *JHEP* **08** (2019) 045, [arXiv:1905.09833].

[1056] S. Girmohanta, R. N. Mohapatra, and R. Shrock, *Neutrino Masses and Mixing in Models with Large Extra Dimensions and Localized Fermions*, *Phys. Rev. D* **103** (2021), no. 1 015021, [arXiv:2011.01237].

[1057] A. E. Nelson and M. J. Strassler, *Suppressing flavor anarchy*, *JHEP* **09** (2000) 030, [hep-ph/0006251].

[1058] S. Weinberg, *Electromagnetic and weak masses*, *Phys. Rev. Lett.* **29** (1972) 388–392.

[1059] H. Georgi and S. L. Glashow, *Attempts to calculate the electron mass*, *Phys. Rev. D* **7** (1973) 2457–2463.

[1060] S. M. Barr and A. Zee, *A New Approach to the electron-Muon Mass Ratio*, *Phys. Rev. D* **15** (1977) 2652.

[1061] B. S. Balakrishna, A. L. Kagan, and R. N. Mohapatra, *Quark Mixings and Mass Hierarchy From Radiative Corrections*, *Phys. Lett. B* **205** (1988) 345–352.

[1062] K. S. Babu, *TASI Lectures on Flavor Physics*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: The Dawn of the LHC Era*, pp. 49–123, 2010. arXiv:0910.2948.

[1063] F. Feruglio, *Pieces of the Flavour Puzzle*, *Eur. Phys. J. C* **75** (2015), no. 8 373, [arXiv:1503.04071].

[1064] F. Feruglio and A. Romanino, *Lepton flavor symmetries*, *Rev. Mod. Phys.* **93** (2021), no. 1 015007, [arXiv:1912.06028].

[1065] A. Smolkovi**v**c, M. Tammaro, and J. Zupan, *Anomaly free Froggatt-Nielsen models of flavor*, *JHEP* **10** (2019) 188, [arXiv:1907.10063]. [Erratum: JHEP 02, 033 (2022)].

[1066] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, *Nucl. Phys. B* **645** (2002) 155–187, [hep-ph/0207036].

[1067] D. Egana-Ugrinovic, S. Homiller, and P. Meade, *Aligned and Spontaneous Flavor Violation*, *Phys. Rev. Lett.* **123** (2019), no. 3 031802, [arXiv:1811.00017].

[1068] Y. Nir and N. Seiberg, *Should squarks be degenerate?*, *Phys. Lett. B* **309** (1993) 337–343, [hep-ph/9304307].

[1069] T. Feldmann and T. Mannel, *Minimal Flavour Violation and Beyond*, *JHEP* **02** (2007) 067, [hep-ph/0611095].

[1070] M. Bordone, O. Catà, and T. Feldmann, *Effective Theory Approach to New Physics with Flavour: General Framework and a Leptoquark Example*, *JHEP* **01** (2020) 067, [arXiv:1910.02641].

[1071] M. Bordone, O. Catà, T. Feldmann, and R. Mandal, *Constraining flavour patterns of scalar leptoquarks in the effective field theory*, *JHEP* **03** (2021) 122, [arXiv:2010.03297].

[1072] I. Dor**v**sner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Ko**v**snik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1–68, [arXiv:1603.04993].

[1073] E. Baver and M. Leurer, *Naturally light leptoquarks*, *Phys. Rev. D* **51** (1995) 260–264, [hep-ph/9407324].

[1074] I. de Medeiros Varzielas and G. Hiller, *Clues for flavor from rare lepton and quark decays*, *JHEP* **06** (2015) 072, [arXiv:1503.01084].

[1075] G. Hiller, D. Loose, and K. Schönwald, *Leptoquark Flavor Patterns & B Decay Anomalies*, *JHEP* **12** (2016) 027, [arXiv:1609.08895].

[1076] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, *Phys. Rev. Lett.* **51** (1983) 1945.

[1077] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (1963) 531–533.

[1078] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.

[1079] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, *Zh. Eksp. Teor. Fiz.* **34** (1957) 247.

[1080] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, *Prog. Theor. Phys.* **28** (1962) 870–880.

[1081] D. Aloni, P. Asadi, Y. Nakai, M. Reece, and M. Suzuki, *Spontaneous CP violation and horizontal symmetry in the MSSM: toward lepton flavor naturalness*, *JHEP* **09** (2021) 031, [arXiv:2104.02679].

[1082] C. Cornella, D. Curtin, E. T. Neil, and J. O. Thompson, *Mapping and Probing Froggatt-Nielsen Solutions to the Quark Flavor Puzzle*, arXiv:2306.08026.

[1083] Y. Nir and R. Rattazzi, *Solving the supersymmetric CP problem with Abelian horizontal symmetries*, *Phys. Lett. B* **382** (1996) 363–368, [hep-ph/9603233].

[1084] Y. Nakai, M. Reece, and M. Suzuki, *Supersymmetric alignment models for (g - 2)$_\mu$*, *JHEP* **10** (2021) 068, [arXiv:2107.10268].

[1085] M. Fedele, A. Mastroddi, and M. Valli, *Minimal Froggatt-Nielsen textures*, *JHEP* **03** (2021) 135, [arXiv:2009.05587].

[1086] A. Davidson and K. C. Wali, *MINIMAL FLAVOR UNIFICATION VIA MULTIGENERATIONAL PECCEI-QUINN SYMMETRY*, *Phys. Rev. Lett.* **48** (1982) 11.

[1087] A. Davidson, V. P. Nair, and K. C. Wali, *Peccei-Quinn Symmetry as Flavor Symmetry and Grand Unification*, *Phys. Rev. D* **29** (1984) 1504.

[1088] F. Wilczek, *Axions and Family Symmetry Breaking*, *Phys. Rev. Lett.* **49** (1982) 1549–1552.

[1089] Y. Ema, K. Hamaguchi, T. Moroi, and K. Nakayama, *Flaxion: a minimal extension to solve puzzles in the standard model*, *JHEP* **01** (2017) 096, [arXiv:1612.05492].

[1090] L. Calibbi, F. Goertz, D. Redigolo, R. Ziegler, and J. Zupan, *Minimal axion model from flavor*, *Phys. Rev. D* **95** (2017), no. 9 095009, [arXiv:1612.08040].

[1091] Q. Bonnefoy, E. Dudas, and S. Pokorski, *Chiral Froggatt-Nielsen models, gauge anomalies and flavourful axions*, *JHEP* **01** (2020) 191, [arXiv:1909.05336].

[1092] B. C. Allanach, J. Davighi, and S. Melville, *An Anomaly-free Atlas: charting the space of flavour-dependent gauged* U(1) *extensions of the Standard Model*, *JHEP* **02** (2019) 082, [arXiv:1812.04602]. [Erratum: JHEP 08, 064 (2019)].

[1093] D. B. Costa, B. A. Dobrescu, and P. J. Fox, *General Solution to the U(1) Anomaly Equations*, *Phys. Rev. Lett.* **123** (2019), no. 15 151601, [arXiv:1905.13729].

[1094] M. B. Green and J. H. Schwarz, *Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory*, *Phys. Lett. B* **149** (1984) 117–122.

[1095] R. Barbier et al., *R-parity violating supersymmetry*, *Phys. Rept.* **420** (2005) 1–202, [hep-ph/0406039].

[1096] J. M. Arnold, B. Fornal, and M. B. Wise, *Phenomenology of scalar leptoquarks*, *Phys. Rev. D* **88** (2013) 035009, [arXiv:1304.6119].

[1097] W. Altmannshofer, J. Davighi, and M. Nardecchia, *Gauging the accidental symmetries of the standard model, and implications for the flavor anomalies*, *Phys. Rev. D* **101** (2020), no. 1 015004, [arXiv:1909.02021].

[1098] G. Buchalla and A. J. Buras, *The rare decays* $K^+ \to \pi^+ \nu\bar{\nu}$ *and* $K_L \to \mu^+\mu^-$ *beyond leading logarithms*, *Nucl. Phys. B* **412** (1994) 106–142, [hep-ph/9308272].

[1099] M. Misiak and J. Urban, *QCD corrections to FCNC decays mediated by Z penguins and W boxes*, *Phys. Lett. B* **451** (1999) 161–169, [hep-ph/9901278].

[1100] G. Buchalla and A. J. Buras, *The rare decays* $K \to \pi\nu\bar{\nu}$, $B \to X\nu\bar{\nu}$ *and* $B \to l^+l^-$*: An Update*, *Nucl. Phys. B* **548** (1999) 309–327, [hep-ph/9901288].

[1101] P. Ball and R. Zwicky, *New results on* $B \to \pi, K, \eta$ *decay formfactors from light-cone sum rules*, *Phys. Rev. D* **71** (2005) 014015, [hep-ph/0406232].

[1102] P. Ball and R. Zwicky, $B_{d,s} \to \rho, \omega, K^*, \varphi$ *decay form-factors from light-cone sum rules revisited*, *Phys. Rev. D* **71** (2005) 014029, [hep-ph/0412079].

[1103] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y. M. Wang, *Charm-loop effect in* $B \to K^{(*)}\ell^+\ell^-$ *and* $B \to K^*\gamma$, *JHEP* **09** (2010) 089, [arXiv:1006.4945].

[1104] J. Brod, M. Gorbahn, and E. Stamou, *Two-Loop Electroweak Corrections for the* $K \to \pi\nu\bar{\nu}$ *Decays*, *Phys. Rev. D* **83** (2011) 034030, [arXiv:1009.0947].

[1105] **HPQCD** Collaboration, C. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, *Rare decay* $B \to K\ell^+\ell^-$ *form factors from lattice QCD*, *Phys. Rev. D* **88** (2013), no. 5 054509, [arXiv:1306.2384]. [Erratum: Phys.Rev.D 88, 079901 (2013)].

[1106] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Lattice QCD calculation of form factors describing the rare decays* $B \to K^*\ell^+\ell^-$ *and* $B_s \to \varphi\ell^+\ell^-$, *Phys. Rev. D* **89** (2014), no. 9 094501, [arXiv:1310.3722].

[1107] W. Altmannshofer, A. J. Buras, D. M. Straub, and M. Wick, *New strategies for New Physics search in* $B \to K^*\nu\bar{\nu}$, $B \to K\nu\bar{\nu}$ *and* $B \to X_s\nu\bar{\nu}$ *decays*, *JHEP* **04** (2009) 022, [arXiv:0902.0160].

[1108] A. J. Buras, J. Girrbach-Noe, C. Niehoff, and D. M. Straub, $B \to K^{(*)}\nu\bar{\nu}$ *decays in the Standard Model and beyond*, *JHEP* **02** (2015) 184, [arXiv:1409.4557].

[1109] T. Blake, G. Lanfranchi, and D. M. Straub, *Rare* B *Decays as Tests of the Standard Model*, *Prog. Part. Nucl. Phys.* **92** (2017) 50–91, [arXiv:1606.00916].

[1110] **Belle** Collaboration, O. Lutz et al., *Search for* $B \to h^{(*)}\nu\bar{\nu}$ *with the full Belle* $\Upsilon(4S)$ *data sample*, *Phys. Rev. D* **87** (2013), no. 11 111103, [arXiv:1303.3719].

[1111] **BaBar** Collaboration, J. P. Lees et al., *Search for* $B \to K^{(*)}\nu\bar{\nu}$ *and invisible quarkonium decays*, *Phys. Rev. D* **87** (2013), no. 11 112005, [arXiv:1303.7465].

[1112] **Belle** Collaboration, J. Grygier et al., *Search for* $B \to h\nu\bar{\nu}$ *decays with semileptonic tagging at Belle*, *Phys. Rev. D* **96** (2017), no. 9 091101, [arXiv:1702.03224]. [Addendum: Phys.Rev.D 97, 099902 (2018)].

[1113] **Belle-II** Collaboration, W. Altmannshofer et al., *The Belle II Physics Book*, *PTEP* **2019** (2019), no. 12 123C01, [arXiv:1808.10567]. [Erratum: PTEP 2020, 029201 (2020)].

[1114] M. Bauer and M. Neubert, *Minimal Leptoquark Explanation for the* $R_{D^{(*)}}$, R_K , *and* $(g-2)_\mu$ *Anomalies*, *Phys. Rev. Lett.* **116** (2016), no. 14 141802, [arXiv:1511.01900].

[1115] D. Bevcirević, N. Kovsnik, O. Sumensari, and R. Zukanovich Funchal, *Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive* $b \to s\ell_1\ell_2$ *modes*, *JHEP* **11** (2016) 035, [arXiv:1608.07583].

[1116] T. S. Roussy et al., *A new bound on the electron's electric dipole moment*, *Science* **381** (2023) 46, [arXiv:2212.11841].

[1117] C. J. Ho, J. A. Devlin, I. M. Rabey, P. Yzombard, J. Lim, S. C. Wright, N. J. Fitch, E. A. Hinds, M. R. Tarbutt, and B. E. Sauer, *New techniques for a measurement of the electron's electric dipole moment*, *New J. Phys.* **22** (2020), no. 5 053031, [arXiv:2002.02332].

[1118] N. J. Fitch, J. Lim, E. A. Hinds, B. E. Sauer, and M. R. Tarbutt, *Methods for measuring the electron's electric dipole moment using ultracold YbF molecules*, *Quantum Sci. Technol.* **6** (2021), no. 1 014006, [arXiv:2009.00346].

[1119] **MEG** Collaboration, A. M. Baldini et al., *Search for the lepton flavour violating decay* $\mu^+ \to e^+\gamma$ *with the full dataset of the MEG experiment*, *Eur. Phys. J. C* **76** (2016), no. 8 434, [arXiv:1605.05081].

[1120] **MEG II** Collaboration, A. Baldini et al., *The design of the MEG II experiment*, *Eur. Phys. J. C* **78** (2018), no. 5 380, [arXiv:1801.04688].

[1121] **SINDRUM II** Collaboration, W. H. Bertl et al., *A Search for muon to electron conversion in muonic gold*, *Eur. Phys. J. C* **47** (2006) 337–346.

[1122] **Mu2e** Collaboration, L. Bartoszek et al., *Mu2e Technical Design Report*, arXiv:1501.05241.

[1123] **Mu2e** Collaboration, F. Abusalma et al., *Expression of Interest for Evolution of the Mu2e Experiment*, arXiv:1802.02599.

[1124] **Belle** Collaboration, A. Abdesselam et al., *Search for lepton-flavor-violating tau-lepton decays to* $\ell\gamma$ *at Belle*, *JHEP* **10** (2021) 19, [arXiv:2103.12994].

[1125] **Belle-II** Collaboration, L. Aggarwal et al., *Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond*, arXiv:2207.06307.

[1126] S. Banerjee, *Searches for Lepton Flavor Violation in Tau Decays at Belle II*, *Universe* **8** (2022), no. 9 480, [arXiv:2209.11639].

[1127] **NA62** Collaboration, E. Cortina Gil et al., *Measurement of the very rare* $K^+ \to \pi^+ \nu\overline{\nu}$ *decay*, *JHEP* **06** (2021) 093, [arXiv:2103.15389].

[1128] **NA62, KLEVER** Collaboration, *Rare decays at the CERN high-intensity kaon beam facility*, arXiv:2009.10941.

[1129] A. Cerri et al., *Report from Working Group 4: Opportunities in Flavour Physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 867–1158, [arXiv:1812.07638].

[1130] A. J. Buras, D. Buttazzo, J. Girrbach-Noe, and R. Knegjens, $K^+ \to \pi^+ \nu\overline{\nu}$ *and* $K_L \to \pi^0 \nu\overline{\nu}$ *in the Standard Model: status and perspectives*, *JHEP* **11** (2015) 033, [arXiv:1503.02693].

[1131] **Particle Data Group** Collaboration, R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[1132] **chiQCD** Collaboration, G. Wang, T. Draper, K.-F. Liu, and Y.-B. Yang, *Muon g-2 with overlap valence fermions*, arXiv:2204.01280.

[1133] M. Cè et al., *Window observable for the hadronic vacuum polarization contribution to the muon g-2 from lattice QCD*, *Phys. Rev. D* **106** (2022), no. 11 114502, [arXiv:2206.06582].

[1134] **Extended Twisted Mass** Collaboration, C. Alexandrou et al., *Lattice calculation of the short and intermediate time-distance hadronic vacuum polarization contributions to the muon magnetic moment using twisted-mass fermions*, *Phys. Rev. D* **107** (2023), no. 7 074506, [arXiv:2206.15084].

[1135] **Fermilab Lattice, MILC, HPQCD** Collaboration, C. T. H. Davies et al., *Windows on the hadronic vacuum polarization contribution to the muon anomalous magnetic moment*, *Phys. Rev. D* **106** (2022), no. 7 074509, [arXiv:2207.04765].

[1136] A. Bazavov et al., *Light-quark connected intermediate-window contributions to the muon* $g-2$ *hadronic vacuum polarization from lattice QCD*, arXiv:2301.08274.

[1137] T. Blum et al., *An update of Euclidean windows of the hadronic vacuum polarization*, arXiv:2301.08696.

[1138] D. Stockinger, *The Muon Magnetic Moment and Supersymmetry*, *J. Phys. G* **34** (2007) R45–R92, [hep-ph/0609168].

[1139] M. Blanke, A. J. Buras, B. Duling, A. Poschenrieder, and C. Tarantino, *Charged Lepton Flavour Violation and (g-2)(mu) in the Littlest Higgs Model with T-Parity: A Clear Distinction from Supersymmetry*, *JHEP* **05** (2007) 013, [hep-ph/0702136].

[1140] M. Pospelov, *Secluded U(1) below the weak scale*, *Phys. Rev. D* **80** (2009) 095002, [arXiv:0811.1030].

[1141] F. Feruglio, C. Hagedorn, Y. Lin, and L. Merlo, *Lepton Flavour Violation in Models with A(4) Flavour Symmetry*, *Nucl. Phys. B* **809** (2009) 218–243, [arXiv:0807.3160].

[1142] R. Dermisek and A. Raval, *Explanation of the Muon g-2 Anomaly with Vectorlike Leptons and its Implications for Higgs Decays*, *Phys. Rev. D* **88** (2013) 013017, [arXiv:1305.3522].

[1143] P. Agrawal, Z. Chacko, and C. B. Verhaaren, *Leptophilic Dark Matter and the Anomalous Magnetic Moment of the Muon*, *JHEP* **08** (2014) 147, [arXiv:1402.7369].

[1144] L. Calibbi, P. Paradisi, and R. Ziegler, *Lepton Flavor Violation in Flavored Gauge Mediation*, *Eur. Phys. J. C* **74** (2014), no. 12 3211, [arXiv:1408.0754].

[1145] L. Calibbi, R. Ziegler, and J. Zupan, *Minimal models for dark matter and the muon g-2 anomaly*, *JHEP* **07** (2018) 046, [arXiv:1804.00009].

[1146] S. Saad, *Combined explanations of* $(g-2)_\mu$, $R_{D^{(*)}}$, $R_{K^{(*)}}$ *anomalies in a two-loop radiative neutrino mass model*, *Phys. Rev. D* **102** (2020), no. 1 015019, [arXiv:2005.04352].

[1147] L. Calibbi, M. L. López-Ibáñez, A. Melis, and O. Vives, *Muon and electron* $g-2$ *and lepton masses in flavor models*, *JHEP* **06** (2020) 087, [arXiv:2003.06633].

[1148] L. Calibbi, M. L. López-Ibáñez, A. Melis, and O. Vives, *Implications of the Muon g-2 result on the flavour structure of the lepton mass matrix*, *Eur. Phys. J. C* **81** (2021), no. 10 929, [arXiv:2104.03296].

[1149] P. Fileviez Perez, C. Murgui, and A. D. Plascencia, *Leptoquarks and matter unification: Flavor anomalies and the muon g-2*, *Phys. Rev. D* **104** (2021), no. 3 035041, [arXiv:2104.11229].

[1150] M. L. López-Ibáñez, A. Melis, M. J. Pérez, M. H. Rahat, and O. Vives, *Constraining low-scale flavor models with* $(g-2)_\mu$ *and lepton flavor violation*, *Phys. Rev. D* **105** (2022), no. 3 035021, [arXiv:2112.11455].

[1151] B. Diaz, M. Schmaltz, and Y.-M. Zhong, *The leptoquark Hunter's guide: Pair production*, *JHEP* **10** (2017) 097, [arXiv:1706.05033].

[1152] M. Schmaltz and Y.-M. Zhong, *The leptoquark Hunter's guide: large coupling*, *JHEP* **01** (2019) 132, [arXiv:1810.10017].

[1153] B. C. Allanach, B. Gripaios, and T. You, *The case for future hadron colliders from* $B \to K^{(*)}\mu^+\mu^-$ *decays*, *JHEP* **03** (2018) 021, [arXiv:1710.06363].

[1154] B. C. Allanach, T. Corbett, and M. Madigan, *Sensitivity of Future Hadron Colliders to Leptoquark Pair Production in the Di-Muon Di-Jets Channel*, *Eur. Phys. J. C* **80** (2020), no. 2 170, [arXiv:1911.04455].

[1155] P. Bandyopadhyay, S. Dutta, M. Jakkapu, and A. Karan, *Distinguishing Leptoquarks at the LHC/FCC*, *Nucl. Phys. B* **971** (2021) 115524, [arXiv:2007.12997].

[1156] G. Hiller, D. Loose, and I. Ni**v**sand**v**zić, *Flavorful leptoquarks at the LHC and beyond: spin 1*, *JHEP* **06** (2021) 080, [arXiv:2103.12724].

[1157] G.-y. Huang, S. Jana, F. S. Queiroz, and W. Rodejohann, *Probing the RK(*) anomaly at a muon collider*, *Phys. Rev. D* **105** (2022), no. 1 015013, [arXiv:2103.01617].

[1158] P. Asadi, R. Capdevilla, C. Cesarotti, and S. Homiller, *Searching for leptoquarks at future muon colliders*, *JHEP* **10** (2021) 182, [arXiv:2104.05720].

[1159] P. Bandyopadhyay, A. Karan, R. Mandal, and S. Parashar, *Distinguishing signatures of scalar leptoquarks at hadron and muon colliders*, *Eur. Phys. J. C* **82** (2022), no. 10 916, [arXiv:2108.06506].

[1160] S. Qian, C. Li, Q. Li, F. Meng, J. Xiao, T. Yang, M. Lu, and Z. You, *Searching for heavy leptoquarks at a muon collider*, *JHEP* **12** (2021) 047, [arXiv:2109.01265].

[1161] S. Parashar, A. Karan, Avnish, P. Bandyopadhyay, and K. Ghosh, *Phenomenology of scalar leptoquarks at the LHC in explaining the radiative neutrino masses, muon g-2, and lepton flavor violating observables*, *Phys. Rev. D* **106** (2022), no. 9 095040, [arXiv:2209.05890].

[1162] **Muon Collider** Collaboration, J. de Blas et al., *The physics case of a 3 TeV muon collider stage*, arXiv:2203.07261.

[1163] **COMET** Collaboration, R. Abramishvili et al., *COMET Phase-I Technical Design Report*, *PTEP* **2020** (2020), no. 3 033C01, [arXiv:1812.09018].

[1164] **COMET** Collaboration, J. Angélique et al., *COMET - A submission to the 2020 update of the European Strategy for Particle Physics on behalf of the COMET collaboration*, arXiv:1812.07824.

[1165] **Mu2e-II** Collaboration, K. Byrum et al., *Mu2e-II: Muon to electron conversion with PIP-II*, in *Snowmass 2021*, 3, 2022. arXiv:2203.07569.

[1166] **C. Group** Collaboration, M. Aoki et al., *A New Charged Lepton Flavor Violation Program at Fermilab*, in *Snowmass 2021*, 3, 2022. arXiv:2203.08278.

[1167] S. A. R. Ellis and A. Pierce, *Impact of Future Lepton Flavor Violation Measurements in the Minimal Supersymmetric Standard Model*, *Phys. Rev. D* **94** (2016), no. 1 015014, [arXiv:1604.01419].

[1168] S. Homiller, Q. Lu, and M. Reece, *Complementary signals of lepton flavor violation at a high-energy muon collider*, *JHEP* **07** (2022) 036, [arXiv:2203.08825].

[1169] A. Baldini et al., *A submission to the 2020 update of the European Strategy for Particle Physics on behalf of the COMET, MEG, Mu2e and Mu3e collaborations*, arXiv:1812.06540.

[1170] **ACME** Collaboration, V. Andreev et al., *Improved limit on the electric dipole moment of the electron*, *Nature* **562** (2018), no. 7727 355–360.

[1171] R. Alarcon et al., *Electric dipole moments and the search for new physics*, in *Snowmass 2021*, 3, 2022. arXiv:2203.08103.

[1172] **UTfit** Collaboration, M. Bona et al., *Model-independent constraints on $\Delta F = 2$ operators and the scale of new physics*, *JHEP* **03** (2008) 049, [arXiv:0707.0636].

[1173] **UTfit** Collaboration, F. Ferrari, *Updates in the Unitarity Triangle fits with UTfit*, *PoS* **CKM2021** (2023) 078.

[1174] E. Goudzovski et al., *New physics searches at kaon and hyperon factories*, *Rept. Prog. Phys.* **86** (2023), no. 1 016201, [arXiv:2201.07805].

[1175] K. Aoki et al., *Extension of the J-PARC Hadron Experimental Facility: Third White Paper*, arXiv:2110.04462.

[1176] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012].

[1177] G. Isidori, G. Ridolfi, and A. Strumia, *On the metastability of the standard model vacuum*, *Nucl.Phys.* **B609** (2001) 387–409, [hep-ph/0104016].

[1178] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, *JHEP* **1208** (2012) 098, [arXiv:1205.6497].

[1179] A. Andreassen, W. Frost, and M. D. Schwartz, *Scale Invariant Instantons and the Complete Lifetime of the Standard Model*, *Phys. Rev. D* **97** (2018), no. 5 056006, [arXiv:1707.08124].

[1180] M. Czakon, A. Mitov, M. Papucci, J. T. Ruderman, and A. Weiler, *Closing the stop gap*, *Phys. Rev. Lett.* **113** (2014), no. 20 201803, [arXiv:1407.1043].

[1181] T. Eifert and B. Nachman, *Sneaky light stop*, *Phys. Lett. B* **743** (2015) 218–223, [arXiv:1410.7025].

[1182] **ATLAS** Collaboration, G. Aad et al., *Measurement of the* $t\bar{t}$ *production cross-section using* $e\mu$ *events with b-tagged jets in pp collisions at* $\sqrt{s}$ *= 7 and 8* TeV *with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014), no. 10 3109, [arXiv:1406.5375]. [Addendum: Eur.Phys.J.C 76, 642 (2016)].

[1183] T. Cohen, W. Hopkins, S. Majewski, and B. Ostdiek, *Magnifying the ATLAS Stealth Stop Splinter: Impact of Spin Correlations and Finite Widths*, *JHEP* **07** (2018) 142, [arXiv:1804.00111].

[1184] T. Cohen, S. Majewski, B. Ostdiek, and P. Zheng, *On the ATLAS Top Mass Measurements and the Potential for Stealth Stop Contamination*, *JHEP* **06** (2020) 019, [arXiv:1909.09670].

[1185] **ATLAS** Collaboration, M. Aaboud et al., *Measurements of top-quark pair spin correlations in the* $e\mu$ *channel at* $\sqrt{s} = 13$ *TeV using* pp *collisions in the ATLAS detector*, *Eur. Phys. J. C* **80** (2020), no. 8 754, [arXiv:1903.07570].

[1186] C. collaboration et al., *Measurement of the tt production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at sqrt s = 13 tev*, .

[1187] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the t-tbar production cross section in the e-mu channel in proton-proton collisions at sqrt(s) = 7 and 8 TeV*, *JHEP* **08** (2016) 029, [arXiv:1603.02303].

[1188] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the* $t\bar{t}$ *production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **79** (2019), no. 5 368, [arXiv:1812.10505].

[1189] **ATLAS** Collaboration, *Measurement of the* $t\bar{t}$ *production cross-section and lepton differential distributions in* $e\mu$ *dilepton events from* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, .

[1190] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of lepton differential distributions and the top quark mass in* $t\bar{t}$ *production in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *Eur. Phys. J. C* **77** (2017), no. 11 804, [arXiv:1709.09407].

[1191] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of* $t\bar{t}$ *normalised multi-differential cross sections in pp collisions at* $\sqrt{s} = 13$ *TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions*, *Eur. Phys. J. C* **80** (2020), no. 7 658, [arXiv:1904.05237].

[1192] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the top quark mass in the* $t\bar{t} \to$ *dilepton channel from* $\sqrt{s} = 8$ *TeV ATLAS data*, *Phys. Lett. B* **761** (2016) 350–371, [arXiv:1606.02179].

[1193] **ATLAS** Collaboration, M. Aaboud et al., *Top-quark mass measurement in the all-hadronic* $t\bar{t}$ *decay channel at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* **09** (2017) 118, [arXiv:1702.07546].

[1194] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the top quark mass in the* $t\bar{t} \to$ *lepton+jets channel from* $\sqrt{s} = 8$ *TeV ATLAS data and combination with previous results*, *Eur. Phys. J. C* **79** (2019), no. 4 290, [arXiv:1810.01772].

[1195] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the top quark mass with lepton+jets final states using* p p *collisions at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **78** (2018), no. 11 891, [arXiv:1805.01428].

[1196] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the top quark mass in the all-jets final state at* $\sqrt{s} = 13$ *TeV and combination with the lepton+jets channel*, *Eur. Phys. J. C* **79** (2019), no. 4 313, [arXiv:1812.10534].

[1197] A. H. Hoang, S. Plätzer, and D. Samitz, *On the Cutoff Dependence of the Quark Mass Parameter in Angular Ordered Parton Showers*, *JHEP* **10** (2018) 200, [arXiv:1807.06617].

[1198] S. Fleming, A. H. Hoang, S. Mantry, and I. W. Stewart, *Jets from massive unstable particles: Top-mass determination*, *Phys. Rev. D* **77** (2008) 074010, [hep-ph/0703207].

[1199] A. H. Hoang and I. W. Stewart, *Top Mass Measurements from Jets and the Tevatron Top-Quark Mass*, *Nucl. Phys. B Proc. Suppl.* **185** (2008) 220–226, [arXiv:0808.0222].

[1200] M. Butenschoen, B. Dehnadi, A. H. Hoang, V. Mateu, M. Preisser, and I. W. Stewart, *Top Quark Mass Calibration for Monte Carlo Event Generators*, *Phys. Rev. Lett.* **117** (2016), no. 23 232001, [arXiv:1608.01318].

[1201] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart, *Extracting a Short Distance Top Mass with Light Grooming*, *Phys. Rev. D* **100** (2019), no. 7 074021, [arXiv:1708.02586].

[1202] A. H. Hoang, *What is the Top Quark Mass?*, arXiv:2004.12915.

[1203] J. Kieseler, K. Lipka, and S.-O. Moch, *Calibration of the top-quark monte carlo mass*, *Phys. Rev. Lett.* **116** (Apr, 2016) 162001.

[1204] A. Andreassen and M. D. Schwartz, *Reducing the Top Quark Mass Uncertainty with Jet Grooming*, *JHEP* **10** (2017) 151, [arXiv:1705.07135].

[1205] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [arXiv:1402.2657].

[1206] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121, [arXiv:1810.05165].

[1207] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, *Deep sets*, in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[1208] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[1209] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [arXiv:1111.6097].

[1210] M. J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, *Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks*, arXiv:2010.09206.

[1211] *ATLAS Pythia 8 tunes to 7 TeV datas*, .

[1212] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, and W. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [hep-ph/0201195].

[1213] G. Watt and R. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, *JHEP* **08** (2012) 052, [arXiv:1205.4024].

[1214] S. Carrazza, S. Forte, and J. Rojo, *Parton Distributions and Event Generators*, in *43rd International Symposium on Multiparticle Dynamics*, pp. 89–96, 2013. arXiv:1311.5887.

[1215] A. M. Cooper-Sarkar, *HERAPDF1.5LO PDF Set with Experimental Uncertainties*, *PoS* **DIS2014** (2014) 032.

[1216] S. Argyropoulos and T. Sjöstrand, *Effects of color reconnection on $t\bar{t}$ final states at the LHC*, *JHEP* **11** (2014) 043, [arXiv:1407.6653].

[1217] P. Z. Skands and D. Wicke, *Non-perturbative QCD effects and the top mass at the Tevatron*, *Eur. Phys. J. C* **52** (2007) 133–140, [hep-ph/0703081].

[1218] F. Chollet et al., "Keras." https://keras.io, 2015.

[1219] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980.

[1220] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, *Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure*, *Phys. Rev. Lett.* **108** (2012) 182003, [arXiv:1201.1914].

[1221] Y.-T. Chien, *Telescoping jets: Probing hadronic event structure with multiple R 's*, *Phys. Rev. D* **90** (2014), no. 5 054008, [arXiv:1304.5240].

[1222] Y.-T. Chien, D. Farhi, D. Krohn, A. Marantan, D. Lopez Mateos, and M. Schwartz, *Quantifying the power of multiple event interpretations*, *JHEP* **12** (2014) 140, [arXiv:1407.2892].

[1223] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[1224] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, *Parameterized neural networks for high-energy physics*, *Eur. Phys. J. C* **76** (2016), no. 5 235, [arXiv:1601.07913].

[1225] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, *MadMiner: Machine learning-based inference for particle physics*, *Comput. Softw. Big Sci.* **4** (2020), no. 1 3, [arXiv:1907.10621].

[1226] M. Erdmann, B. Fischer, D. Noll, Y. Alexander Rath, M. Rieger, and D. Josef Schmidt, *Adversarial Neural Network-based data-simulation corrections for jet-tagging at CMS*, *J. Phys. Conf. Ser.* **1525** (2020), no. 1 012094.

[1227] C. Badiali, F. Di Bello, G. Frattari, E. Gross, V. Ippolito, M. Kado, and J. Shlomi, *Efficiency Parameterization with Neural Networks*, arXiv:2004.02665.

[1228] A. Blance and M. Spannowsky, *Unsupervised Event Classification with Graphs on Classical and Photonic Quantum Computers*, arXiv:2103.03897.

[1229] T. Dorigo, M. Fumanelli, C. Maccani, M. Mojsovska, G. C. Strong, and B. Scarpa, *RanBox: Anomaly Detection in the Copula Space*, arXiv:2106.05747.

[1230] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [arXiv:1011.2268].

[1231] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Metric Space of Collider Events*, *Phys. Rev. Lett.* **123** (2019), no. 4 041801, [arXiv:1902.02346].

[1232] P. T. Komiske, E. M. Metodiev, and J. Thaler, *The Hidden Geometry of Particle Collisions*, *JHEP* **07** (2020) 006, [arXiv:2004.04159].

[1233] M. A. Kramer, *Nonlinear principal component analysis using autoassociative neural networks*, *AIChE Journal* **37** (1991), no. 2 233–243, [https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690370209].

[1234] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, arXiv:1312.6114.

[1235] J. An and S. Cho, *Variational autoencoder based anomaly detection using reconstruction probability*, *Special Lecture on IE* **2** (2015) 1.

[1236] K. Dohi, *Variational Autoencoders for Jet Simulation*, arXiv:2009.04842.

[1237] J. H. Collins, *An Exploration of Learnt Representations of W Jets*, arXiv:2109.10919.

[1238] T. Cheng, *Test sets for jet anomaly detection at the lhc*, Mar., 2021.

[1239] J. Leissner-Martin, T. Cheng, and J.-F. Arguin, *Qcd jet samples with particle flow constituents*, July, 2020.

[1240] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [arXiv:1307.6346].

[1241] M. Cacciari and G. P. Salam, *Dispelling the* N^3 *myth for the* k_t *jet-finder*, *Phys. Lett. B* **641** (2006) 57–61, [hep-ph/0512210].

[1242]

[1243] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118, [arXiv:1407.5675].

[1244] C. Villani, *Optimal Transport, Old and New*. Springer, 2009.

[1245] C. Cesarotti and J. Thaler, *A Robust Measure of Event Isotropy at Colliders*, *JHEP* **08** (2020) 084, [arXiv:2004.06125].

[1246] C. Cesarotti, M. Reece, and M. J. Strassler, *Spheres To Jets: Tuning Event Shapes with 5d Simplified Models*, arXiv:2009.08981.

[1247] C. Cesarotti, M. Reece, and M. J. Strassler, *The Efficacy of Event Isotropy as an Event Shape Observable*, arXiv:2011.06599.

[1248] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), pp. 8024–8035. Curran Associates, Inc., 2019.

[1249] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, *Fast and accurate deep network learning by exponential linear units (elus)*, arXiv:1511.07289.

[1250] D. P. Kingma and M. Welling, *An Introduction to Variational Autoencoders*, arXiv:1906.02691.

[1251] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, *Variational inference: A review for statisticians*, *Journal of the American Statistical Association* **112** (4, 2017) 859–877.

[1252] D. Jimenez Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, arXiv:1505.05770.

[1253] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, *Improving Variational Inference with Inverse Autoregressive Flow*, arXiv:1606.04934.

[1254] R. van den Berg, L. Hasenclever, J. M. Tomczak, and M. Welling, *Sylvester Normalizing Flows for Variational Inference*, arXiv:1803.05649.

[1255] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980.

[1256] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouve, and G. Peyré, *Interpolating between optimal transport and mmd using sinkhorn divergences*, in *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2681–2690, 2019.

[1257] P. T. Komiske, R. Mastandrea, E. M. Metodiev, P. Naik, and J. Thaler, *Exploring the Space of Jets with CMS Open Data*, *Phys. Rev. D* **101** (2020), no. 3 034009, [arXiv:1908.08542].

[1258] M. Crispim Romão, N. F. Castro, J. G. Milhano, R. Pedro, and T. Vale, *Use of a generalized energy Mover's distance in the search for rare phenomena at colliders*, *Eur. Phys. J. C* **81** (2021), no. 2 192, [arXiv:2004.09360].

[1259] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, *Weakly Supervised Classification in High Energy Physics*, *JHEP* **05** (2017) 145, [arXiv:1702.00414].

[1260] T. Cohen, M. Freytsis, and B. Ostdiek, *(Machine) Learning to Do More with Less*, *JHEP* **02** (2018) 034, [arXiv:1706.09451].

[1261] M. Reece, *Axion-gauge coupling quantization with a twist*, *JHEP* **10** (2023) 116, [arXiv:2309.03939].

[1262] Y. Choi, M. Forslund, H. T. Lam, and S.-H. Shao, *Quantization of Axion-Gauge Couplings and Noninvertible Higher Symmetries*, *Phys. Rev. Lett.* **132** (2024), no. 12 121601, [arXiv:2309.03937].

[1263] J. W. Foster, Y. Park, B. R. Safdi, Y. Soreq, and W. L. Xu, *Search for dark matter lines at the Galactic Center with 14 years of Fermi data*, *Phys. Rev. D* **107** (2023), no. 10 103047, [arXiv:2212.07435].

[1264] P. De La Torre Luque, J. Smirnov, and T. Linden, *Gamma-ray lines in 15 years of Fermi-LAT data: New constraints on Higgs portal dark matter*, *Phys. Rev. D* **109** (2024), no. 4 L041301, [arXiv:2309.03281].

[1265] M. Eto, Y. Hamada, R. Jinno, M. Nitta, and M. Yamada, *Neutrino zeromodes on electroweak strings in light of topological insulators*, arXiv:2402.19417.

[1266] L. Trefethen, *Approximation Theory and Approximation Practice*, vol. 128. Siam, 2013.

[1267] J.-P. Berrut and L. Trefethen, *Barycentric Lagrange Interpolation*, *SIAM review* **46** (2004) 501–517.

[1268] N. Higham, *The Numerical Stability of Barycentric Lagrange Interpolation*, *IMA Journal of Numerical Analysis* **24** (2004) 547–556.

[1269] R. Baltensperger, *Improving the Accuracy of the Matrix Differentiation Method for Arbitrary Collocation Points*, *Applied Numerical Mathematics* **33** (2000) 143–149.

[1270] J. Hisano, T. Moroi, K. Tobe, and M. Yamaguchi, *Lepton flavor violation via right-handed neutrino Yukawa couplings in supersymmetric standard model*, *Phys. Rev. D* **53** (1996) 2442–2459, [hep-ph/9510309].

[1271] J. R. Ellis, J. S. Lee, and A. Pilaftsis, *Electric Dipole Moments in the MSSM Reloaded*, *JHEP* **10** (2008) 049, [arXiv:0808.1819].

[1272] A. Crivellin, M. Hoferichter, and P. Schmidt-Wellenburg, *Combined explanations of* $(g-2)_{\mu,e}$ *and implications for a large muon EDM*, *Phys. Rev. D* **98** (2018), no. 11 113002, [arXiv:1807.11484].

[1273] C. Cesarotti, Q. Lu, Y. Nakai, A. Parikh, and M. Reece, *Interpreting the Electron EDM Constraint*, *JHEP* **05** (2019) 059, [arXiv:1810.07736].

[1274] Y. Okada, K.-i. Okumura, and Y. Shimizu, *Mu –> e gamma and mu –> 3 e processes with polarized muons and supersymmetric grand unified theories*, *Phys. Rev. D* **61** (2000) 094001, [hep-ph/9906446].

[1275] Y. Kuno and Y. Okada, *Muon decay and physics beyond the standard model*, *Rev. Mod. Phys.* **73** (2001) 151–202, [hep-ph/9909265].

[1276] R. Kitano, M. Koike, and Y. Okada, *Detailed calculation of lepton flavor violating muon electron conversion rate for various nuclei*, *Phys. Rev. D* **66** (2002) 096002, [hep-ph/0203110]. [Erratum: Phys.Rev.D 76, 059902 (2007)].

[1277] S. Fajfer, J. F. Kamenik, I. Nisandzic, and J. Zupan, *Implications of Lepton Flavor Universality Violations in B Decays*, *Phys. Rev. Lett.* **109** (2012) 161801, [arXiv:1206.1872].

[1278] D. Bevcirević, O. Sumensari, and R. Zukanovich Funchal, *Lepton flavor violation in exclusive* $b \to s$ *decays*, *Eur. Phys. J. C* **76** (2016), no. 3 134, [arXiv:1602.00881].

[1279] Y. Sakaki, M. Tanaka, A. Tayduganov, and R. Watanabe, *Testing leptoquark models in* $\bar{B} \to D^{(*)}\tau\bar{\nu}$, *Phys. Rev. D* **88** (2013), no. 9 094012, [arXiv:1309.0301].

[1280] M. Freytsis, Z. Ligeti, and J. T. Ruderman, *Flavor models for* $\bar{B} \to D^{(*)}\tau\bar{\nu}$, *Phys. Rev. D* **92** (2015), no. 5 054018, [arXiv:1506.08896].

[1281] Y. Cai, J. Gargalionis, M. A. Schmidt, and R. R. Volkas, *Reconsidering the One Leptoquark solution: flavor anomalies and neutrino mass*, *JHEP* **10** (2017) 047, [arXiv:1704.05849].

[1282] M. Tanaka and R. Watanabe, *New physics in the weak interaction of* $\bar{B} \to D^{(*)}\tau\bar{\nu}$, *Phys. Rev. D* **87** (2013), no. 3 034028, [arXiv:1212.1878].

[1283] P. Asadi, M. R. Buckley, and D. Shih, *Asymmetry Observables and the Origin of* $R_{D^{(*)}}$ *Anomalies*, *Phys. Rev. D* **99** (2019), no. 3 035015, [arXiv:1810.06597].

[1284] D. Bardhan, P. Byakti, and D. Ghosh, *A closer look at the* R_D *and* R_{D^*} *anomalies*, *JHEP* **01** (2017) 125, [arXiv:1610.03038].

[1285] A. J. Buras, F. Schwab, and S. Uhlig, *Waiting for precise measurements of* $K^+ \to \pi^+ \nu\bar{\nu}$ *and* $K_L \to \pi^0 \nu\bar{\nu}$, *Rev. Mod. Phys.* **80** (2008) 965–1007, [hep-ph/0405132].

[1286] Y. Grossman and Y. Nir, $K_L \to \pi^0 \nu\bar{\nu}$ *beyond the standard model*, *Phys. Lett. B* **398** (1997) 163–168, [hep-ph/9701313].

[1287] **KOTO** Collaboration, J. K. Ahn et al., *Study of the* $K_L \to \pi^0 \nu\bar{\nu}$ *Decay at the J-PARC KOTO Experiment*, *Phys. Rev. Lett.* **126** (2021), no. 12 121801, [arXiv:2012.07571].

[1288] **ATLAS** Collaboration, G. Aad et al., *Search for the lepton flavor violating decay* $Z \to e\mu$ *in pp collisions at* $\sqrt{s}$ *TeV with the ATLAS detector*, *Phys. Rev. D* **90** (2014), no. 7 072010, [arXiv:1408.5774].

[1289] D. Bevcirević and O. Sumensari, *A leptoquark model to accommodate* $R_K^{exp} < R_K^{SM}$ *and* $R_{K^*}^{exp} < R_{K^*}^{SM}$, *JHEP* **08** (2017) 104, [arXiv:1704.05835].

[1290] P. Arnan, D. Becirevic, F. Mescia, and O. Sumensari, *Probing low energy scalar leptoquarks by the leptonic* W *and* Z *couplings*, *JHEP* **02** (2019) 109, [arXiv:1901.06315].

[1291] A. Efrati, A. Falkowski, and Y. Soreq, *Electroweak constraints on flavorful effective theories*, *JHEP* **07** (2015) 018, [arXiv:1503.07872].

[1292] A. Juttner, *Progress in kaon physics on the lattice*, *PoS* **LATTICE2007** (2007) 014, [arXiv:0711.1239].

[1293] R. J. Dowdall, C. T. H. Davies, R. R. Horgan, G. P. Lepage, C. J. Monahan, J. Shigemitsu, and M. Wingate, *Neutral B-meson mixing from full lattice QCD at the physical point*, *Phys. Rev. D* **100** (2019), no. 9 094508, [arXiv:1907.01025].

[1294] B. Grinstein, *TASI-2013 Lectures on Flavor Physics*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Particle Physics: The Higgs Boson and Beyond*, 1, 2015. arXiv:1501.05283.

[1295] A. Lenz, U. Nierste, J. Charles, S. Descotes-Genon, A. Jantsch, C. Kaufhold, H. Lacker, S. Monteil, V. Niess, and S. T'Jampens, *Anatomy of New Physics in* $B - \bar{B}$ *mixing*, *Phys. Rev. D* **83** (2011) 036004, [arXiv:1008.1593].

[1296] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, *Rev. Mod. Phys.* **68** (1996) 1125–1144, [hep-ph/9512380].

[1297] A. J. Buras, *Flavor physics and CP violation*, in *2004 European School of High-Energy Physics*, pp. 95–168, 5, 2005. hep-ph/0505175.

[1298] E. Golowich, J. Hewett, S. Pakvasa, and A. A. Petrov, *Implications of* D^0 - $\bar{D}^0$ *Mixing for New Physics*, *Phys. Rev. D* **76** (2007) 095009, [arXiv:0705.3650].

[1299] A. Bazavov et al., *Short-distance matrix elements for* D^0*-meson mixing for* $N_f = 2 + 1$ *lattice QCD*, *Phys. Rev. D* **97** (2018), no. 3 034513, [arXiv:1706.04622].