Selection Techniques for Optimal Meta-analysis of Beyond Standard Model Physics

Jamie Yellen

A dissertation submitted to the University of Glasgow
for the degree of Doctor of Philosophy

Abstract

This thesis addresses the development of selection techniques tailored for optimising meta-analyses in Beyond Standard Model (BSM) physics, focusing on three key objectives: (i) identifying a minimally overlapping set of results, (ii) implementing these selections for model exclusion, and (iii) extending them to anomaly detection applications. Central to this study are the HDFS and WHDFS algorithms—graph-based methods that systematically address the combinatorial challenge of selecting optimal result combinations.

In the context of model exclusion, the thesis applies the WHDFS algorithm within the TACO project to optimise combinations of analyses by estimating overlaps in signal regions (SRs). Using simplified model spectra looking at SUSY-like processes, the project demonstrates a measurable increase in exclusion. The proto-models project, an extension to the TACO project and previous work by the SModelS collaboration, focused on anomaly detection, adapting the WHDFS algorithm to construct a test statistic for identifying significant deviations from the Standard Model (SM) hypothesis. Through iterative improvements in the algorithms' weighting mechanisms, the study presents a self-regulating test statistic for the measure of significance.

The findings highlight the dual utility of the HDFS and WHDFS algorithms across domains, from collider-based physics applications to machine learning contexts. This work thus contributes a computationally robust framework that enhances reinterpretation capacity in particle physics and supports further integration with reinterpretation tools like SModelS, MadAnalysis 5 Rivet and Contur. The research underscores the increasing importance of efficient, adaptable algorithms for data-intensive BSM analyses. It lays the groundwork for future reinterpretation methodologies necessary for maximising data utility in HL-LHC and related high-energy physics experiments.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Jamie Yellen

Acknowledgements

The completion of this thesis was achieved through the collaboration, support, and encouragement of many individuals. First, I am deeply grateful to my supervisor, Prof. Andy Buckley, who provided thoughtful guidance, expertise, and motivation throughout my Ph.D. studies.

I am fortunate to have had the support of Dr. Wolfgang Waltenburger, Dr. Sabine Kraml and Dr. Christian Gutschow who provided assistance, technical help, and friendship over the last four years. I would also like to recognise the financial and institutional support provided by UKRI and ScotDIST, which was essential in carrying out this research.

On a personal note, I am incredibly grateful for the encouragement and patience of my family and friends. To my Mother, thank you for instilling in me the values of perseverance and dedication. To my partner, Emily, thank you for your love and encouragement and for standing by me through the highs and lows of this journey.

Preface

This thesis represents the culmination of a decade-long academic journey I began as a mature student in 2014. My early educational opportunities had been restricted due to severe dyslexia, but through the Scottish Wider Access Program (SWAP), I gained the qualifications required to enter higher education. These initial studies formed the groundwork for my subsequent MPhys degree in Astrophysics at the University of Edinburgh, which I completed with First-Class honours in 2020. Driven to broaden my knowledge into new areas, I commenced my PhD at the University of Glasgow, concentrating on data selection methodologies within the framework of Beyond Standard Model (BSM) physics. This research has been supported by the Scottish Data-Intensive Science Triangle (ScotDIST) funding program, which allowed me to develop my skills in statistical data analysis.

My research interest in data analysis stems from a strong background in statistical methodologies developed during my MPhys studies. My focus on astrophysics and cosmology involved significant engagement with advanced Bayesian statistical techniques. Through the ScotDIST program, the postgraduate researcher position in Glasgow offered an exciting opportunity to expand my knowledge and develop these skills in new settings, specifically frequentist statistics and BSM phenomenology.

This thesis can be divided into two halves. The first half explores the foundational research and subject matter that underpins the subsequent analysis. It comprises two chapters dedicated to collider physics and supersymmetry, two chapters focusing on statics and hypothesis testing, and a chapter addressing the current state of reinterpretation.

The second half focuses on the development of algorithms and their application to two distinct projects: (1) model exclusion, known as the taco project, and (2) anomaly detection through the proto-model project. While the chapter on algorithm development stands alone, it was carried out in parallel with the taco and proto-model projects, which provided critical test cases and validation opportunities. This integrated approach

proved to be both intellectually enriching and demanding, as it necessitated a careful balance between theoretical advancements and practical applications.

Regarding the publication status of this thesis, several sections are derived from both published work and ongoing research. Chapter 7, which details the taco project, is based on the paper "*Strength in Numbers: Optimal and Scalable Combination of LHC New-Physics Searches*," in which I served as a lead author. Chapter 8, which outlines the proto-models project, is currently under collaborative development with the SModelS group, and data collection is actively in progress, with a projected publication time frame in early to mid-2025. Within Chapter 7, all result plots, except those pertaining to t-channel dark matter, were produced exclusively by myself, thereby underscoring my direct contributions to this collaborative research effort.

Collaboration was central to the success of this research. Chapter 7 reflects joint efforts with the other authors of the "*Strength in Numbers*" paper, while Chapter 8 benefited from collaborative input and resources provided by the SModelS collaboration. These partnerships significantly enriched the research process, providing diverse perspectives and access to specialised tools critical for realising these projects.

This thesis's target audience includes graduate researchers interested in phenomenological particle physics, particularly those exploring methods for meta-analysis and data-driven approaches to model exclusion and anomaly detection. I hope the methods and results presented herein will provide useful insights for future investigations.

This thesis represents a culmination of years of academic and research training and a step forward in developing robust methodologies for analysing data in the search for new physics beyond the Standard Model. I hope that the work presented here will contribute to the scientific community's ongoing efforts to advance our understanding of the fundamental nature of the universe.

Contents

"Some ideas are so stupid that only intellectuals believe them"
— George Orwell

Chapter 1.

Collider Physics

Experimental particle physics is a juxtaposition of extremes, and this is rarely more apparent when considering the experimental apparatus that constitutes the Large Hadron Collider(LHC), where femtometre-sized particles are accelerated to relativistic velocities, producing the conditions necessary to probe fundamental particles and their interactions.

This chapter will cover the fundamentals of collider physics, starting with the Standard Model (SM), exploring its theoretical foundations and the vital experimental discoveries that have validated its predictions. Section 1.1 will examine the mathematical formulation of the SM, focusing on the underlying principles of quantum field theory and gauge symmetry. Additionally, Section 1.2 will discuss the significance of electroweak symmetry breaking and the Higgs mechanism.

1.1. The standard model of particle physics

The Standard Model of particle physics, or SM for short, stands as one of the most successful theories in the history of physics. Developed throughout the 20^{th} century, it provides a comprehensive framework for understanding the fundamental particles and the forces that govern their interactions. This theory elegantly unifies the electromagnetic, weak, and strong nuclear forces under a single theoretical umbrella, excluding only gravity.

At its core, the Standard Model describes a collection of quantum fields that are components of an irreducible unitary representation of a symmetry group. [1] In this description, particles are viewed not as individual, isolated entities but as excitations or quanta of underlying fields. Quantum field theory (QFT), and for the case of collider physics, Relativistic QFT (RQFT), is the foundation of the Standard Model. In QFT, a

Lagrangian density $\mathcal{L}$ is constructed from the quantum fields; this is used to calculate the action and the transition probabilities.

The full symmetry group of the Standard Model splits into two component products, the local internal and global space-time symmetry. Following Noether's theorem, "For each symmetry of the Lagrangian, there is a conserved quantity" [2]. Thus, for each component symmetry of the SM, a corresponding set of conserved quantities exists; a good example of how the symmetries lead to conservation laws, which in turn lead to physically observable quantities, is the Lorentz group (or Poincare group), the space-time component of the SM. The representations of the Lorentz group are labelled by spin and mass, which correspond to the conservation of energy and angular momentum. For our purposes, mass is treated as a positive real number, while spin (s) may assume half-integer and integer values, which correspond to fermions and bosons, respectively. The Standard Model contains $s = 1/2$ fermions, vector bosons of $s = 1$ and a single scalar boson with $s = 0$ (the Higgs [3, 4]).

The local symmetries of the SM are the gauge groups, which are a group of invariant transformations, i.e. the U(1) transformations of the electromagnetic potential in electromagnetism. The gauge group, or local symmetry, of the SM, is $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$, the individual products are referred to as colour, weak isospin and weak hypercharge, respectively. These groups correspond to the strong, weak, and electromagnetic forces. The gauge group SU(3) is associated with the strong force, which binds quarks together to form protons, neutrons, and other hadrons. This group includes eight gauge bosons (components) known as gluons, which mediate the interactions between quarks. The SU(2) group is linked to the weak force responsible for processes like beta decay. It involves three gauge bosons: (W^+), (W^-), and (Z^0), which interact with particles such as electrons and neutrinos. Lastly, the U(1) gauge group governs the electromagnetic force via the photon, which mediates the force between charged particles. Unifying SU(2) and U(1) leads to the electroweak theory, explaining how these forces merge at high energies. Particles are labelled by how they transform under this gauge group, meaning that they are labelled by representations of SU(3), SU(2) and U(1). Table 1.1 shows the three gauge fields and how they relate to their associated groups, coupling constants and components (gauge bosons).

The Standard Model has five types of fermionic, chiral fields, usually referred to as matter fields: $q_\mathrm{L}, u_\mathrm{R}, d_\mathrm{R}, l_\mathrm{L}, l_\mathrm{R}$, where q_L and l_L are left-handed fields and $u_\mathrm{R}, d_\mathrm{R}, l_\mathrm{R}$ are the right-handed fields (where L, R denotes a left and right-handed). Table 1.2 lists the

Gauge Fields				
Symbol	Group	Associated Charge	Coupling	Gauge Bosons
B	U(1)	Weak hypercharge (Y)	g_1	γ
W	SU(2)	Weak isospin (T_3)	g_2	W^+, W^-, Z^0
G	SU(3)	Colour (C)	g_3	g

Table 1.1.: The three gauge fields with associated charges, couplings and number of components

Matter Fields				
Classification		Representations		
Symbol	Name	SU(3)	SU(2)	U(1)
l_{L} (L)	Left-handed lepton	**1**	**2**	-1
l_{R} (E)	Right-handed lepton	**1**	**1**	-2
q_{L} (Q)	Left-handed quark	**3**	**2**	$+\frac{1}{3}$
u_{R} (U)	Right-handed quark (up)	**3**	**1**	$+\frac{4}{3}$
d_{R} (D)	Right-handed quark (down)	**3**	**1**	$-\frac{2}{3}$
Scalar Boson				
H	Higgs Boson	**1**	**2**	$+1$

Table 1.2.: Representations of the five matter fields and the Scalar Higgs boson [3, 4] Symbols are provided in two common conventions, and the bold numbers shown under SU(N) are not ordinary abelian charges but labels of Lie group representations [5].

fields and shows under which representation of the gauge group each field transforms (the value of the weak hypercharge (Y) is listed under U(1)). The fields q_{L} and l_{L} are weak isospin doublets, and therefore, their components are indexed by the quantum number T_3, traditionally corresponding to what is referred to as the third generator of the SU(2) or the component along the 3-axis. It is worth taking a step back here to define what a generator is, as the term will be extensively used over the next two chapters. In Lie algebra, generators are the fundamental elements from which the entire algebra can be constructed through the Lie bracket operation. They form a basis for the algebra, meaning any element of the Lie algebra can be expressed as a linear combination of these generators [6]. The relations between generators, captured by commutators (or Lie brackets), define the structure and properties of the algebra. Each symmetry group within the SM has its own set of generators, which define the interactions of fundamental particles, γ^a ($a = 1, ..., 8$) for SU(3), σ^i ($i = 1, 2, 3$) for SU(2) and Y for U(1). In the

context of symmetry groups, generators correspond to infinitesimal transformations, making them crucial in describing the underlying symmetry of physical systems [5].

Gauge fields correspond to the connections associated with a symmetry group, and they mediate interactions by ensuring local gauge invariance. Matter fields, on the other hand, transform under specific representations, which dictate their transformation properties and interactions with gauge bosons. Tables 1.1 and 1.2 give a sense of how the gauge fields relate to the matter fields through the group representations. For instance, in quantum electrodynamics (QED), the gauge group is $\mathrm{U}(1)_{\mathrm{em}}$, and matter fields correspond to charged leptons (l), such as the electron (e). These fields are chiral and following l_{R} and l_{L} are denoted as e_{R} and e_{L}. In QED, both transform under the same electric charge $-e$, though they behave differently under the electroweak interaction. Looking at Table 1.2, the right-handed electron, e_{R} is a singlet under the weak $SU(2)_L$ group and transforms under $\mathrm{U}(1)_{\mathrm{em}}$ as $e_{\mathrm{R}} \to e^{-i\alpha} e_{\mathrm{R}}$ where α is a constant. The left-handed electron is part of an electroweak doublet (with the neutrino), but in QED alone, it also transforms under $\mathrm{U}(1)_{\mathrm{em}}$ as $e_{\mathrm{L}} \to e^{-i\alpha} e_{\mathrm{L}}$ [7].

Just as the representations of the Lorentz group give rise to conserved quantities of spin and mass, the conserved quantities associated with the combined gauge group $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$ are weak isospin (T), weak hypercharge (Y), electric charge ($\mathcal{Q}$) and colour charge (C). The electric charge of the component field is calculated via the Gell-Mann-Nishijima relation [8], a phenomenologically motivated formula showing the linear combination of weak hypercharge Y and the third component of weak isospin:

$$\mathcal{Q} = T_3 + \frac{1}{2} Y \,. \tag{1.1}$$

The fermionic weak isospin T value is closely related to chirality (handedness). Left-handed fermions (and right-handed anti-fermions) have a weak isospin of $+\frac{1}{2}$, which are grouped into doublets with the third component T_3 taking the value of $\pm\frac{1}{2}$. Right-handed fermions (and left-handed anti-fermions) have $T = 0$ and thus $T_3 = 0$, resulting in singlets that do not undergo charged weak interactions. Table 1.3 lists electroweak quantum numbers for all the SM matter particles. The values of electromagnetic charge are a direct application of the Gell-Mann-Nishijima relation from Equation (1.1). For example, the left-handed lepton field in Table 1.2 has a weak hypercharge of -1; thus, the left-handed isospin doublet has two instances of charge, the $T_3 = \frac{1}{2}$ is the electrically neutral neutrino ν_{L}, while the $T_3 = -\frac{1}{2}$ component has a charge of -1 which happens to take the same value as the right-handed lepton, singlet field, both are the charged lepton fields of the

Field	Y	T	T_3	$\mathcal{Q}$
e^-_{L}, μ^-_{L}, τ^-_{L}	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1
$\nu^-_{e,L}$, $\nu^-_{\mu,L}$, $\nu^-_{\tau,L}$	-1	$\frac{1}{2}$	$\frac{1}{2}$	0
e^-_{R}, μ^-_{R}, τ^-_{R}	-2	0	0	-1
$\nu^-_{e,R}$, $\nu^-_{\mu,R}$, $\nu^-_{\tau,R}$	0	0	0	0
u_{L}, c_{L}, t_{L}	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$
d_{L}, s_{L}, b_{L}	$\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$
u_{R}, c_{R}, t_{R}	$\frac{4}{3}$	0	0	$\frac{2}{3}$
d_{R}, s_{R}, b_{R}	$-\frac{2}{3}$	0	0	$-\frac{1}{3}$

Table 1.3.: Electroweak ($\mathrm{SU}(2) \times \mathrm{U}(1)$) quantum numbers for the fermions, where Y is the weak hypercharge T the total weak isospin, T_3 its component along the 3-axis and Q is the electromagnetic charge. Subscript L & R denotes a left-handed & right-handed state respectively [7]

.

electron. For each of the five matter fields, three generations of particles are present in the theory. For the leptons, the electron (e) and electron-neutrino (ν_e) are extended by the muon (μ) and muon-neutrino (ν_μ) in the second generation as well as the tau (τ) and its neutrino (ν_τ) in the third. For the quarks, a charm and strange, as well as top and bottom, are added as second and third generations.

Using the information from Tables 1.1 - 1.3, we can begin to build the kinetic terms of the Lagrangian for the electroweak theory. A generalised form of dynamic terms in the fermion field can be summarised by:

$$\mathcal{L}_f \supset \sum_{\phi} \mathrm{i}\bar{\phi}\gamma^\mu \boldsymbol{D}_\mu \phi; \quad \phi \in \{L, Q, E, \nu, U, D\}\,, \tag{1.2}$$

where $\boldsymbol{D}_\mu$ is the covariant derivative, a modified invariant derivative under local gauge transformations. The specific form that this derivative will take depends on the field, ϕ, however, by defining Λ_i and g_i as the gauge groups and coupling constants (given in Table 1.1), the covariant derivative can be represented as

$$\boldsymbol{D}_\mu = \partial_\mu + \mathrm{i}\sum_i g_i \Lambda_{\mu,i}\,. \tag{1.3}$$

It is now possible to construct the fermionic kinetic terms of $\mathcal{L}$ by combining Equations (1.2) and (1.3) with the gauge fields and matter fields of 1.1 and 1.3 [7]:

$$\begin{aligned}\mathcal{L}_{f,\text{kin}} = &\,\mathrm{i}\bar{L}_i\gamma^\mu(\partial_\mu + \mathrm{i}g_2 W_\mu^a\sigma^a + \mathrm{i}g_1 Y_\text{L} B_\mu)L_i + \mathrm{i}\bar{Q}_i\gamma^\mu(\partial_\mu + \mathrm{i}g_2 W_\mu^a\sigma^a + \mathrm{i}g_1 Y_Q B_\mu)Q_i \\ &+\mathrm{i}\bar{E}^i_\text{R}\gamma^\mu(\partial_\mu + \mathrm{i}g_1 Y_e B_\mu)E^i_\text{R} + \mathrm{i}\bar{\nu}^i_\text{R}\gamma^\mu(\partial_\mu + \mathrm{i}g_1 Y_\nu B_\mu)\nu^i_\text{R} \\ &+\mathrm{i}U^i_\text{R}\gamma^\mu(\partial_\mu + \mathrm{i}g_1 Y_U B_\mu)U^i_\text{R} + \mathrm{i}D^i_\text{R}\gamma^\mu(\partial_\mu + \mathrm{i}g_1 Y_D B_\mu)D^i_\text{R}\,,\end{aligned} \tag{1.4}$$

where the subscript f defines the fermionic terms. Looking closer at Equation (1.3) and (1.4), the effect of the covariant derivative is to weakly couple the matter fields to a set of gauge fields according to the representation. This results in a single gauge-field component B for U(1), a three-component field W for SU(2), and an eight-component field G for SU(3). Consequently, there are as many gauge bosons as there are components of the gauge field: the photon (γ) for the B field, the W^+, W^- and Z^0 for the W field and eight gluons for the G field. In addition to the fermionic terms, we also require the kinetic terms of the gauge fields,

$$\mathcal{L}_\text{kin} = -\frac{1}{4}\sum_{a=1}^{8} G^{a\mu\nu}G_{a\mu\nu} - \frac{1}{4}\sum_{a=1}^{3} W^{a\mu\nu}W_{a\mu\nu} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\,, \tag{1.5}$$

where $G^a_{\mu\nu}$, $W^a_{\mu\nu}$ and $F_{\mu\nu}$ are the field-strength tensors associated with the gauge fields G^a_μ, W^a_μ and B^a_μ respectively. The standard convention of using Einstein summation notation has been ignored here to highlight the number of component bosons for each gauge field. The kinetic terms of Equation (1.5) give rise to massless gauge fields; however, from experimental results, it has long been known that some of the gauge bosons in the electroweak theory are massive [9]. A scalar field with a non-zero vacuum expectation value is introduced to account for this discrepancy. This enables the generation of non-zero gauge-boson masses through spontaneous symmetry breaking.

1.2. Electroweak symmetry breaking

Fundamental forces in nature are associated with abstract local gauge symmetries. Requiring theories to possess such symmetries necessitates the introduction of vector fields, which, in quantum theory, give rise to force-carrying particles. The symmetries considered thus far include the U(1) symmetry of the original QED theory and the SU(3) symmetry of QCD. In both contexts, the symmetry is exact. However, it is also possible that a local gauge symmetry can appear in nature but be broken. This phenomenon

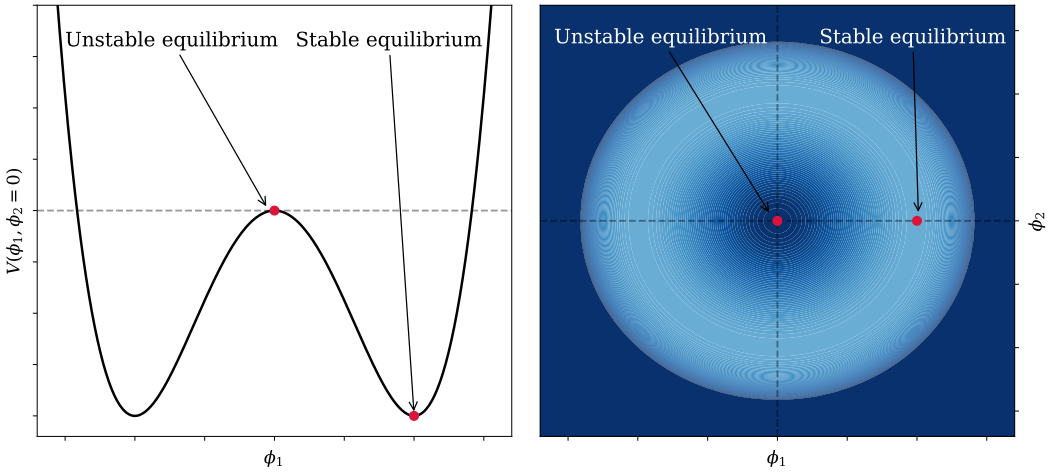


Figure 1.1.: Orthogonal slices through the potential $V(\phi_1, \phi_2)$ (pointing out of the page in the right-hand plot). The plots demonstrate that placing a particle at the apex results in a system that exhibits symmetry under rotation around its central axis

is referred to as spontaneous symmetry breaking and is intimately connected with the challenge of describing the non-zero mass of fundamental particles.

Spontaneous symmetry breaking is when a physical system, initially in a state of symmetry, ultimately transitions to an asymmetric state without any external influence. This term is often used to characterise systems where the Lagrangian has well-defined symmetries in one state, but the lowest-energy, or vacuum-state, solutions do not display the same symmetry. Consequently, when the system adopts one of these vacuum solutions, the symmetry is broken for small disturbances around that vacuum, despite the entire Lagrangian maintaining corresponding charge symmetries. An illustration of such a potential is shown in Figure 1.1, where the left and right-hand plots show orthogonal slices through the potential $V(\phi_1, \phi_2)$. The plots demonstrate that placing a particle at the apex results in a system that exhibits symmetry under rotation around its central axis. However, the particle can disrupt this symmetry by spontaneously descending into a lower-energy state. Subsequently, the particle settles at a fixed location. In this final state, both the potential and the particle maintain their respective symmetries. However, the system as a whole does not. Goldstone's theorem describes a general continuous symmetry that is spontaneously broken [10]. For example, in the context of the gauge fields described so far, the charge currents are conserved; however, the corresponding charges do not change the ground state. This kind of spontaneous breaking of a continuous symmetry inevitably gives rise to massless scalar fields known as Goldstone bosons, where there is one such boson for every degree of freedom for each of the symmetries that have

been broken [11]. By choosing a suitable gauge, the Goldstone bosons can be set to vanish; this leaves one remaining degree of freedom, the massive scalar particle, the Higgs boson. This can be shown using the U(1) gauge theory with a single gauge field [12] taken from Equation (1.4). For the case of the photon (γ), the Lagrangian is simply

$$\mathcal{L}_\gamma = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} \,. \tag{1.6}$$

Defining the field strength tensor for the photon as

$$F_{\mu\nu} = \partial_\mu A_\mu - \partial^\mu A^\mu \,. \tag{1.7}$$

The local symmetry of the U(1) group is a statement of invariance under the transformation $A_\mu(x) \to A_\mu(x) - \partial_\mu \Lambda(x)$ for any Λ and x [7]. It is easily verified that applying this transformation to Equation (1.7) leaves the $\mathcal{L}$ unchanged. Let us now add a mass term to Equation (1.6):

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} m^2 A_\mu A^\mu \,. \tag{1.8}$$

Applying the transformation to the second term violates the local gauge invariance, suggesting that U(1) gauge invariance requires the photon to be massless. Referencing Tables 1.2 and 1.3, the photon Lagrangian can be extended by adding a single complex scalar field with charge $-e$, which couples to the photon

$$\mathcal{L}_\gamma = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + (\boldsymbol{D}_\mu \phi)^\dagger (\boldsymbol{D}_\mu \phi) - V(|\phi|) \,, \tag{1.9}$$

where $\boldsymbol{D}_\mu$ is the covariant derivative from Equation (1.3) for the photon field with a coupling constant $-e$. The potential $V(|\phi|)$ is chosen such that it is invariant under global U(1) rotations, $\phi \to e^{i\theta}\phi$, and under local gauge transformations

$$V(|\phi|) = -\mu^2 \phi^* \phi + \lambda (\phi^* \phi)^2 \,, \tag{1.10}$$

where μ and λ are arbitrary parameters with the caveat that $\lambda > 0$, this ensures that the potential energy is bounded from below. Equation (1.9) now has two possibilities: if $\mu^2 < 0$, the Lagrangian symmetries are preserved as the state with the lowest energy $\phi = 0$. This is quantum electrodynamics (QED) with a massless photon and a charged scalar field ϕ with mass μ. However, a very different theory is revealed when $\mu^2 > 0$, the minimum energy state is not at $|\phi| = 0$ but at some other value called vacuum

expectation value (VEV). The VEV for Equation (1.9) for $\mu^2 > 0$ [1]

$$|\phi| = \phi_0 = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{\upsilon}{\sqrt{2}}\,. \tag{1.11}$$

Note that ϕ is a complex field; thus, there is an arbitrary choice as to which direction the vacuum is chosen. By convention, the vacuum is chosen to lie along the direction of the real part of ϕ:

$$\phi = \frac{1}{\sqrt{2}}(\upsilon + h)e^{\mathrm{i}\chi/\upsilon}\,, \tag{1.12}$$

where h and χ are real fields that represent deviations from the modulus and phase of the vacuum field, respectively [7]. Substituting Equation (1.12) into $\mathcal{L}_\gamma$ (excluding cubic and higher in the fields) gives

$$\begin{aligned}\mathcal{L}_\gamma = &-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - g_1\upsilon A_\mu\partial_\mu\chi + \frac{g_1^2\upsilon^2}{2}A_\mu A^\mu \\ &\frac{1}{2}\left(\partial_\mu h\partial^\mu h + 2\lambda\upsilon^2 h^2\right) + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi + \cdots\,.\end{aligned} \tag{1.13}$$

The terms of this Lagrangian include the kinetic terms from Equation (1.9) together with two kinetic terms for the scalar fields h and χ. Mass values are given to the photon with $m_\gamma = g_1\upsilon$, and the scalar field where $m_h^2 = 2\lambda\upsilon^2$ is the squared mass. Interestingly, there is no mass term for χ, which is consistent with Goldstone's Theorem, meaning that χ is the massless scalar Goldstone boson resultant from the breaking of rotational symmetry. The final term in the equation implies a mixing between the gauge field and χ; this lacks physical interpretation and can be understood by considering the degrees of freedom. Before gauge symmetry breaking, a massless vector field with two polarisation states exists. This configuration provides four degrees of freedom in conjunction with the complex scalar field ϕ. Upon symmetry breaking, the system gains a massive vector field with three polarisation states and two real scalar fields, resulting in five degrees of freedom. One of these degrees of freedom is redundant and can be eliminated. This elimination is achieved by performing an appropriate gauge transformation.

$$\phi \to \phi e^{-\mathrm{i}\chi/\upsilon}\,, \tag{1.14}$$

This particular choice of gauge is referred to as the unitary gauge; applying it to Equation (1.13) gives:

$$\begin{aligned}\mathcal{L}_\gamma = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{g_1 v}{2}A^\mu A_\mu + \frac{1}{2}\partial^\mu h \partial_\mu h - \lambda v^2 h^2 \\ + g_1^2\left(vh + \frac{1}{2}h^2\right)A^\mu A_\mu - \lambda v h^3 - \frac{\lambda}{4}h^4 .\end{aligned} \tag{1.15}$$

Neglecting constants and retaining all non-linear interaction terms, it is evident that the extra degree of freedom χ has been completely removed, leaving four manifest degrees of freedom: the three polarisation states of the vector field and a remaining massive scalar boson h known as the Higgs boson [3, 4, 13, 14].

1.3. The Higgs mechanism: vector bosons

The Higgs mechanism is a fundamental process in the Standard Model of particle physics that explains how certain gauge bosons acquire mass while preserving gauge invariance. In the context of electroweak interactions, the Higgs field undergoes spontaneous symmetry breaking, leading to the generation of mass for the $W^\pm$ and Z bosons while leaving the photon massless. This mechanism, first proposed by Higgs, Brout and Englert [3, 4] in 1964, is a cornerstone of modern particle physics and was experimentally confirmed with the discovery of the Higgs boson at the LHC in 2012 [13, 14].

Considering the ground state behaviour of the Lagrangian with an invariant potential spontaneously breaks the local electroweak symmetry SU(2) $\times$ U(1)$_Y$ to $U(1)_{\text{EM}}$. This allows for the introduction of Equation (1.12) as the magnitude of a scalar field $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, with $\chi = 0$, known as the Higgs doublet $\mathbf{\Phi}$. The kinematic terms of the Lagrangian can now be written as [15]

$$\mathcal{L}_{kin} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}\sum_{a=1}^{3} W^{a\mu\nu}W_{\mu\nu} + (\boldsymbol{D}^\mu\mathbf{\Phi})^\dagger(\boldsymbol{D}_\mu\mathbf{\Phi}) - V(\mathbf{\Phi}) , \tag{1.16}$$

where the covariant derivative acting on the Higgs doublet is defined as:

$$\boldsymbol{D}_\mu\mathbf{\Phi} = \left(\partial_\mu + \mathrm{i}g_2 W_\mu^a \sigma^a + \frac{\mathrm{i}g_1}{2}B_\mu\right)\mathbf{\Phi} . \tag{1.17}$$

The final term, $V(\mathbf{\Phi})$, is Equation (1.10) ($V(|\phi|)$) evaluated in terms of the Higgs doublet; this defines the Higgs potential.

$$V(\mathbf{\Phi}) = -\mu^2 \mathbf{\Phi}^\dagger \mathbf{\Phi} + \lambda(\mathbf{\Phi}^\dagger \mathbf{\Phi})^2 \,. \tag{1.18}$$

Expanding the covariant derivatives from Equation (1.16) results in quadratic terms for the gauge bosons

$$|\boldsymbol{D}_\mu \mathbf{\Phi}|^2 = \frac{\upsilon^2}{8}\left[(g_2 W_\mu^3 - g_1 B_\mu)(g_2 W^{3\mu} - g_1 B^\mu) + 2g_2^2 W_\mu^- W^{+\mu}\right] , \tag{1.19}$$

where the standard convention defines $W^\pm$ as

$$W^{\mu\pm} = \frac{1}{\sqrt{2}}\left(W_1^\mu \pm \mathrm{i} W_2^{\mu\pm}\right) . \tag{1.20}$$

Looking at Equation (1.19), it's clear that spontaneous symmetry breaking results in the mixing of the gauge fields associated with the weak isospin and hypercharge (B_μ, W_μ^3). The mixing terms can be removed by performing a rotation defined by

$$\begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} B_\mu \\ W_\mu^3 \end{bmatrix} , \tag{1.21}$$

$$\begin{aligned} A_\mu &= B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \,, \\ Z_\mu &= W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \,, \end{aligned} \tag{1.22}$$

where θ_W is the weak mixing angle, which follows the relation $\tan\theta_W = g_1/g_2$. Together with Equation (1.20), these equations form a mass basis where A is the massless gauge boson of the electric charge with $W^\pm$ and Z bosons being the charged and neutral vector bosons which, due to spontaneous symmetry breaking, now have mass. The masses of the $W^\pm$ and Z boson are given by

$$m_W = \frac{g_2 \upsilon}{2}, \quad m_Z = m_W \frac{\sqrt{g_1^2 + g_2^2}}{g_2} = \frac{g_2 \upsilon}{2\cos\theta_W} \,. \tag{1.23}$$

The interaction terms between the Higgs boson and massive vector bosons can now be written in the new mass basis:

$$\mathcal{L}_{h,\mathrm{VB}} = \frac{2h}{\upsilon}\left(m_W^2 W^{+\mu}W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu\right) + \left(\frac{h}{\upsilon}\right)^2\left(m_W^2 W^{+\mu}W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu\right). \quad (1.24)$$

1.4. Yukawa's interaction: fermions

To understand the fermion masses in terms of the electroweak symmetry-breaking, it is necessary to introduce the Yukawa interactions and couplings [16, 17]. The Yukawa interactions were first introduced by Hideki Yukawa in 1935 in the context of nuclear forces [16], though their role in fermion masses became clear later with the development of the Standard Model. The Yukawa interaction terms in the Lagrangian describe the coupling between the Higgs and Fermion fields. When the Higgs field acquires a non-zero VEV, it breaks the electroweak symmetry and gives rise to mass terms for the fermions. Starting with the leptonic fields, the Yukawa Lagrangian for the lepton fields interacting with the Higgs field ϕ is given by:

$$\mathcal{L}_{l,\mathrm{Yuk}} = -\sum_i y_i \bar{L}_i \phi E_i + \mathrm{h.c}\,, \quad (1.25)$$

where the abbreviation 'h.c.' indicates the hermitian conjugates of the preceding Yukawa terms, y_f is the Yukawa coupling constant and L, E are the left-handed and right-handed components of the lepton field (consistent with Table 1.2). Upon spontaneous symmetry breaking, the Higgs field acquires a VEV, giving rise to mass terms in the Lagrangian:

$$\mathcal{L}_{l,\mathrm{Yuk}} = -\sum_{\mathrm{L}}\left[m_{\mathrm{L}}\bar{l}_{\mathrm{L}} l_E + \frac{1}{\upsilon}h\bar{l}_{\mathrm{L}} l_E\right] + \mathrm{h.c}\,, \quad (1.26)$$

where the sum is over the lepton generations $l \in \{e, \mu, \tau\}$. Thus, the mass of the lepton is given by:

$$m_{\mathrm{L}} = \frac{y_{\mathrm{L}}\upsilon}{\sqrt{2}}\,. \quad (1.27)$$

This equation shows that the mass of each lepton is proportional to its Yukawa coupling to the Higgs field. Different fermions have different Yukawa couplings, which accounts for, but does not explain, the varying masses of the three lepton generations. The hierarchical structure of lepton masses (with $m_e \ll m_\mu \ll m_\tau$) arises from the fact that the Yukawa

couplings are free parameters in the SM and are not predicted by the theory itself. Their observed values suggest an unexplained hierarchy, hinting at deeper physics beyond the SM, such as flavour symmetries or extra dimensions. While the Yukawa couplings successfully parameterise mass differences, they do not explain why they take on their specific values, which remains an open question in particle physics [18].

Due to their association with the lower components of SU(2), the same procedure can be applied to the down-type quarks (D). The up-type quarks require changing the Higgs doublet via a charge conjugation operation, which reverses the quantum numbers, giving the conjugate Higgs doublet.

$$\tilde{\mathbf{\Phi}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{\Phi}^* = \mathrm{i}\sigma_2 \mathbf{\Phi}^* = \frac{(\upsilon + h)}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} . \tag{1.28}$$

The most general set of mass terms for the quark sector can now be written as

$$\mathcal{L}_{q,\mathrm{Yuk}} = -Y^U_{ij} \bar{Q}^i \tilde{\mathbf{\Phi}} U^j_{\mathrm{R}} - Y^D_{ij} \bar{Q}^i \mathbf{\Phi} D^j_{\mathrm{R}} + h.c\,, \tag{1.29}$$

where Y^u_{ij} and Y^d_{ij} are the Yukawa matrices, which specify the strength of the interaction between the up and down fields with the Higgs boson [5].

1.5. Quantum chromodynamics

Following the review of electroweak symmetry breaking, the resulting vector boson masses, and the mechanism by which fermions acquire mass, it is time to explore the strong interaction described by Quantum Chromodynamics (QCD). At the core of QCD lies the SU(3) group, often referred to as $\mathrm{SU(3)_C}$, where the C stands for the colour charge. This group has non-abelian gauge symmetries, i.e. the group contains elements that do not necessarily commute. This results in a highly non-linear theory, resulting in self-interactions between the gauge bosons. Additionally, unlike the $W^\pm$ and Z^0 gauge bosons of the U(1) and SU(2) gauge groups, which become massive after symmetry breaking, the gauge bosons in QCD, the gluons remain massless.

The gluons are the force carriers of the strong interaction, mediating the binding force between quarks. Eight gluons correspond to the eight generators of the $\mathrm{SU(3)_C}$ group.

This group's non-abelian nature means that gluons themselves carry colour charge and can interact with each other.

To show this, it is necessary to generalise the field strength tensor $F_{\mu\nu}$ from Equation (1.7) to something not invariant under gauge transformations. To this end, consider the commutator of two covariant derivatives acting on an arbitrary electron field:

$$\begin{aligned}\left[\boldsymbol{D}_\mu, \boldsymbol{D}^\mu\right]\Psi_f &= (\partial_\mu + \mathrm{i}eA_\mu)\Psi_f - (\partial_\nu + \mathrm{i}eA_\nu)\Psi_f \\ &= [\partial_\mu\partial_\nu + \mathrm{i}eA_\mu\partial_\nu + \mathrm{i}e(\partial_\mu A_\nu) + \mathrm{i}eA_\nu\partial_\mu - e^2A_\mu A_\nu]\Psi_f - [\mu \leftrightarrow \nu] \\ &= \mathrm{i}e(\partial_\mu A_\nu - \partial_\nu A_\mu)\Psi_f \\ &= \mathrm{i}eF_{\mu\nu}\Psi_f\,.\end{aligned} \tag{1.30}$$

Applying two covariant derivatives to the electron field, reversing their order, and then taking the difference results in multiplying the electron field by the electromagnetic field strength tensor. Additionally, since the electron field under consideration is arbitrary, we can formally define the field strength tensor as $[\boldsymbol{D}_\mu, \boldsymbol{D}_\nu] = \mathrm{i}eF_{\mu\nu}$. Before defining the QCD covariant derivative, a gauge field needs to be defined as

$$\boldsymbol{A}_\mu = A^a_\mu \boldsymbol{T}^a\,, \tag{1.31}$$

where A^a_μ are the gauge fields, and $\boldsymbol{T}^a$ are the generators of the non-Abelian Lie algebra. It is now straightforward to generalise to QCD by reevaluating Equation (1.30) in terms of the new gauge field

$$\begin{aligned}\left[\boldsymbol{D}_\mu, \boldsymbol{D}^\mu\right] &= (\partial_\mu + \mathrm{i}g_3\boldsymbol{A}_\mu) - (\partial_\nu + \mathrm{i}g_3\boldsymbol{A}_\nu) \\ &= [\partial_\mu\partial_\nu + \mathrm{i}g_3(\boldsymbol{A}_\mu\partial_\nu + \boldsymbol{A}_\nu\partial_\mu) + \mathrm{i}g_3(\partial_\mu\boldsymbol{A}_\nu) - g_3^2\boldsymbol{A}_\mu\boldsymbol{A}_\nu] - [\mu \leftrightarrow \nu] \\ &= \mathrm{i}g_3(\partial_\mu\boldsymbol{A}_\nu - \partial_\nu\boldsymbol{A}_\mu - \mathrm{i}g_3[\boldsymbol{A}_\mu, \boldsymbol{A}_\nu]) \\ &= \mathrm{i}g_3\boldsymbol{G}_{\mu\nu}\,,\end{aligned} \tag{1.32}$$

where the QCD field strength tensor is

$$\boldsymbol{G}_{\mu\nu} = \partial_\mu\boldsymbol{A}_\nu - \partial_\nu\boldsymbol{A}_\mu + \mathrm{i}g_3[\boldsymbol{A}_\mu, \boldsymbol{A}_\nu]\,. \tag{1.33}$$

Expanding the field strength tensor in terms of the QCD gauge field in Equation (1.31) and applying the computation relation $[\boldsymbol{T}^a, \boldsymbol{T}^b] = \mathrm{i}f^{abc}\boldsymbol{T}_c$ yields

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_3 f^{abc} A^b_\mu A^c_\nu\,, \tag{1.34}$$

where f^{abc} are the structure constants of the $SU(3)_{\mathrm{C}}$ group. With these terms in mind, the QCD Lagrangian is given by

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \sum_f \bar{\Psi}_f \left(i\gamma^\mu \boldsymbol{D}_\mu - m_f \right) \Psi_f \,, \tag{1.35}$$

where $G^a_{\mu\nu}$ is the gluon field strength tensor, Ψ_f represents the quark fields, and $\boldsymbol{D}_\mu$ is the covariant derivative containing the gluon fields.

One of the most striking features of QCD is the phenomenon of confinement. Unlike the electromagnetic force mediated by photons, quarks and gluons cannot be isolated as free particles; they are always confined within hadrons. This confinement arises from the property of the QCD potential, which increases with distance, preventing the separation of quarks over macroscopic scales.

While Equation (1.35) does not explicitly imply confinement itself, it provides the foundation for confinement to emerge as a non-perturbative phenomenon. The key feature of QCD is the non-Abelian nature of the SU(3) gauge group, which allows gluons to interact with themselves due to the presence of terms like $G^a_{\mu\nu} G^{\mu\nu}_a$. These self-interactions lead to asymptotic freedom, where the strong coupling constant α_s decreases at short distances but increases at long distances.

Asymptotic freedom at very short distances or high energies is a property first discovered by Gross, Wilczek [19], and Politzer [20]. Asymptotic freedom is characterised by the diminishing interaction strength between quarks and gluons as the energy scale increases, thereby allowing the application of perturbative techniques to QCD at high energies. This phenomenon is quantitatively described by the QCD beta function, which governs the running of the strong coupling constant, $\alpha_s = g_3^2/(4\pi)$, with respect to the energy scale, μ. The QCD beta function is given by

$$\beta(\alpha_s) \equiv \mu \frac{d\alpha_s}{d\mu} = -2\alpha_s \left[\left(\frac{\alpha_s}{4\pi} \right) b_0 + \left(\frac{\alpha_s}{4\pi} \right)^2 b_1 + \left(\frac{\alpha_s}{4\pi} \right)^3 b_2 + \mathcal{O}(\alpha_s^4) \right] , \tag{1.36}$$

where the first coefficient b_0 can be considered as the one-loop correction of the renormalisation group equation (RGE) given as [15]

$$b_0 = 11 - \frac{2n_f}{3} \,, \tag{1.37}$$

where n_f is the number of flavours. Crucially, this shows that b_0 is positive, so as long as $n_f < 17$, the beta function will be negative at low orders, which implies that α_s decreases

logarithmically as μ increases. Thus, interactions are strong at low energies and weaken when sufficiently high, leading to asymptotic freedom. The duality between confinement at low energies and asymptotic freedom at high energies is a defining feature of QCD, encapsulating the complex interplay between the strong force's strength across different scales.

The properties of confinement and asymptotic freedom have important phenomenological implications. Asymptotic freedom implies that at sufficiently high energies, such as those encountered in deep inelastic scattering experiments, quarks behave as nearly free particles, allowing perturbative QCD calculations to predict cross-sections and parton distribution functions accurately [19], where the parton is a constituent of a hadron (quark or gluon). This property is crucial for interpreting experimental results from particle colliders, where high-energy processes can be analysed using perturbative techniques [21].

Conversely, confinement explains why quarks and gluons are never observed as free particles but are always bound within hadrons. This leads to hadronization, where quarks and gluons produced in high-energy collisions form hadrons before they can be detected [22]. The inability to isolate quarks also impacts the formation and decay of hadronic states, contributing to the mass spectrum of hadrons and the dynamics of bound states, such as mesons and baryons [23]. Additionally, confinement is essential for understanding the phase transition between hadronic matter and the quark-gluon plasma, which occurs at extremely high temperatures or densities, as studied in heavy-ion collisions and early universe cosmology [24].

Quantum Chromodynamics, with its massless gauge bosons and non-Abelian symmetry, forms a cornerstone of our understanding of the strong force. The interplay between quarks and gluons under the principles of confinement and asymptotic freedom explains the binding of quarks within protons and neutrons but also provides a framework for exploring the deeper structure of matter. This section lays the groundwork for a deeper exploration of QCD, its experimental validations, and its implications for particle physics.

1.6. Renormalization

The renormalisation group (RG) is an important feature of QFT that provides insights into how physical systems behave at different energy scales. This concept originated from the need to address infinities that arise in QFT calculations. When calculating specific

quantities, such as the mass and charge of particles, integrals often diverge at high energies, leading to nonphysical infinite results. The renormalisation process systematically removes these infinities by absorbing them into redefined (or "renormalised") physical quantities, leaving finite, physically meaningful predictions. The RG extends this concept by studying how these renormalised quantities change as the energy scale at which the theory is applied varies[1].

At its core, the RG approach is rooted in the idea that physical laws might exhibit different behaviours depending on the scale at which they are examined. This is often called "running" coupling constants evolving with the energy scale. In QED, for instance, the fine-structure constant $\alpha = g_1^2/(4\pi)$ is not a fixed quantity but depends logarithmically on the momentum transfer q^2. This running of α is described by the beta function $\beta(\alpha)$ (Equation (1.36) evaluated for α), which quantifies how a coupling constant changes with the logarithm of the energy scale [1]. In QED, the beta function is positive, indicating that the coupling constant increases with energy, leading to the phenomenon known as the Landau pole [25], where the theory becomes non-perturbative at extremely high energies. In QCD, the beta function leads to asymptotic freedom. As particles interact at higher energies or equivalently at shorter distances, the strong force becomes weaker, allowing quarks to behave almost like free particles. This discovery was crucial in explaining the results of high-energy particle collisions and earned Gross, Politzer, and Wilczek the Nobel Prize in Physics in 2004 [19, 20].

Another key application of the RG is in the study of phase transitions. In statistical mechanics, which strongly parallels particle physics, when a system nears a critical point, it exhibits self-similarity across different scales. The RG flow quantitatively captures this concept. This flow describes how physical systems transform under changes in scale, revealing the universal behaviour of phase transitions. The RG also provides a framework for understanding phase transitions in particle physics, particularly in systems where symmetry breaking occurs. In the context of electroweak symmetry breaking, the RG flow can describe how the Higgs field's vacuum expectation value evolves with energy, influencing the masses of gauge bosons and fermions. Moreover, the RG approach is essential in effective field theory, where it guides the integration of heavy degrees of freedom, leading to a low-energy effective theory with appropriately renormalised parameters [26].

1.7. Cross-sections and scattering amplitudes

Understanding the interaction between particles requires a thorough analysis of scattering processes. These interactions can be quantified using cross-sections and scattering amplitudes, which provide critical insights into the probabilities of various interaction outcomes. The S-matrix (scattering matrix) is a central concept in this analysis, encapsulating the transition probabilities between different quantum states.

The S-matrix formalism provides a comprehensive framework for describing particle interactions. It relates the initial state of a system, comprising particles A and B, to the final state, consisting of particles with specific momenta. Mathematically, the S-matrix element S_{fi} between an initial state $|i\rangle$ and a final state $|f\rangle$ is given by:

$$S_{fi} = \langle f|S|i\rangle \,, \tag{1.38}$$

where S is the operator that governs the system's evolution from the initial to the final state. To extract physically meaningful information, it is useful to consider the transfer matrix (or T-matrix), which isolates the interacting part of the S-matrix [15]:

$$\langle f|S|i\rangle = \mathbb{1} + \mathrm{i}T \,, \tag{1.39}$$

where $\mathbb{1}$ is the identity matrix, i is the initial state and i is the imaginary number. The matrix element $\langle f|T|i\rangle$ is directly related to the scattering amplitude $\mathcal{M}$, which in turn is used to compute cross-sections:

$$\langle f|S - \mathbb{1}|i\rangle = \mathrm{i}T = \mathrm{i}\mathcal{M}(2\pi)^4\delta^4(\textstyle\sum p_f - \sum p_i) \,. \tag{1.40}$$

For the scattering of two particles, A and B, into a final state with particles of specific momenta, the scattering amplitude $\mathcal{M}$ can be derived from the matrix elements of the transfer matrix. Given initial momenta $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ and final momenta $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$, the scattering amplitude is given by:

$$\mathcal{M}(A + B \to 1 + 2 + \cdots + n) = \langle \boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n|T|\boldsymbol{p}_A, \boldsymbol{p}_B\rangle \,. \tag{1.41}$$

This scattering amplitude $\mathcal{M}$ encapsulates the interaction dynamics. It is computed using Feynman rules [27] derived from the underlying quantum field theory, such as QCD or the electroweak theory [28]. The differential cross-section measures the likelihood of scattering into a particular final state and phase-space point at a certain luminosity. It

is directly related to the scattering amplitude by:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \propto |\mathcal{M}|^2 \,, \tag{1.42}$$

where $d\Omega$ represents the differential solid angle. The observable of choice is the cross-section of the initial state evolving into a given set of "final state" particles. For such a $2 \to N$ transition, the full differential cross-section for the scattering of two particles A and B, with energies E, velocity v and momenta p, into a final state with particles of momenta p_i is given by

$$\mathrm{d}\sigma = \frac{1}{2E_A 2E_B |v_A - v_B|} |\mathcal{M}(p_A, p_B \to \{p_i\}|^2 \mathrm{d}\Pi_{\mathrm{LIPS}} \,, \tag{1.43}$$

where $\mathrm{d}\Pi_{\mathrm{LIPS}}$ is the Lorentz-Invariant Phase-Space

$$\mathrm{d}\Pi_{\mathrm{LIPS}} = (2\pi)^4 \delta^4(p_a + p_b - \sum p_i) \left(\prod_i \frac{\mathrm{d}{p_i}^3}{(2\pi)^3} \frac{1}{2E_i} \right) . \tag{1.44}$$

Integrating this over all possible final states yields the total cross-section, a key observable in experiments.

When dealing with proton-proton collisions, it is essential to account for the internal structure. Unlike fundamental particles, protons are composite particles of quarks, antiquarks, and gluons. As discussed, these constituents interact according to Quantum Chromodynamics (QCD) principles. The key to simplifying the calculations for these collisions lies in the factorisation theorem, which allows us to separate the complex proton structure from the fundamental interactions of the partons [15]. The interactions of the proton's constituents, quarks and gluons, collectively known as partons, allow the overall calculation to be factorised into largely independent components: the hard scattering process, calculable using perturbative QCD, and the non-perturbative parton distribution functions (PDFs). The PDF is a critical element in this extended calculation as it describes the probability of finding a parton with a specific fraction of the proton's momentum at a given energy scale. These functions encapsulate the non-perturbative aspects of QCD, which cannot be calculated directly but are extracted from experimental data. Mathematically, the cross-section for a proton-proton collision can be expressed as

$$\sigma(pp \to X) = \sum_{a,b} \int dx_a \, dx_b \, f_a(x_a, Q^2) \, f_b(x_b, Q^2) \, \hat{\sigma}(ab \to X) \,, \tag{1.45}$$

where $f_a(x_a, Q^2)$ and $f_b(x_b, Q^2)$ are the PDFs for partons a and b inside the protons, x_a and x_b are the momentum fractions carried by the partons, Q^2 is the energy scale, and $\hat{\sigma}(ab \to X)$ is the parton-level cross-section. PDFs are often functions of both the momentum fraction x and a factorisation scale μ_F, which is typically linked to a characteristic energy scale Q in the scattering process where the PDFs are applied. This scale varies depending on the context, requiring PDFs to be provided across the two-dimensional parameter space (x, Q^2), where the squared energy scale Q^2 is conventionally used. Several global collaborations generate these distributions and offer computer codes that compute PDF values for any given input. This is often achieved by reading a precomputed PDF grid containing values at discrete points in the (x, Q^2) plane and using interpolation algorithms to estimate values between these points [7].

The factorisation theorem assumes that the cross-section of the entire proton-proton collision can be decomposed into a convolution of PDFs and the hard scattering cross-section of the partons [15]. This theorem greatly simplifies the calculations, allowing us to handle the proton's complex internal dynamics separately from the high-energy interactions of its constituents. The factorised form allows physicists to use perturbative QCD to calculate the parton-level cross-sections while using empirical data to parameterise the PDFs.

This factorisation approach has profound implications for particle physics research. It enables precise predictions for outcomes of high-energy proton collisions, such as those occurring in the Large Hadron Collider (LHC). Understanding these interactions at the parton level allows researchers to probe deeper into the fundamental forces and particles that constitute the universe. It also aids in the search for new physics beyond the Standard Model by providing a robust framework to compare theoretical predictions with experimental results.

1.8. Current status of the SM

The Standard Model is one of the most successful scientific theories, accurately predicting many high-precision experimental observations. It has not only described known experimental results but also made significant predictions, such as the existence of a third generation of quarks and leptons [29] and the Higgs boson [3, 4], which were later confirmed through experiments [13, 14]. The Standard Model is also theoretically

appealing as it represents the most comprehensive theory possible, given its symmetry group and particle content [30].

Despite its success, the Standard Model has notable limitations. One of the most significant shortcomings is its inability to incorporate gravity into a unified framework alongside the strong and electroweak forces. This limitation suggests that the Standard Model may be a low-energy effective field theory, similar to Fermi's theory of weak interactions, which could be a limiting case of a more complete theory that includes quantum gravity [31]. Developing a consistent theory of quantum gravity that also encompasses the particle and symmetry structure of the Standard Model remains an unresolved challenge in modern physics [32].

One of the problems neglected in the previous discussion of fermions is that, in practice, the neutrinos observed in nature possess mass, as evidenced by neutrino oscillation experiments [33]. However, neutrinos are considered massless within the Standard Model due to the absence of right-handed neutrinos [31]. Without these right-handed counterparts, neutrino masses cannot be generated through the Yukawa interactions in the same way as quarks and charged leptons [34]. Additionally, the observed neutrino masses and their associated Yukawa couplings are significantly smaller than those of quarks and charged leptons [33]. More broadly, the origin of the mass values for all fundamental particles in the Standard Model remains unexplained. Currently, these masses are treated as external parameters, encapsulated in the Yukawa couplings, and must be determined experimentally [35]. While the described mechanism accounts for the masses of fundamental fermions within the Standard Model, it does not explain the origins of the Yukawa coupling constants themselves [35]. These discrepancies suggest the existence of a mechanism beyond the Standard Model that generates neutrino masses and potentially provides an underlying principle explaining the pronounced mass hierarchy among the three generations of quarks and leptons [34].

The discovery of the Higgs boson at a mass of 125.11 ± 0.11 GeV [13, 14, 36] was not without complications, one of which is the hierarchy problem. As a scalar field in the theory, the physical mass of the Higgs exhibits quadratic sensitivity to the cutoff scale of the Standard Model, which is treated as an effective field theory. This implies that at large cutoff scales, even small changes in the input parameters can result in substantial variations in the Higgs mass as predicted at collider scales. Therefore, to achieve the observed Higgs mass, the parameters at high scales must be finely tuned, giving an unexpected dependency of low-energy physics on the high-energy details of the theory. Furthermore, the spontaneously broken $\mathrm{SU}(2)\times\mathrm{U}(1)$ symmetry underpinning

the Higgs mechanism lacks a dynamical explanation for why this symmetry is broken. The mechanism is introduced without intrinsic justification within the theory, although this may be more fundamentally linked to the philosophy of representation between mathematics and reality [37]

One major problem raised by astrophysical observations is that the Standard Model accounts for only 4.9% of the universe's energy density. The missing energy is split between dark energy, at 68.3%, and dark matter, at 26.8% [38]. Dark matter is thought to be comprised of non-baryonic matter, which is believed to consist of electrically neutral, weakly interacting particles. Consequently, extensions of the Standard Model that propose potential dark matter candidates and can be tested at collider energies are of great interest and are being studied intensively.

This gap in understanding has led to ongoing research into extensions of the Standard Model, such as supersymmetry or theories involving extra dimensions. These extensions aim to address the deeper questions surrounding the origin of particle masses and provide a more fundamental understanding of the observed mass hierarchies.

Chapter 2.

Supersymmetry

Supersymmetry (SUSY) is a theoretical framework that extends the Standard Model of particle physics by positing a symmetry between fermions and bosons such that a corresponding boson exists for every fermion and vice versa. This elegant symmetry aims to address some of the limitations of the Standard Model and offers potential solutions to problems mentioned in the previous chapter.

This chapter covers the historical development of supersymmetry, tracing its origins from early theoretical proposals to its formalisation in the 1970s. It examines the key motivations behind the development of SUSY, such as the hierarchy problem, a natural candidate for dark-matter and a potential pathway towards unifying the fundamental forces.

The chapter also explores the theoretical constructs of supersymmetry, including the algebraic structures that define SUSY transformations and the mathematical framework that allows for the extension of space-time symmetries. Key concepts such as superpartners, superfields, and superspace will be discussed to understand the theory comprehensively.

Various models within supersymmetry are presented, ranging from the Minimal Supersymmetric Standard Model (MSSM) to more complex extensions. The MSSM is presented as an example SUSY extension of the Standard Model, introducing superpartners for all known particles and providing mechanisms for electroweak symmetry breaking. More intricate models, such as those involving extra dimensions or higher symmetries, are also considered, illustrating the rich landscape of possibilities within supersymmetric theories.

Throughout this chapter and beyond, there will be references to experimental results from direct measurements and searches; these terms refer to two distinct but complementary approaches used to explore beyond the standard model (BSM). "Search" refers to the systematic effort to discover new particles or phenomena predicted by a BSM theory, such as SUSY. These searches involve analysing data for unexpected signatures, such as missing energy or new particle decays, indicating the presence of supersymmetric particles like neutralinos or gluinos. On the other hand, "measurement" pertains to the precise quantification of known physical processes, such as Standard Model interactions, to detect deviations from expected values. Precise measurements of particle masses, cross-sections, and branching ratios can also be useful for indirectly testing SUSY models, especially in scenarios where direct evidence is elusive.

Through this exploration, the chapter aims to provide a comprehensive overview of supersymmetry, its theoretical underpinnings, and its significance in the broader context beyond the Standard Model (BSM) particle physics.

2.1. An introduction to SUSY

The previous chapter demonstrated just how central the concept of symmetry was in guiding the construction of the SM. As it turns out, the SM splitting into component products of the local and global symmetries is a characteristic fundamental to group theories in general and provides valuable constraints. The Coleman-Mandula theorem states that for a relativistic quantum field theory with well-defined operators (i.e. an operator that gives the same result when the representation of the input is changed without changing the value of the input), a finite number of particle types and a non-trivial S-matrix, the symmetries of the S-matrix must be a direct product of the Poincaré and internal symmetries [39, 40]. The theorem, formulated by Sidney Coleman and Jeffrey Mandula in 1967, is a "no-go" theorem, which is one that states that a particular formulation is not physically possible. The theorem implies that any attempt to unify space-time symmetries with internal symmetries into a larger symmetry group must fail, except in the trivial case where they form a direct product. This was a significant obstacle to developing unified theories that combined these different types of symmetries.

The concept of supersymmetry emerged in the early 1970s, primarily within the context of string theory. In 1974, early contributions by Julius Wess and Bruno Zumino led to the formulation of the first four-dimensional supersymmetric field theory, now

known as the Wess-Zumino model [41]. While the theorem prohibits the unification of internal symmetries with space-time symmetries in a non-trivial way, it does not apply to graded Lie algebras. A graded Lie algebra is a Lie algebra decomposed into a direct sum of vector spaces indexed by integers, called grades. The Lie bracket operation respects this grading, meaning the bracket of elements from grades p and q belongs to grade $p+q$. This structure allows for the study of symmetries that are more complex than those in ordinary Lie algebras. These extended symmetries are characterised by algebras encompassing commutation and anti-commutation relations. Specifically, symmetry generators in this framework are categorised as bosonic or fermionic, depending on their statistical properties. The theorem's restrictions remain intact for relations involving solely bosonic generators. However, when fermionic generators are introduced, the situation becomes more nuanced. The Haag-Łopuszański-Sohnius theorem [42] provides crucial guidance here: It dictates that fermionic generators must transform consistently under both the internal symmetry group and the Lorentz group, adhering to specific commutation and anti-commutation relations. For realistic theories incorporating chiral fermions, where the left-handed and right-handed components of fermions transform differently under the gauge group, parity-violating interactions become possible. In such contexts, the Haag-Łopuszański-Sohnius theorem [42] necessitates that the supercharge generators Q and $Q^\dagger$ that generate transformations between bosonic and fermionic states adhere to a specific algebra. This algebra consists of both anti-commutation and commutation relations,

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle\,, \quad |Q|\text{Fermion}\rangle = |\text{Boson}\rangle\,. \tag{2.1}$$

The anti-commutation relations reflect the fermionic nature of the supersymmetry generators are as follows:

$$\begin{aligned} \{Q_i, Q_j^\dagger\} &= 2\sigma_{ij}^\mu P_\mu \\ \{Q_i, Q_j\} &= \{Q_i^\dagger, Q_j^\dagger\} = 0\,, \end{aligned} \tag{2.2}$$

where σ_{ij}^μ are the Pauli matrices and P_μ represents the four-momentum operator. The commutation relations are

$$\begin{aligned} [Q_i, P_\mu] &= [Q_i^\dagger, P_\mu] = 0 \\ [Q_i, J_{\mu\nu}] &= (\sigma_{\mu\nu})_i^j Q_j \\ [Q_i^\dagger, J_{\mu\nu}] &= (\sigma_{\mu\nu})_i^j Q_j^\dagger\,, \end{aligned} \tag{2.3}$$

where $J_{\mu\nu}$ are the generators of the Lorentz group. These commutation relations ensure that the supersymmetry generators transform appropriately under Lorentz transformations and commute with the momentum operator. It is worth noting that the anti-commutation relation involving the four-momentum P_μ directly ties the supersymmetry transformations to space-time translations. Additionally, the commutation relations with the Lorentz generators $J_{\mu\nu}$ ensure that the supersymmetric partners of particles transform correctly under rotations and boosts. This is crucial for maintaining the covariance of the theory and ensuring that physical observables, such as scattering amplitudes, behave consistently under Lorentz transformations [43, 44].

In a supersymmetric theory, single-particle states are organised into irreducible representations of the supersymmetry algebra, known as supermultiplets. Each supermultiplet encompasses both fermion and boson states, referred to as superpartners. By definition, if $|\Omega\rangle$ and $|\Omega'\rangle$ belong to the same supermultiplet, then $|\Omega'\rangle$ can be generated by applying a combination of the supersymmetry generators Q_i and $Q_i^\dagger$ to $|\Omega\rangle$, modulo a space-time translation or rotation.

The squared mass operator, P^2, commutes with the supersymmetry generators Q_i and $Q_i^\dagger$, as well as with all space-time translation and rotation operators. Consequently, particles within the same irreducible supermultiplet must share the same eigenvalues of P^2, implying that they have identical masses. This mass degeneracy is a significant result, as it ensures that a supermultiplet's bosonic and fermionic components are indistinguishable in their mass. However, since supersymmetry must be a broken symmetry in nature (given the absence of observed mass-degenerate superpartners for known particles), the actual masses of superpartners in realistic models are split, typically by mechanisms associated with supersymmetry breaking. Understanding these mechanisms is crucial for predicting the mass spectrum of superparticles.

The supersymmetry generators Q_i and $Q_i^\dagger$ also commute with the generators of gauge transformations; therefore, particles grouped in the same supermultiplet must transform identically under gauge transformations. Consequently, all particles within a given supermultiplet must be in the same representation of the gauge group. This means that these particles share identical gauge quantum numbers. For instance, they must possess the same electric charge, reflecting their transformation properties under the $U(1)$ gauge symmetry. Similarly, they must have identical weak isospin values, corresponding to $SU(2)_L$; if they are subject to the strong interaction, they must share the same colour charge, as dictated by the $SU(3)_C$ of quantum chromodynamics (QCD).

This requirement ensures consistent transformation behaviour under the respective gauge groups and maintains the internal symmetry structure of the supersymmetric theory. The identical gauge quantum numbers within a supermultiplet highlight the connection between the gauge symmetry and supersymmetry, enforcing stringent constraints on the particles' properties and interactions. This consistency is crucial for the coherence and predictive power of supersymmetric models as it guarantees that the introduction of superpartners does not violate the established gauge symmetry dynamics of the SM.

In practical terms, understanding the algebra of Q and $Q^\dagger$ helps construct realistic supersymmetric models that can be tested experimentally. For instance, the relations guide the development of supersymmetric extensions of the Standard Model, such as the Minimal Supersymmetric Standard Model (MSSM). These models predict many new particles, including neutralinos, charginos, and squarks, whose interactions and decays are governed by supersymmetric algebra.

2.2. Supersymmetric models

It is beneficial to outline a brief description of how a supersymmetric theory is constructed to gain a qualitative understanding of the potential interactions. This section will cover the fundamental principles and mechanisms underlying the construction of a supersymmetric theory and provide examples of the most widely studied supersymmetric extensions.

2.2.1. The superpotential

The superpotential represents a fundamental concept in supersymmetric theories, encapsulating the interactions among chiral superfields. It is characterised as a holomorphic function, meaning it is complex and differentiable and depends solely on the chiral superfields, excluding their conjugates. This characteristic is crucial as it plays a significant role in defining both the dynamics and the overall structure of the theory [45, 46]. The superpotential W can be generally expressed in the form

$$W = L^i\phi_i + \frac{1}{2}M^{ij}\phi_i\phi_j + \frac{1}{6}y^{ijk}\phi_i\phi_j\phi_k\,. \tag{2.4}$$

In this expression, L^i denotes parameters with dimensions of $[\text{mass}]^2$, M_{ij} represents a symmetric mass matrix for the fermion fields, and y^{ijk} are the Yukawa couplings

associated with the scalar fields ϕ_i. These parameters and coefficients delineate the interaction terms and mass structures within supersymmetric models [43].

The holomorphic property of the superpotential is essential for preserving supersymmetry after spontaneous symmetry breaking. This is because the potential derived from the superpotential remains dependent solely on the scalar components of the chiral superfields. By defining the interactions among the chiral superfields, the superpotential significantly influences the mass spectrum of the particles within the theory.[47]. The superpotential induces interaction terms in the Lagrangian as

$$\mathcal{L}_{\text{int}} \supset -\frac{1}{2}W^{ij}\phi_i\phi_j + |W^i W_i^*|\,, \tag{2.5}$$

where the terms W^{ij} and W^i are polynomials in the scalar fields, with degrees 1 and 2 respectively, and are defined as

$$W^{ij} = \frac{\delta^2 W}{\delta\phi_i, \delta\phi_j}\,, \qquad W^i = \frac{\delta W}{\delta\phi_i}\,. \tag{2.6}$$

Evaluating these derivatives in terms of the superpotential defined in Equation (2.4) we obtain

$$\begin{aligned} \mathcal{L}_{\text{int}} \supset W^{ij} &= M^{ij} + y^{ijk}\phi_k \\ \mathcal{L}_{\text{int}} \supset |W^i W_i^*| &= |M|^2\phi^{*k}\phi_j + \frac{1}{2}M^{kn}y^*_{jin}\phi_k\phi^{*j}\phi^{*i} \\ &\quad + \frac{1}{2}M^*_{kn}y^{jin}\phi^{*k}\phi_j\phi_i + \frac{1}{4}y^{kjn}y^*_{iln}\phi_k\phi_j\phi^{*i}\phi^{*l}\,. \end{aligned} \tag{2.7}$$

These represent the basic renormalisable set of interaction terms available. Supersymmetric field theories exhibit several fundamental types of interactions. These include mass-like vertices for both fermions and bosons, which involve two fermionic fields and one bosonic field, and vertices consisting of three and four scalar fields. An example of such an interaction is the Yukawa-type interaction in Equation (2.7) and found in the SM.

A superfield is a composite entity that integrates all the bosonic, fermionic, and auxiliary fields associated with a given supermultiplet. This is illustrated by Φ_i, which encompasses components ϕ_i and W_i. This concept is analogous to describing a weak isospin doublet or a colour triplet using a multi-component field. The gauge quantum numbers and the mass dimension of a chiral superfield match those of its scalar component.

In the superfield formulation, the expression used is

$$W = L^i \Phi_i + \frac{1}{2} M^{ij} \Phi_i \Phi_j + \frac{1}{6} y^{ijk} \Phi_i \Phi_j \Phi_k \,. \tag{2.8}$$

2.2.2. The gauge sector

In supersymmetry, the gauge sector consists of vector superfields that correspond to the gauge bosons present in the Standard Model. A vector superfield V includes three main components: a gauge boson, its fermionic superpartner called the gaugino and an auxiliary field. These components combine to form a superfield that transforms appropriately under supersymmetry transformations.

The Lagrangian for the gauge sector incorporates kinetic terms for the fields and interaction terms that maintain gauge invariance. The kinetic terms describe the propagation and self-interaction of the gauge fields, while the interaction terms ensure that the theory respects the symmetry principles. A central feature of this formulation is the gauge kinetic function $f_{ab}(\Phi)$, a holomorphic function of the chiral superfields Φ, where the ab indices correspond to the different gauge group factors and are associated with the generators of the gauge group. Specifically, in a theory with a gauge group G that can be decomposed into several simple or abelian factors (such as $G = G_1 \times G_2 \times \ldots \times G_n$), the indices a and b label the components of the field strengths and gauge fields corresponding to different factors of this gauge group. The gauge kinetic function f_{ab} is typically a matrix-valued function, where each entry $f_{ab}(\Phi)$ determines the coupling between the gauge fields associated with generators of the gauge group. The holomorphicity of f_{ab} implies that it depends on the chiral superfields Φ in a way that preserves the supersymmetry of the theory [48–50]. The gauge kinetic function determines the coupling of the gauge fields and their interactions. The Lagrangian for the gauge sector is given by [44]

$$\mathcal{L}_{gauge} = \int d^2\theta \, f_{ab}(\Phi) W^a W^b + \text{h.c.}\,, \tag{2.9}$$

where W^a represents the gauge field strength superfields, which describe the field strengths of the gauge bosons along with their superpartners. The integral is defined over $d^2\theta$ which are Grassmann coordinates. In the context of supersymmetric theories, Grassmann coordinates are used to extend conventional spacetime into a higher-dimensional space known as superspace. Superspace combines ordinary spacetime coordinates with additional Grassmann coordinates to incorporate supersymmetry. The coordinates, denoted

typically as θ and $\bar{\theta}$, are anticommuting variables, meaning they satisfy the property $\theta\theta' = -\theta'\theta$. This anticommutative nature is essential for representing fermionic degrees of freedom within the supersymmetric framework [51]. Superspace is therefore, a space that includes both the usual four spacetime dimensions x^μ (where μ runs over the spacetime indices) and the Grassmann coordinates θ and $\bar{\theta}$. A point in superspace is described by the coordinates $(x^\mu, \theta, \bar{\theta})$. In this setting, supersymmetric fields, or superfields, are functions of both the spacetime coordinates and the Grassmann coordinates. These superfields encapsulate the components of a supersymmetric multiplet, including both bosonic fields (like gauge fields) and their fermionic superpartners.

The integral over the Grassmann coordinates is used to select specific components of a superfield, typically the highest component, which corresponds to the physical field content of the theory after integrating out the Grassmann variables. For our case of the gauge field strength superfield W^a, the $d^2\theta$ integral effectively isolates the component that describes the gauge boson field strengths and their supersymmetric partners, contributing to the Lagrangian of the theory. Thus, the form of the gauge kinetic function $f_{ab}(\Phi)$ determines how the gauge fields couple to the matter fields in the theory. This function can vary depending on the specific model of supersymmetry being considered, leading to different physical implications and predictions. In many models, f_{ab} is often considered a constant or a function of the scalar components of the chiral superfields, which can result in different gauge interactions and the running of coupling constants. An example of this is in $\mathcal{N} = 1$ Super Yang-Mills theory where $f_{ab}(\Phi) = \tau$, with τ being the natural combination of the gauge coupling to θ [52].

Including auxiliary fields in the vector superfield V simplifies the supersymmetric Lagrangian. These auxiliary fields do not have kinetic terms. They can be eliminated through their equations of motion, resulting in the standard form of the Lagrangian with the gauge fields and gauginos.

Overall, the structure of the gauge sector in supersymmetry is designed to extend the Standard Model by incorporating superpartners for the gauge bosons. Thus, it maintains consistency with the supersymmetric framework while potentially addressing some of its limitations and unanswered questions.

2.2.3. Supersymmetry breaking

If supersymmetry were a realised, unbroken symmetry, the superpartners of known particles would have identical masses to the particles of the Standard Model; however, this is not observed experimentally. This necessitates SUSY breaking, introducing a characteristic energy scale where superpartners acquire masses different from their SM counterparts. The specific scale of SUSY breaking depends on the mechanism employed. In gravity-mediated SUSY breaking, the soft SUSY-breaking scale is typically around the Planck scale M_{Pl}, suppressed by a mediation factor, leading to soft terms at $\mathcal{O}(\text{TeV})$ [53]. In gauge-mediated SUSY breaking, the scale can be much lower, with soft masses arising from loop effects of messenger fields coupling to the SUSY-breaking sector [54]. While these mechanisms constrain the mass range of superpartners, they do not yield a single, unique value for each mass; rather, they depend on model parameters such as coupling constants, mediation scales, and details of SUSY breaking. Therefore, while SUSY models predict that new particle masses should be within a certain energy range—typically near the electroweak scale for naturalness reasons—there is no strict lower or upper bound beyond theoretical consistency and experimental limits. This flexibility has been a challenge for direct experimental detection, as different scenarios can push the SUSY particle masses to multi-TeV scales or even higher, beyond current collider reach [43]. Hence, while SUSY breaking determines characteristic scales, it does not impose a single fixed mass value for new particles.

In supersymmetric theories, the P^2 operator, which corresponds to the mass, commutes with all elements of the supersymmetry algebra. This characteristic allows the mass to serve as a label for supermultiplets, similar to how it functions in the Poincaré symmetry context. If supersymmetry were a realised symmetry, the superpartners of known particles would have identical masses to the particles of the Standard Model. Consequently, these superpartners would have already been detected in experiments. However, since such particles have not been observed, it indicates that supersymmetry must be broken at energy scales currently accessible to collider experiments.

This supersymmetry breaking can be analogous to the spontaneous symmetry breaking observed in the Standard Model. In section 1.2, spontaneous symmetry breaking was described as when a physical system, initially in a state of symmetry, ultimately transitions to an asymmetric state without any external influence. Applying this concept to supersymmetry implies that the theory can remain fundamentally supersymmetric, even if the observed phenomena do not exhibit supersymmetry.

To preserve the beneficial properties of supersymmetric theories, such as the cancellation of quadratic divergences in radiative corrections to scalar masses, the breaking of supersymmetry must occur so that the resulting mass differences between particles and their superpartners are relatively small. This condition ensures the breaking is "soft", meaning minimal mass splitting. The set of soft supersymmetry-breaking terms in the Lagrangian of a general theory are

$$\mathcal{L}_{soft} \supset -\frac{1}{2}M_a\lambda^a\lambda^a - \frac{1}{6}a^{ikj}\phi_i\phi_j\phi_k - \frac{1}{2}b^{ij}\phi_i\phi_j - t^i\phi_i - (m^2)^i_j\phi^{*j}\phi i\,. \tag{2.10}$$

The terms include explicit soft gaugino masses M_a, tri-linear couplings denoted as a^{ijk}, and both holomorphic and non-holomorphic scalar mass terms represented by $(m^2)^i_j$ and b^{ij}, respectively. The coupling term t^i can only be realised for singlets of the internal symmetries [43].

At the scales probed by current colliders, spontaneous and soft supersymmetry breaking can be described using effective operators that explicitly break supersymmetry. These effective operators encapsulate the low-energy consequences of the underlying supersymmetry-breaking dynamics. Consequently, they provide a phenomenological description that can be tested against experimental data, allowing for exploring the supersymmetric parameter space and searching for potential signs of supersymmetry in collider experiments.

The literature on supersymmetry breaking identifies three main modes, each characterised by a distinct procedure. These modes include Planck-Scale-Mediated, Gauge-Mediated, Extra-Dimensional and Anomaly-Mediated; they can be summarised as follows:

Planck-Scale-Mediated Breaking

Often referred to as gravity-mediated models, assume that supersymmetry breaking occurs in a hidden sector and is communicated to the visible sector through gravitational interactions [55]. These models typically operate at the Planck scale ($M_{\text{Pl}} \approx 10^{19}$ GeV), where gravitational interactions become significant. The primary mechanism involves the mediation of supersymmetry breaking through Planck-suppressed operators in the supergravity Lagrangian, leading to soft supersymmetry breaking terms in the observable sector [53]. This mediation generates soft masses for superpartners proportional to the gravitino mass, typically in the range of hundreds of GeV to a few TeV [56]. These models, including mSUGRA (minimal supergravity), posit that all fields feel supersymmetry-

breaking effects universally due to the gravitational force [57]. The scalar masses, gaugino masses, and tri-linear couplings in these models are typically of the same order of magnitude as the gravitino mass, and the hierarchy problem is partially addressed by these soft terms stabilising the electroweak scale [58].

Gauge-Mediated Supersymmetry Breaking Models

Often referred to as GMSB models, these propose an alternative to Planck-scale-mediated scenarios, where the supersymmetry breaking is transmitted to the visible sector via gauge interactions instead of gravitational ones [59]. In these models, the communication occurs through messenger fields charged under the Standard Model gauge groups, which acquire supersymmetry-breaking masses. The resulting soft terms in the visible sector are generated at a lower scale, usually around 100 TeV, significantly lower than the Planck scale. This leads to a spectrum where the gravitino is the lightest supersymmetric particle (LSP), with typical masses ranging from a few eV to keV, making it a candidate for dark-matter [60].

Extra-Dimensional and Anomaly-Mediated Breaking

Extra-dimensional and anomaly-mediated models explore the idea that supersymmetry is broken in a higher-dimensional space, and its effects are felt in our four-dimensional world through mechanisms involving compactified dimensions [61]. These models often utilise the framework of string theory or brane-world scenarios, where the hidden and visible sectors are localised on different branes and communicate via bulk fields.

On the other hand, anomaly-mediated supersymmetry breaking (AMSB) is a distinct mechanism where the supersymmetry breaking is mediated by superconformal anomalies [62]. Superconformal anomalies are a specific type of anomaly that occurs in theories with superconformal symmetry, which is an extension of conformal symmetry incorporating supersymmetry. In field theory, an anomaly refers to the breaking of a symmetry that is preserved at the classical level but violated upon quantisation. In the context of AMSB, superconformal anomalies arise when the superconformal symmetry of the theory is broken by quantum effects, specifically by the renormalisation group flow of the theory [63]. In a superconformal theory, the anomaly is associated with breaking the dilatation (scaling) symmetry and the superconformal symmetry. The quantum corrections that

lead to these anomalies are proportional to the gauge couplings' beta functions and the matter fields' anomalous dimensions.

AMSB exploits these anomalies to generate soft supersymmetry-breaking terms in the low-energy effective theory [64]. The mechanism works as follows: even if the supersymmetry is unbroken in a higher energy regime, the quantum corrections associated with the superconformal anomalies induce soft-breaking terms when moving to a lower energy scale. These terms are proportional to the beta functions and the anomalous dimensions, making AMSB highly predictive because the pattern of supersymmetry breaking is tightly linked to the known RG flow of the theory. A good example of such predictions is the gravitino in AMSB models, which typically has a mass of tens to hundreds of TeV, making it less accessible for collider experiments but relevant for cosmological considerations [65].

2.2.4. R-parity

R-parity is a quantum number introduced in supersymmetric theories to prevent baryon and lepton number-violating interactions, which, if allowed, would predict rapid decay of the proton (a phenomenon that has not been observed [66, 67]). R-parity is multiplicatively conserved, which means that the corresponding product, rather than the sum, is preserved. The introduction of R-parity avoids the proton decay issues while ensuring the stability of the lightest supersymmetric particle (LSP). R-parity is defined as

$$R_p = (-1)^{3(B-L)+2S}, \tag{2.11}$$

where B is the baryon number, L is the lepton number, and S is the particle's spin [40]. Under this definition, all Standard Model (SM) particles have $R_p = +1$, and all supersymmetric partners (sparticles) have $R_p = -1$. R-parity conservation implies that interactions must involve an even number of sparticles. This conservation has significant consequences for the stability of the LSP and the decay channels of heavier sparticles. The lightest particle with $R_p = -1$ cannot decay into any SM particle because it would violate R-parity conservation. Consequently, the LSP is stable, making it a viable candidate for dark-matter, as it can only annihilate or decay into other supersymmetric particles. Heavier sparticles can only decay through cascades that produce the LSP and other SM particles. For instance, a neutralino ($\tilde{\chi}_2^0$)–a hypothetical particle that arises as a mass eigenstate from the mixing of the neutral supersymmetric partners (gauginos and

higgsinos) of the photon ($\tilde{\gamma}$), Z boson ($\tilde{Z}$), and neutral Higgs bosons ($\tilde{H}_u^0, \tilde{H}_d^0$)–might decay into a lighter neutralino ($\tilde{\chi}_1^0$), and a photon or a Z boson, provided that these processes are kinematically allowed [68].

R-parity provides some interesting phenomenological consequences; for example, the LSP is a stable dark-matter candidate in R-parity conservation. The nature of the LSP, whether it be the neutralino, gravitino, or another sparticle, significantly affects dark-matter detection strategies and relic-density calculations. Additionally, at colliders such as the LHC, R-parity conservation implies that supersymmetric particle production must result in events with missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$), as the LSP escapes detection. In contrast, R-parity violating (RPV) scenarios can lead to distinctive signatures without $E_{\mathrm{T}}^{\mathrm{miss}}$, often involving multiple leptons or jets due to the additional decay-channels available. Finally, in RPV models, proton-decay constraints are relaxed compared to scenarios with R-parity conservation. The specific RPV couplings must be small enough to comply with experimental bounds on proton lifetime, yet they allow for new decay channels that can be probed experimentally [40].

2.2.5. The minimal supersymmetric standard model (MSSM)

Supersymmetry as a framework can be implemented in various quantum field theories. In collider physics, phenomenologically viable alternative theories encompass the experimentally verified gauge structure and matter content of the SM. The Minimal Supersymmetric Standard Model (MSSM) is the most extensively examined supersymmetric extension of the SM. The MSSM incorporates the minimal set of new particle states necessary to organise the existing matter fields into supermultiplets while maintaining neutrality regarding the exact mechanism of symmetry breaking. The superpotential for the MSSM is given by

$$W_{\mathrm{MSSM}} = \bar{u}\boldsymbol{y}_u Q H_u - \bar{d}\boldsymbol{y}_d Q H_d - \bar{e}\boldsymbol{y}_e L H_d + \mu H_u H_d \,. \tag{2.12}$$

In this expression, the fields Q, L, $\bar{e}$, $\bar{u}$, and $\bar{d}$ represent the familiar quark and lepton doublets and singlets of the Standard Model. In their supersymmetric form, these fields also contain scalar partners, squarks and sleptons. The dimensionless Yukawa coupling parameters $\boldsymbol{y}_u$, $\boldsymbol{y}_d$ and $\boldsymbol{y}_e$ are 3×3 matrices in the vector like family space [69]. The Higgs sector is augmented to include two Higgs doublets, denoted H_u and H_d. This is necessitated by the requirement for the superpotential to be holomorphic, which prohibits

Gauge Supermultiplets		
Names	Spin 1	Spin 1/2
gluon, gluino	g	$\widetilde{g}$
W-bosons, wino	$W^{\pm}$, W^0	$\widetilde{W}^{\pm}$ $\widetilde{W}^0$
B-bosons, bino	B^0	$\widetilde{B}^0$

Table 2.1.: The gauge supermultiplets of the Minimal Supersymmetric Standard Model.

Matter Fields of the MSSM					
Classification			Representations		
Symbol	spin 0	spin 1/2	SU(3)	SU(2)	U(1)
L	$(\tilde{\nu}_L,\, \tilde{e}_L)$	$(\nu_L,\, e_L)$	**1**	**2**	-1
E	$\tilde{e}_R^*$	$e_R^\dagger$	**1**	**2**	-2
Q	$(\tilde{u}_L,\, \tilde{d}_L)$	$(u_L,\, d_L)$	**3**	**2**	$+\frac{1}{3}$
U	$\tilde{u}_R^*$ (up)	$u_R^\dagger$	**3**	**1**	$+\frac{4}{3}$
D	$\tilde{d}_R^*$ (down)	$d_R^\dagger$	**3**	**1**	$-\frac{2}{3}$
H_u	$(H_u^+,\, H_u^0)$	$(\widetilde{H}_u^+,\, \widetilde{H}_u^0)$	**1**	**2**	-1
H_d	$(H_d^0,\, H_d^-)$	$(\widetilde{H}_d^0,\, \widetilde{H}_d^-)$	**1**	**2**	$+1$

Table 2.2.: Matter Fields in the Minimal Supersymmetric Standard Model

using a single Higgs field and its complex conjugate to generate Yukawa interactions for both up-type and down-type quarks [43]. The fermionic superpartners of the Higgs fields are referred to as Higgsinos. The supermultiplets of the MSSM are shown in Tables 2.1 and 2.2 [40]. The gauge sector in the MSSM mirrors that of the Standard Model but includes additional superpartners. The gauge groups $SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ correspond to the gauge bosons G, W, and B, respectively. In the MSSM, the superpartners of these gauge bosons are termed gluinos, winos, and binos, respectively. The soft supersymmetry

breaking operators that correspond to the MSSM can be built by applying Tables 2.1 and 2.2 to Equation (2.10) and writing down the most general set of such terms:

$$\begin{aligned}
\mathcal{L}_{\text{soft}}^{\text{MSSM}} &\supset \left(-\frac{1}{2}\sum_{\in\{\tilde{B},\tilde{W},\tilde{g}\}} m_\alpha \alpha\alpha\right) - \tilde{\bar{u}}\boldsymbol{a}_u \tilde{Q} H_u + \tilde{\bar{d}}\boldsymbol{a}_d \tilde{Q} H_d + \tilde{\bar{e}}\boldsymbol{a}_e \tilde{L} H_d \\
\mathcal{L}_{\text{soft}}^{\text{MSSM}} &\supset \left(-\sum_{\in\{L,E,Q,U,D\}} \tilde{\phi}^\dagger m_\phi^2 \tilde{\phi}\right) - m_{H_u}^2 H_u^* H_u - m_{H_d}^2 H_d^* H_d - b H_u H_d \,,
\end{aligned} \tag{2.13}$$

Where m_α corresponds to the gluino, wino, and bino mass terms followed by the tri-linear couplings ($\boldsymbol{a}$), which are in one-to-one correspondence with the Yukawa couplings of the superpotential (Equation (2.4)). The second line contains squark and slepton mass terms where each m_ϕ is a 3×3 real hermitian matrix, followed by the Higgs squared-mass terms $m^2_{H_u}$ and $m^2_{H_d}$. The final term b is the non-holomorphic scalar mass term that can occur in the MSSM according to Equation (2.4).

2.3. Simplifying supersymmetry

Various simplifying assumptions and models are employed to manage the complexity of SUSY models, such as the constrained MSSM (cMSSM), phenomenological MSSM (pMSSM), and simplified models.

2.3.1. Constrained MSSM (cMSSM)

The Constrained Minimal Supersymmetric Standard Model (cMSSM), often called the minimal supergravity (mSUGRA) model, is a well-motivated extension of the MSSM. This model is characterised by its enforcement of universality conditions on the soft SUSY breaking parameters at the Grand Unified Theory (GUT) scale [55]. "Soft" in this context refers to a type of supersymmetry breaking that does not cause ultraviolet divergences to appear in scalar masses. The purpose of the softly broken supersymmetry is to protect the scalar mesons from quadratic mass renormalisations, preventing their appearance in the physics below the GUT mass scale [70]. These constraints significantly reduce the parameter space, enhancing the model's predictive power and facilitating analytical and numerical investigations.

In the cMSSM, the soft SUSY-breaking parameters are assumed to be unified at the GUT scale, typically around 10^{16}GeV. Specifically, the universality conditions require that the scalar masses m_0, gaugino masses $M_{1/2}$, and tri-linear couplings A_0 are each set to be shared values for all respective particles [71]. Consequently, the cMSSM is often characterised by the set of parameters shown in Table 2.3. These conditions significantly reduce the number of independent parameters compared to the general MSSM, which has over 100 free parameters. This parameter space reduction makes the model more manageable and allows for more straightforward phenomenological analysis and experimental testing.

Parameter	Description
m_0	Universal scalar mass.
$M_{1/2}$	Universal gaugino mass.
A_0	Universal tri-linear coupling.
$\tan\beta$	Ratio of the two Higgs doublets VEVs.
$\text{sgn}(\mu)$	Sign of the Higgsino mass.

Table 2.3.: Common representation of the five parameters that define the cMSSM [71].

The cMSSM provides a framework that naturally incorporates the mechanism of supersymmetry breaking through gravitational interactions, linking it to the more fundamental theory of supergravity [72]. The gravitational mediation of SUSY-breaking suggests that the soft terms are generated at the GUT scale and then evolve to the electroweak scale via the renormalisation-group equations (RGEs) [53]. The RGEs for the soft SUSY-breaking parameters ensure that the low-energy spectrum of superpartners can be computed once the initial conditions at the GUT scale are specified. The resultant spectrum and its properties are subject to experimental constraints from collider searches, dark-matter detection experiments, and precision measurements of SM observables.

2.3.2. Phenomenological MSSM (pMSSM)

The Phenomenological Minimal Supersymmetric Standard Model (pMSSM) represents an extension of the cMSSM by relaxing several assumptions and constraints that define the cMSSM. This relaxation allows for a larger parameter space, essential for reconciling theoretical models with the increasing volume of experimental data.

As covered in the previous section, the cMSSM makes certain assumptions to reduce the number of free parameters, which leads to a highly constrained model with only a few free parameters. The cMSSM's strict constraints can limit its flexibility and applicability in light of experimental results from collider experiments and dark-matter searches. In the pMSSM, the constraints of universality at the GUT scale are lifted, allowing the soft SUSY-breaking parameters to vary independently at the electroweak scale. This results in a model with 19 free parameters [73], provided in Table 2.4. Decoupling these parameters from each other means that the pMSSM can accommodate a greater range of phenomenological scenarios. For instance, the mass hierarchy between the sfermions and gauginos can be more finely tuned, and the model can be better adjusted to match the

Parameter	Description
M_A	Mass of the pseudoscalar Higgs boson
M_1, M_2, M_3	Bino, wino and gluino mass.
$m_{\tilde{q}}, m_{\tilde{u}_R}, m_{\tilde{b}_R}, m_{\tilde{l}}, m_{\tilde{e}_R}$	First/second generation sfermion masses
$m_{\tilde{Q}}, m_{\tilde{t}_R}, m_{\tilde{b}_R}, m_{\tilde{L}}, m_{\tilde{\tau}_R}$	Third generation sfermion masses
A_t, A_b, A_τ	Universal tri-linear coupling parameter
$\tan\beta$	Ratio of the two Higgs doublets VEVs.
μ	Higgs–Higgsino mass.

Table 2.4.: Common representation of the 19 parameters that define the pMSSM [73].

observed Higgs boson mass and other experimental constraints such as those from direct and indirect dark-matter detection experiments, electroweak precision tests, and flavour physics.

Additionally, the pMSSM allows for a more detailed exploration of the SUSY breaking mechanism. While the cMSSM's universality conditions imply a specific mechanism, such as gravity-mediated SUSY breaking [53, 72], the pMSSM can accommodate various mediation scenarios, including gauge-mediated and anomaly-mediated SUSY breaking. This flexibility is particularly valuable for phenomenological studies comparing theoretical predictions with experimental data.

The parameter space of the pMSSM is typically explored using a combination of analytical techniques and numerical simulations. Parameter scans are conducted to identify regions consistent with experimental data, often employing sophisticated statistical methods and global fits to account for uncertainties in the experimental measurements and theoretical predictions. Tools such as Markov Chain Monte Carlo (MCMC) methods [74] or nested-sampling [75] algorithms are frequently used to sample the high-dimensional parameter space of the pMSSM efficiently. A new search approach is presented in Chapter 7, where the limits on the pMSSM are improved by combining data from multiple experiments.

2.3.3. Simplified models

Simplified models in supersymmetry (SUSY) research are designed to focus on specific scenarios by considering only a limited subset of superpartners and their interactions.

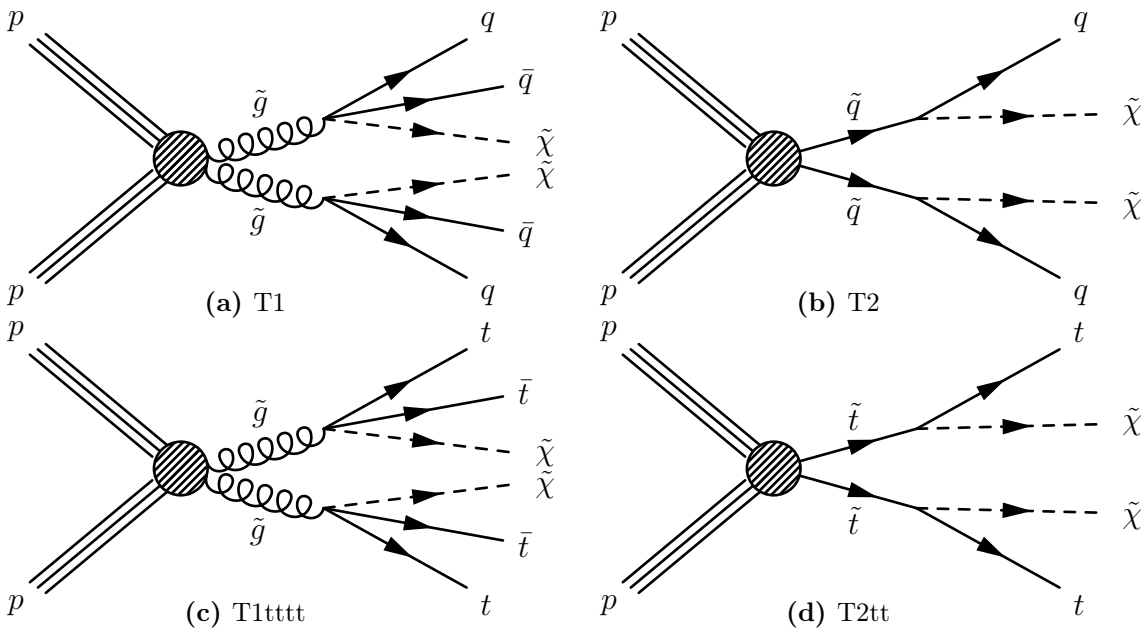


Figure 2.1.: Feynman diagrams of four simplified-model examples. The diagrams are split into T1 and T2 type topologies with (a) and (c) showing gluino pair production followed by the decay into a qq (tt) pair and a neutralino $\tilde{\chi}$. Diagrams (b) and (d) show squark (top-squark) pair production, where each squark decays into a quark and a neutralino.

By reducing the number of particles and interactions under consideration, these models provide a more tractable framework for both theoretical and experimental studies. Simplified models facilitate targeted searches for SUSY signatures at colliders, such as the Large Hadron Collider (LHC), by enabling a clear interpretation of experimental results in the context of specific SUSY predictions [76, 77]. In these models, assumptions are made to isolate particular SUSY processes, thereby avoiding the full complexity of more comprehensive models like the MSSM, cMSSM, or the pMSSM. For example, a simplified model might focus on the production and decay of a gluino and a squark, ignoring other possible SUSY particles. Figure 2.1 shows the Feynman diagrams for the T1 and T2 simplified models (sometimes called topologies), along with an alternative top-quark (t) decay mode.

Simplified models are related to full Simplified Model Spectra (SMS) through projection techniques, where the higher-dimensional parameter space of the SMS is projected onto a lower-dimensional subspace defined by the simplified model and, in turn, full SUSY models project into sets of orthogonal SMS [78]. The projection process is a method used to map high-dimensional parameter spaces onto lower-dimensional representations that capture the essential phenomenological features of the model. The projection is

typically performed by selecting a subset of parameters, such as the masses of the lightest supersymmetric particles and their dominant production and decay modes, while marginalising over less relevant details (further detail on likelihoods and marginalisation can be found in Section 3.3). Mathematically, this can be expressed as an integration over the marginalised parameters in a likelihood function,

$$\mathcal{L}_{\text{proj}}(m_{\tilde{\chi}}, m_{\tilde{g}}, \dots) = \int \mathcal{L}(m_{\tilde{\chi}}, m_{\tilde{g}}, \theta)\, d\theta, \tag{2.14}$$

where $\mathcal{L}(m_{\tilde{\chi}}, m_{\tilde{g}}, \theta)$ represents the full likelihood function dependent on both the primary mass parameters ($m_{\tilde{\chi}}$, $m_{\tilde{g}}$) and the marginalised variables θ, which may include couplings, branching fractions, or higher-order corrections. This projection preserves the leading kinematic and cross-section information while reducing the complexity of the model space, allowing for efficient reinterpretation of experimental data. The validity of the projection depends on the assumption that the marginalised parameters do not significantly alter the experimental observables and must be carefully evaluated in specific scenarios [78–80].

This projection process allows researchers to effectively constrain and interpret the broader parameter space by focusing on critical interactions and characteristics captured by the simplified models. For example, tools like SModelS [81] and Fastlim [82] can be used to project and constrain natural SUSY scenarios within the SMS framework, highlighting the practical utility of projections in making complex models more computationally and experimentally accessible. This approach aligns with general projection-based methods, such as those discussed in other fields like spectral projection models for scattering, which similarly reduce complexity by projecting onto subspaces that capture the most relevant features for specific analyses. This targeted approach aids in understanding how specific superpartners could manifest in collider data and allows for more precise measurements of their properties, such as masses and cross-sections.

Constraints derived from simplified models must be considered with caution as they depend on several factors, such as:

- Minimal Assumptions: Simplified models assume a limited set of production and decay modes, making them useful for setting conservative constraints that apply to broad classes of SUSY scenarios. However, real SUSY models often have additional decay channels or interference effects that can weaken or modify these bounds.
- Kinematic Representativity: if the dominant production mechanisms and kinematic features of the simplified model match those in a full SUSY theory, constraints

would be relatively robust. However, differences in mass spectra, branching ratios, or missing energy distributions may lead to overestimated or underestimated bounds.

- Reinterpretation in Complete Models: A full SUSY model may include additional particles and decay paths that evade simplified model constraints, making it crucial to reinterpret bounds in the context of the full parameter space.
- Experimental Coverage: Simplified models typically provide upper limits on cross-sections and mass exclusions, but they do not account for the full complexity of a SUSY parameter space.

Constraints are most reliable when derived from multiple complementary searches across different final states. Thus, while simplified models provide valuable guidance, they must be used carefully, keeping in mind the assumptions made and the need for reinterpretation within a full SUSY framework.

Simplified models serve as effective tools for comparing theoretical predictions with experimental data. By narrowing down the parameter space, these models help identify regions consistent with observed signals and rule out those that are not. This methodology significantly enhances the efficiency of SUSY searches, enabling a systematic exploration of potential SUSY phenomena without the computational and analytical burdens of more complex models [83].

2.4. Motivations for supersymmetry

2.4.1. Solution to the hierarchy problem

One of the primary motivations for supersymmetry is its potential to address the hierarchy problem, a fundamental question arising from the observed disparity between the gravitational scale, characterised by the Planck mass $M_{Pl} \approx 10^{19}$ GeV, and the electroweak scale, marked by the Higgs boson mass $m_H \approx 125$ GeV [36]. This problem highlights why the Higgs boson mass is so much lighter than the Planck mass despite quantum corrections that naively suggest it should be much heavier. This disparity is at odds with the concept of naturalness [37, 84], which indicates that any physics model should work without requiring ad-hoc coincidences or the fine-tuning of its parameters.

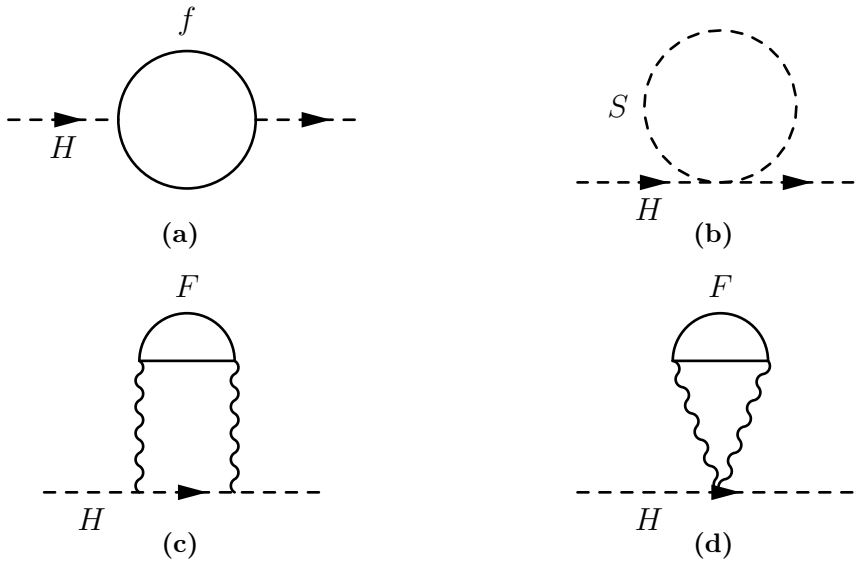


Figure 2.2.: Feynman diagrams showing one and two-loop quantum corrections to the Higgs squared mass parameter m_h^2. The upper plots show one-loop due to (a) a Dirac fermion f and (b) a scalar S. lower plots show two-loop corrections involving a heavy fermion F that couples only indirectly to the SM Higgs through gauge interactions [43].

Chapter 1 explored the process of electroweak symmetry breaking (Section 1.2) by introducing the potential $V(|\phi|)$ (Equation (1.18)) and evaluating the minima with $\mu^2 > 0$ and $\lambda > 0$, which resulted in a non-zero VEV. From this, it was shown that the quarks and leptons and the electroweak gauge bosons Z^0, $W^\pm$ all obtain masses through interactions with the Higgs field. Thus, the mass of the Higgs boson is subject to significant radiative corrections due to interactions with other particles in the SM.

The Higgs squared-mass, defined as $m_h^2 = 2\lambda v^2$, receives significant quantum corrections from each particle that couples, directly or indirectly, to the Higgs field. Figure 2.2 shows one and two-loop corrections to m_h^2, with the upper two plots showing one-loop cases with a Dirac fermion f (a) and a complex scalar S (b). The two lower plots show two-loop corrections involving a heavy fermion F that couples indirectly to the SM Higgs through gauge interactions. The quadratically divergent contributions to the Higgs mass from the loop diagrams necessitate fine-tuning the bare Higgs mass to cancel out these large corrections and yield the observed Higgs mass.

The corrected Higgs squared-mass can be expressed as $m_h^2 = m_{h0}^2 + \Delta m_h^2$, where m_{h0}^2 is the bare Higgs mass squared and Δm_h^2 represents the radiative corrections. The latter is dominated by contributions proportional to the cutoff scale Λ_{UV}, which is assumed to

be around the Planck scale. The correction terms corresponding to Figure 2.2 are [43]

$$\begin{aligned}
\Delta m_h^2 &= -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + ... && \text{(a)} \\
\Delta m_h^2 &= \frac{|\lambda_S|^2}{16\pi^2}\left[\Lambda_{UV}^2 - 2m_S^2\left(\frac{\Lambda_{UV}}{m_S}\right) + ...\right] && \text{(b)} \\
\Delta m_h^2 &= C_H T_F \frac{g_F^2}{16\pi^2}\left[a\Lambda_{UV}^2 - 24m_F^2\left(\frac{\Lambda_{UV}}{m_F}\right) + ...\right] && \text{(c, d)},
\end{aligned} \tag{2.15}$$

where λ_α, m_α are the coupling and mass terms for the fermion ($\alpha = f$), heavy scalar ($\alpha = S$) and heavy fermion ($\alpha = F$), and where C_H and T_F are group-theory factors of order one. Given $\Lambda_{\text{UV}} \sim 10^{19}$ GeV, the correction term Δm_h^2 is many orders of magnitude larger than the observed Higgs mass squared $m_h^2 \approx (125\text{GeV})^2$, suggesting an unnatural cancellation unless new physics is introduced at or near the electroweak scale. Supersymmetry introduces superpartners to the Standard Model particles, where the fermionic and bosonic loops cancel out the quadratic divergences, stabilising the Higgs mass naturally.

2.4.2. Dark-matter candidate

One of the most compelling motivations for SUSY is its capacity to provide a viable candidate for dark-matter, the unidentified matter component constituting approximately 27% of the universe's mass-energy content. In supersymmetric models, every SM particle has a corresponding superpartner with differing spin. A crucial feature of many SUSY models, particularly those with R-parity (Equation (2.11)) conservation, is the Lightest Supersymmetric Particle (LSP) stability. For all SM particles, R-parity is +1, while for their superpartners, it is -1. R-parity conservation implies that the LSP cannot decay into SM particles, ensuring stability. This stability is a crucial criterion for the LSP to be a dark-matter candidate since dark-matter must be long-lived on cosmological timescales [85].

Among the possible LSPs, the neutralino is the most extensively studied candidate. The neutralino is electrically neutral and weakly interacting, fitting the profile of a weakly interacting massive Particle (WIMP), a leading class of dark-matter candidates. The neutralino's mass and interaction strength can be fine-tuned within SUSY models to yield the correct relic-density, matching the observed dark-matter abundance [86].

The relic-density of neutralinos is determined through a process called 'thermal freeze-out' in the early universe. Initially, neutralinos were in thermal equilibrium with the SM particles. As the universe expanded and cooled, the interaction rate of neutralinos fell below the expansion rate, causing them to decouple or freeze out. The relic-density depends on the annihilation cross-section of neutralinos: a larger cross-section results in a lower relic-density and vice versa. SUSY models allow parameter adjustments that reproduce the observed dark-matter relic-density ($\Omega_{\mathrm{DM}}h^2 \approx 0.12$) [87, 88].

Another candidate is the gravitino, the superpartner of the graviton. In scenarios where the gravitino is the LSP, its interactions are suppressed by the Planck scale, making it a viable dark-matter candidate. Gravitino dark-matter models, however, can face challenges related to their production mechanisms and implications for Big Bang nucleosynthesis [89].

The superpartners of neutrinos, known as sneutrinos, are also potential dark-matter candidates. However, traditional sneutrino dark-matter is often disfavoured due to direct detection constraints, as their interaction cross-sections with nuclei are typically large. Nonetheless, variants such as right-handed sneutrinos can evade these constraints and remain viable candidates [90].

2.4.3. Gauge coupling unification

Another significant motivation for SUSY is its potential to facilitate gauge-coupling unification. In the context of the SM, the three fundamental gauge couplings do not converge to a single value at any energy scale. This lack of unification presents a challenge for constructing a cohesive theory that encompasses all fundamental interactions. However, within a supersymmetric framework, the presence of additional particles modifies the renormalisation-group equations governing the running of these couplings. Specifically, in the MSSM, these modifications converge the gauge couplings at a GUT scale, approximately $\mu = 10^{16}$ GeV. At the two-loop level, the renormalisation group equations are given by

$$\begin{aligned} \frac{\mathrm{d}\omega_i}{\mathrm{d}\log\mu} &= \frac{a_i}{2\pi} - \sum_j \frac{b_{ij}}{8\pi^2\omega_j} \\ \omega_i = \alpha_i^{-1} &= \frac{4\pi}{g_i^2}\,, \end{aligned} \tag{2.16}$$

where the i, j indices denote the various subgroups at energy scale μ, with the gauge couplings g_1, g_2, and g_3 representing the usual U(1), SU(2), and SU(3) gauge interactions. The coefficients a_i and b_{ij} contain the normalisation of the generators in the representation of the fermionic and scalar states [91].

The unification of gauge couplings at the GUT scale provides a more coherent and aesthetically appealing picture of the fundamental forces, suggesting that they may be manifestations of a single underlying force at high energies. This convergence supports the idea of a unified theory, such as those proposed in GUTs, which attempt to describe the three gauge interactions within a single framework.

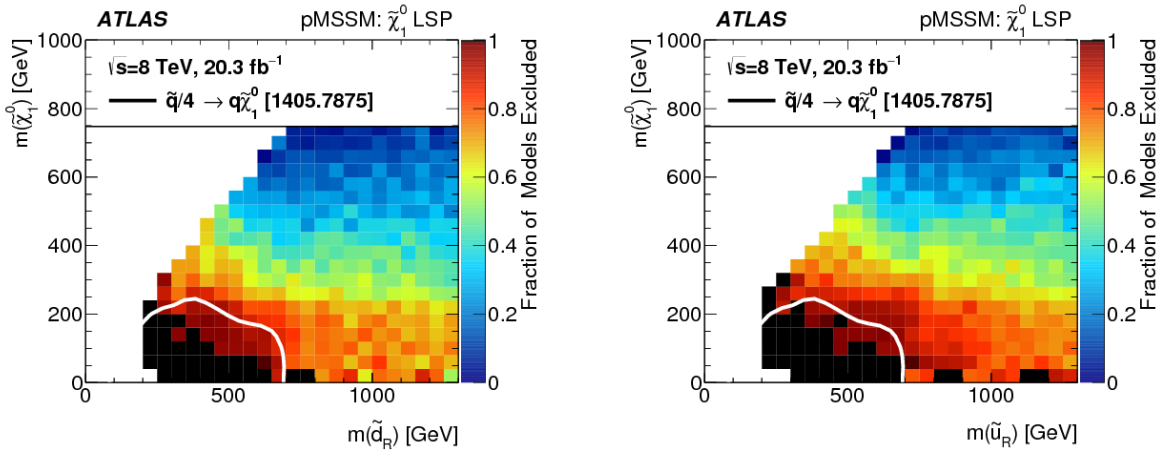
Furthermore, gauge coupling unification in SUSY models addresses several theoretical issues within the SM, such as the hierarchy problem and the stability of the Higgs boson mass. By incorporating SUSY, these models offer a more robust theoretical foundation for exploring BSM physics, reinforcing the appeal of supersymmetry as a candidate for new physics. Thus, the ability of supersymmetry to facilitate gauge coupling unification constitutes a compelling argument for its consideration in the search for an extension on the SM [53].

2.5. SUSY in summary

SUSY continues to be a compelling extension of the Standard Model (SM), addressing several outstanding issues, such as the hierarchy problem [37, 84], the nature of dark-matter[88], and the unification of gauge couplings [91].

The LHC has played an important role in testing SUSY models, particularly through its experiments like ATLAS and CMS, which have focused on searching for signatures predicted by these models. Among the various SUSY frameworks, the pMSSM has been extensively studied due to the reduced parameter space.

In 2015, the ATLAS collaboration performed a comprehensive reinterpretation campaign on the full pMSSM parameter space with 22 Run-1 analyses covering 200 signal regions [93–115]. The study explored millions of pMSSM model points and compared the predicted event yields against the observed data across various channels. The results from the ATLAS pMSSM study place significant constraints on the SUSY parameter space [92]. The absence of observed excesses corresponding to SUSY particles in the data has led to the exclusion of large portions of the parameter space, particularly in regions



(a) Right down squark **(b)** Right up squark

Figure 2.3.: Fraction of model points excluded. Upper plots a and b are in the planes of the masses of the right-handed squarks (up and down) versus the neutralino mass. (Figure taken from Ref. [92])

corresponding to lighter supersymmetric particles. For instance, scenarios where gluinos and squarks are below a few TeV are heavily constrained, especially in models where the lightest neutralino is assumed to be the lightest supersymmetric particle (LSP). The ATLAS study has shown that the exclusion limits are particularly strong in simplified models where certain sparticles are assumed to be decoupled, narrowing the viable SUSY parameter space considerably. Figure 2.3 shows the fraction of model points excluded in the pMSSM for the case of the right up and down squark. The plots were generated using simplified models and run 1 data from the ATLAS experiment.

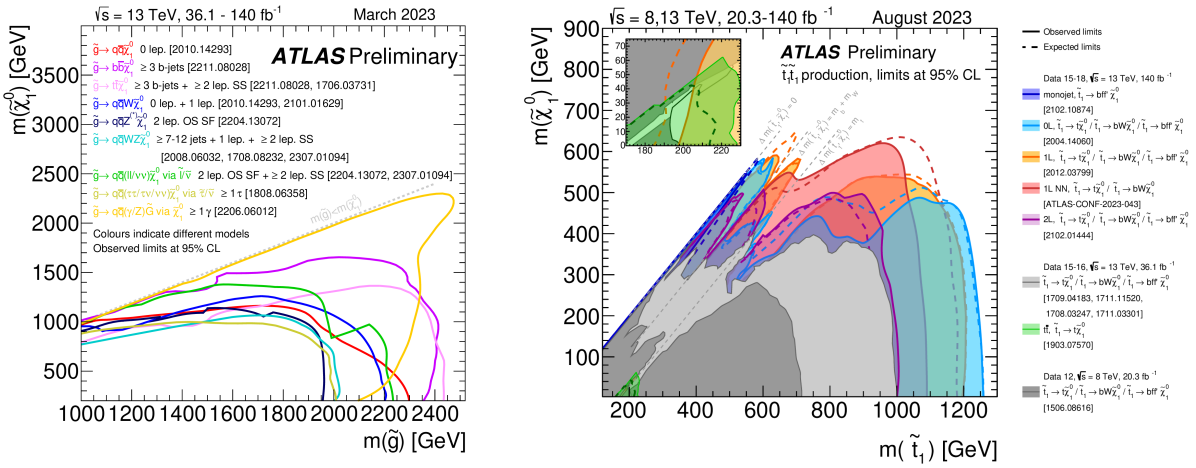
Despite these significant constraints, the ATLAS pMSSM study highlighted several limitations. One critical issue is the reliance on certain assumptions that may bias the results; for instance, the exclusion limits depend heavily on the assumed mass hierarchy and the specific decay chains considered. In scenarios where the mass difference between sparticles is small, the resulting soft leptons or jets that may evade detection, thus leaving certain regions of the parameter space unconstrained. The constraints derived from these results suffer from many of the factors discussed in the simplified model case (Section 2.3.3) such as minimal assumptions, kinematic representativity and reinterpretation in complete Models. Moreover, the pMSSM framework, while more manageable than the full MSSM, still involves a 19-dimensional parameter space. Such large dimensionality is too large for a Cartesian grid, so statistical sampling techniques were required to select model points. Such a complex study inevitably introduces potential issues, such as

the choices made in the parameter scans, such as flat priors that may overlook regions with non-trivial correlations between parameters. Then, there is data reduction and filtration; this is common practice when the processing chains contain bottlenecks like complicated calculations or simulation stages. In these situations, the model point being considered must be of interest and not be one that has already been excluded due to some other experimental result or theoretical condition. For this reason, conditionals are imposed to filter the selected model points; an example would be upper limits on cross-sections, as only points with low cross-sections are considered not excluded. These kinds of conditionals inevitably introduce some bias, as thresholds must be implemented by hand.

Another limitation is the experimental sensitivity, which is finite and dependent on the integrated luminosity and the specific analysis techniques employed. As the LHC continues to accumulate more data and as experimental techniques evolve, it is plausible that regions of the parameter space previously deemed excluded could reemerge or that new constraints could be imposed on currently allowed regions. Moreover, the ATLAS study primarily focuses on specific decay channels that are most accessible experimentally, but alternative decay modes or more exotic SUSY scenarios, such as those involving long-lived particles, are less stringently tested and could provide viable paths within the SUSY framework that evade current constraints.

The paper "Status of Searches for Electroweak-Scale Supersymmetry after LHC Run 2" [116] provides a comprehensive review of the results from the LHC following its second data-taking period. LHC Run 2 delivered proton-proton collisions at a $\sqrt{s} = 13$ TeV and an integrated luminosity of 140fb^{-1}. Despite the extensive data collected by the ATLAS and CMS experiments, no conclusive evidence was found for the production of supersymmetric particles. The search encompassed various experimental signatures and phenomenological scenarios, yet the null results have led to increasingly stringent constraints on SUSY models. For instance, Figure 2.4 the masses of the gluinos in gluino to LSP, plot (a), and top squarks, stop to LSP plot (b), are now constrained to be above 2.2 TeV and 1.1 TeV, respectively. The left-hand plot (a) shows the gluino-neutralino plane for different simplified models, while the right-hand plot (b) shows top squark (stop) pair production for dedicated ATLAS searches. These constraints challenge older notions of naturalness but align with more modern theoretical frameworks that consider the string landscape and refined SUSY breaking scenarios.

The paper also highlights the importance of the LHC's future phases, especially Run 3, which is expected to provide higher sensitivity due to detector and accelerator



(a) Gluino & lightest neutralino mass plane

(b) Squark & lightest neutralino mass plan

Figure 2.4.: Gluino to LSP (a) and stop to LSP (b) summary plot of mass exclusion. Left-hand plot (a) shows the gluino-neutralino plane for different simplified models. Right-hand plot (b) shows top squark (stop) pair production for dedicated ATLAS searches (b) (Figures 1 and 6 from ATLAS [117]). According to the legend, each curve refers to a different decay mode and is assumed to proceed with 100% branching ratio. All data is based on pp collision data taken at $\sqrt{s} = 13$ TeV

upgrades. These improvements are essential for probing deeper into unknown parameter spaces during Run 2. The focus is shifting towards more challenging signatures, such as compressed mass spectra, which are harder to detect but remain consistent with the absence of signals in previous searches. In summary, while SUSY has not yet been observed, its exclusion is far from complete, and the upcoming Run 3 offers new opportunities to potentially uncover supersymmetric particles.[116, 117]

The implications of LHC Run 3 for SUSY are significant, particularly in the quest to probe more challenging regions of parameter space that were less accessible in earlier runs. The primary objective of Run 3 is to continue the search for SUSY particles with improved sensitivity. With a higher centre-of-mass energy of $\sqrt{s} = 14$ TeV and an anticipated increase in integrated luminosity beyond 300fb^{-1}, Run 3 aims to explore scenarios that involve compressed mass spectra and more elusive particles, such as electroweakinos, which were not excluded in previous searches.

One key area of focus is the search for SUSY models that involve near-degenerate mass states, where the mass difference between supersymmetric particles and their Standard Model counterparts is small, making detection more difficult. These models have become more relevant due to the stringent constraints placed on more traditional SUSY searches

that assumed larger mass differences. Detecting new particles with near-degenerate mass states presents significant experimental challenges due to the difficulty in distinguishing their signals from background noise and each other. In the context of SUSY, near-degenerate mass spectra often arise in scenarios such as compressed SUSY models, where the mass difference between the lightest supersymmetric particles is very small. This small mass difference leads to low-energy decay products, such as soft leptons or jets, which are challenging to detect [118].

Another implication of Run 3 is the potential to explore models of SUSY that are embedded within broader theoretical frameworks, such as the string theory landscape or models incorporating extra dimensions, which predict higher sparticle masses or non-standard signatures.

Additionally, Run 3 will provide an opportunity to refine exclusion limits on the masses of key particles like gluinos and squarks, which are expected to push current bounds even higher, further challenging theories based on older naturalness arguments. If no evidence of SUSY is found, this would deepen the exclusion of certain SUSY models but would not necessarily rule out supersymmetry as a whole, especially in light of theoretical developments that suggest SUSY may manifest in more subtle or indirect ways.

Despite the absence of experimental evidence for superpartners, the framework remains fertile for phenomenological investigation. Ongoing and future experimental efforts persist in probing the extensive parameter space inherent to SUSY models, seeking potential signals that may corroborate its predictions [81]. As detailed in this chapter, theoretical constructs such as the superpotential, supersymmetry breaking mechanisms, and the MSSM's introduction refine our theoretical comprehension and guide experimental searches by delineating the parameter space experiments should target. The models discussed, particularly the simplified ones, serve as essential tools in testing the principles of SUSY. They provide a systematic approach to studying potential signatures and their implications in particle physics[76, 77].

The ongoing absence of empirical verification of SUSY does not undermine its theoretical elegance or utility as a research paradigm. On the contrary, it underscores the necessity for comprehensive and nuanced approaches in theoretical modelling and experimental design. There are currently hundreds of search analyses available for reinterpretation, providing a wealth of data to test new SUSY models and develop the statistical methods and tools that can be repurposed for future experimental data. As new data become available from current and future high-energy physics experiments, the insights gained

from the study of SUSY will continue to inform and shape the trajectory of particle physics research. Thus, while supersymmetry has not yet been observed, its theoretical frameworks and the rich phenomenology they encompass ensure that it remains a critical area of study. The exploration of supersymmetric theories continues to be instrumental in pushing the boundaries of our understanding of fundamental physics.

Chapter 3.

Statistics for Collider Physics

The interpretation of data from high-energy particle collisions is inherently statistical, necessitating the development and application of statistical techniques to discern signals from background noise and to make precise predictions. The LHC at CERN represents the forefront of this field, providing a wealth of data from particle collisions at unprecedented energy scales. Analysing such data involves overcoming several challenges, such as the sheer volume of data, the complexity of the physical processes involved, and the need for rigorous statistical methods to identify known processes and validate discoveries.

This chapter will explore the statistical methodologies employed to understand collider physics data. Starting from the fundamentals of probability, it will introduce concepts and notation that will be used extensively throughout this thesis. We will move through the axioms of probability, Bayesian and frequentist interpretations, into the derivations of distributions that describe experimental observations. This chapter intends to cement the theoretical foundations and the practical implementations of these methods, providing a comprehensive overview of their application in the analysis of collider data. This exploration will be framed within the context of recent experimental results from the LHC, highlighting the ongoing developments and challenges in the field.

3.1. Fundamentals of probability

Defining the term "probability" is essential as it forms the foundation of statistical analysis. While probability has been intuitively understood and applied in various forms for centuries, its formal mathematical definition was established in the early 20th century by Andrey Kolmogorov [119]. Let's consider a set of random events, denoted as E_i, where

each event E_i is mutually exclusive, i.e. if one event occurs, none of the other events can co-occur; they are disjoint or mutually exclusive. The probability P associated with a particular event E_i is defined according to the following axioms, which are known as the Kolmogorov axioms [7, 120]:

1. **Non-negativity** For any event E_i, the probability $P(E_i)$ is a non-negative number:

$$P(E_i) \geq 0\,. \tag{3.1}$$

2. **Normalisation** The probability of the entire sample space S, which represents the set of all possible outcomes, is equal to one

$$\sum_i P(E_i) = P(S) = 1\,. \tag{3.2}$$

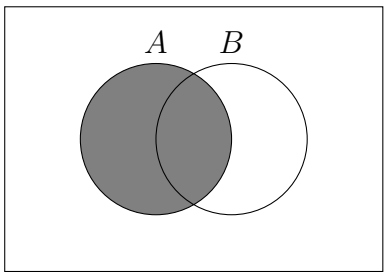
3. **Additivity** For any countable sequence of mutually exclusive events $E_1, E_2, \ldots$, the probability of the union of these events is equal to the sum of their probabilities:

$$P\left(\bigcup_{i=1} E_i\right) = \sum_{i=1} P(E_i)\,. \tag{3.3}$$

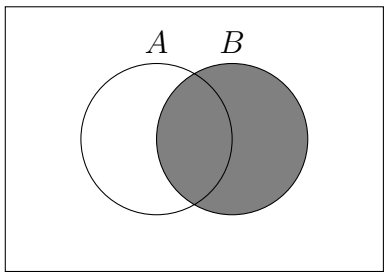
These axioms collectively define the mathematical properties of a probability measure. They provide the basis for calculating the likelihood of events within a well-defined probability space. The axioms are fundamentally grounded in set theory and provide the foundational basis from which various properties of probability can be systematically derived. To illustrate, consider two subsets of events, A and B, within the universal event space S. These subsets are non-exclusive, meaning that at least one event element, E_i, is common to A and B. In other words, A and B are not disjoint, allowing for overlap between the two subsets. The inclusion-exclusion principle describes the probability of an event occurring that belongs to either A alone, B alone, or both A and B. According to this principle, the probability of the union of A and B is given by:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,, \tag{3.4}$$

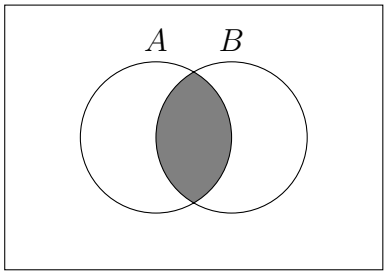
In this expression, $P(A)$ denotes the probability of an event in subset A, $P(B)$ denotes the probability of an event in subset B, and $P(A \cap B)$ represents the probability of an event that is common to both A and B. The subtraction of $P(A \cap B)$ ensures that the



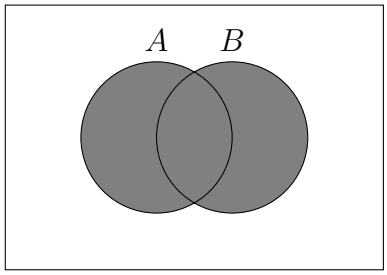
(a) $P(A)$



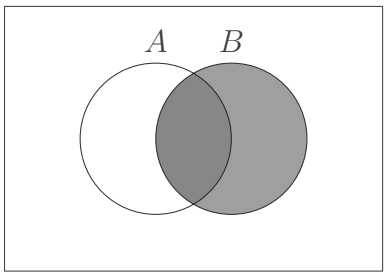
(b) $P(B)$



(c) Intersection $P(A \cap B)$



(d) Union $P(A \cup B)$



(e) Conditional $P(A \mid B)$

Figure 3.1.: Plots a through d show the probability relations that emerge from the Kolmogorov axioms [119], plot e introduces the conditional relation that is the probability of A given B where the darker shaded area illustrates the conditional overlap.

probability of events occurring in both subsets is not double-counted. This relationship is shown as a series of Venn diagrams in Figure 3.1 where $P(A)$, $P(B)$, and $P(A \cup B)$ are subplots a, b and d, respectively.

3.1.1. Bayes' theorem

Expanding on these relations, consider an event E_i that is known to belong to the set B; what would be the probability that the event E_i also belongs to the set A? This conditional probability is denoted by $P(A \mid B)$ and is interpreted as "the probability of

A given B." Conditional probability provides a mechanism to update the probability estimates based on additional information, in this case, the occurrence of an event in B. This relation is depicted in the lower plot (e) of Figure 3.1; the formal definition of conditional probability is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \,. \tag{3.5}$$

This equation states that the conditional probability $P(A \mid B)$ is equal to the probability of the intersection of events A and B (i.e., the event that occurs in both A and B) divided by the probability of B. It is important to note that this definition assumes $P(B) > 0$ since the conditional probability is only meaningful when the conditioning event B has a non-zero probability. Looking at Figure 3.1, it is easy to see that the intersection $P(A \cap B)$ is equivalent to $P(B \cap A)$, putting this into Equation (3.5) provides the the following relation

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A) \,. \tag{3.6}$$

This is a significant result as it provides a way to invert conditional probabilities, dividing through by $P(A)$ or $P(B)$, which results in the central equation of Bayes' theorem:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} \Rightarrow P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \,. \tag{3.7}$$

This equation can be generalised by considering that the event E_i belongs to a single partition Λ_i of the sample space. If B is any event, the probability of B can now be written as

$$P(B) = \sum_i P(B \cap \Lambda_i) = \sum_i P(B \mid \Lambda_i) \cdot P(\Lambda_i) \,. \tag{3.8}$$

Putting this relation back into Equation (3.7) gives an extended, generalised form of Bayes' theorem

$$P(\Lambda_i \mid B) = \frac{P(B \mid \Lambda_i) \cdot P(\Lambda_i)}{\sum_i P(B \mid \Lambda_i) P(\Lambda_i)} \,. \tag{3.9}$$

A common representation of this equation is often used where the individual components have symbolic representations

$$\pi(\Lambda_i \mid B) = \frac{\mathcal{L}(B \mid \Lambda_i) \cdot \pi(\Lambda_i)}{\mathcal{Z}(B)}\,, \tag{3.10}$$

where $\pi(\Lambda_i \mid B)$ is referred to as the posterior, $\mathcal{L}(B \mid \Lambda_i)$ the likelihood, $\pi(\Lambda_i)$ the prior and $\mathcal{Z}(B)$ as the evidence. Up to this point, the language has been restricted to the language of events and sets, which has provided a solid framework for statistical derivations. However, to bring the statistics into the context of experimental physics, it is practical to consider Equation (3.10) in terms of hypotheses H and Data D

$$\pi(\text{Hypothesis} \mid \text{Data}) = \pi(H \mid D) = \frac{\mathcal{L}(D \mid H) \cdot \pi(H)}{\mathcal{Z}(D)}\,. \tag{3.11}$$

In this context, it is helpful to break down Equation (3.11) and describe what each component represents in terms of H and D.

Posterior, $\boldsymbol{\pi(H \mid D)}$: The posterior distribution is the probability of the hypothesis H given the observed data B. The posterior combines prior knowledge about H with the information provided by the data. In Bayesian inference, the posterior is of primary interest, as it reflects the revised beliefs about the parameter H given the observed data D. Using π for the symbolic representation identifies the posterior as an updated probability distribution of the parameter H that can be used as the prior for in an iterative procedure.

Likelihood $\boldsymbol{\mathcal{L}(D \mid H)}$: The likelihood quantifies how well the hypothesis H explains the observed data D. It is the probability of the data given the hypothesis. In practice, the likelihood function plays a crucial role in the inference process, as it indicates the relative plausibility of different hypotheses based on the observed evidence. The likelihood is not a probability distribution over H but a function of H for fixed D. The likelihood is covered in more detail in Section 3.3, as it plays a large role in frequentist analysis and particle physics, especially in hypothesis testing.

Prior $\boldsymbol{\pi(H)}$: The prior probability represents the initial degree of belief in the hypothesis before any data is observed. It reflects prior knowledge or assumptions about the parameter or hypothesis. The choice of prior can be described as informative, incorporating specific knowledge, or non-informative, representing a state of relative ignorance about H.

An informative prior incorporates specific knowledge or assumptions about the hypothesis. This type of prior is typically derived from previous research and is often a previously calculated posterior result. An informative prior will influence, and possibly bias, the posterior distribution, especially when the amount of new data is limited. For instance, if previous experiments have suggested that a particular parameter is likely within a specific range, the prior distribution can be given a higher weight to that range, reflecting a higher degree of confidence in that parameter value.

Conversely, a non-informative prior, also known as an objective or reference prior, is designed to have minimal influence on the posterior distribution. Non-informative priors are employed when there is little to no previous knowledge about the hypothesis, representing a state of relative ignorance. Such priors are often chosen to be uniform over the parameter space or to have large variances, allowing the data to play a more dominant role in shaping the posterior distribution. A non-informative prior tries to avoid introducing subjective biases, aiming instead for an analysis driven primarily by the observed data. A classic choice of such a prior is the Jeffreys prior, which is defined by the square root of the determinant of the Fisher information matrix, $I(\theta)$ (see Section 3.3.2), for a parameter θ. Formally, the Jeffreys prior $\pi(\theta)$ is expressed as

$$\pi(\theta) \propto \sqrt{\det(I(\theta))}\,, \tag{3.12}$$

where I represents the Fisher information with respect to the likelihood $p(x|\theta)$. This formulation is invariant under reparameterisation, making it a popular choice in objective Bayesian analysis [121, 122].

There is a continuum of choices between maximally and minimally informative priors. The degree of informativeness of the prior should be carefully considered in light of the specific context and objectives of the analysis, as it can significantly impact the resulting inferences [120]

Evidence $\mathcal{Z}(D)$: The evidence, or marginal likelihood $\mathcal{Z}(D)$, is the normalising constant in Bayes' theorem. It is the total probability of the observed data D under all possible hypotheses. It is obtained by summing (or integrating, in the case of continuous hypotheses) the likelihood over the entire space of hypotheses, weighted by their prior probabilities:

$$\mathcal{Z}(D) = \mathcal{L}(D) = \sum_i \mathcal{L}(D \mid H) \cdot \pi(H) \quad \text{(discrete case)}\,, \tag{3.13}$$

$$\mathcal{Z}(D) = \int \mathcal{L}(D \mid H) \cdot \pi(H)\, dH \quad \text{(continuous case)}\,. \tag{3.14}$$

The evidence is crucial for model comparison as it allows for the assessment of how well different hypotheses explain the observed data, taking into account both the fit of the model (through the likelihood) and the complexity or plausibility of the model (through the prior).

A fundamental concept in Bayesian inference is the posterior odds ratio, which quantifies how the odds of one hypothesis relative to another update after observing data. Given two competing hypotheses, H_1 and H_2, the posterior odds ratio is defined as

$$\text{Posterior Odds} = \frac{P(H_1|D)}{P(H_2|D)} = \left(\frac{P(D|H_1)}{P(D|H_2)}\right) \times \left(\frac{P(H_1)}{P(H_2)}\right), \tag{3.15}$$

where $P(H_1|D)$ and $P(H_2|D)$ are the posterior probabilities given data D, while $P(H_1)/P(H_2)$ is the prior odds ratio. The term $P(D|H_1)/P(D|H_2)$ represents how well the data supports one hypothesis over the other and is known as the *Bayes factor* [122]. The posterior odds ratio incorporates both prior beliefs and the evidence from the data, allowing for a comprehensive update of beliefs in light of new observations. The Bayes factor is a likelihood ratio that serves as a measure of the strength of evidence provided by the data in favour of one hypothesis relative to another. It is formally defined as

$$BF_{12} = \frac{P(D|H_1)}{P(D|H_2)}\,, \tag{3.16}$$

where $BF_{12} > 1$ indicates that the data favours H_1 over H_2, while $BF_{12} < 1$ suggests the opposite [123]. The Bayes factor plays a crucial role in Bayesian model comparison as it quantifies the relative plausibility of competing hypotheses independent of prior beliefs. However, its interpretation depends on the prior odds: a high Bayes factor can significantly shift posterior odds if the prior odds are not strongly against the favoured hypothesis. Since the posterior odds ratio is the product of the prior odds and the Bayes factor, the latter functions as a bridge between prior beliefs and posterior inference, directly influencing the extent to which the observed data modifies the initial hypothesis preference.

In Bayesian inference, Bayes' theorem provides a systematic way to update the probability of a hypothesis as new data becomes available. The posterior probability $P(H \mid B)$ combines the prior information $\pi(H)$ with the new evidence provided by the

likelihood $\mathcal{L}(B \mid H)$, with the evidence $\mathcal{Z}(B)$ ensuring that the posterior distribution sums (or integrates) to one, maintaining a valid probability distribution.

3.1.2. Frequentist interpretation

So far, probability has been presented as a stochastic interpretation of abstract events within a region or domain. Extending the logic into conditional probabilities allowed us to derive the Bayesian interpretation. However, there is a definition of probability that requires much less abstraction and is arguably more intuitive

$$P(X) = \lim_{N\to\infty} \frac{1}{N} \sum_i X_i, \quad X_i \in \{0, 1\}\,. \tag{3.17}$$

This is Frequentist probability where $P(X)$ is the asymptotic proportion of trials for which an X is true, or in the boolean sense, $X = 1$. The classic example of this is the simple coin toss, which, under ideal conditions, provides a 50 % chance of either heads or tails. In the frequentist interpretation of probability, the probability of an event is defined as the long-run relative frequency with which the event occurs when an experiment is repeated under identical conditions. Consequently, this conceptualisation of probability is inherently tied to repeatable experiments. For instance, when considering the probability of obtaining a specific number of heads in one hundred tosses of a fair coin, the frequentist framework is applicable because the experiment of tossing the coin can be repeated an arbitrary number of times, allowing the probability to be estimated through the observed frequency of outcomes.

Frequentist statistics is dominant in experimental collider physics due to its well-established framework for hypothesis testing (Chapter 4) and its historical alignment with the field's experimental practices. Particle physics experiments are, by design, repeatable, produce large amounts of data and involve testing specific hypotheses about the existence or properties of particles, which aligns naturally with the frequentist approach. But still, there is only one universe and only one true set of priors.

3.2. Probability mass and density functions

In statistical theory, probability mass functions (PMFs) and probability density functions (PDFs) are fundamental tools for characterising random variables. These functions provide

a mathematical framework for describing the distribution of discrete and continuous variables. For a discrete random variable X, the probability mass function $p_X(x)$ assigns a probability to each possible value of X, where x represents a possible outcome of X, meaning that $P(X = x)$ denotes the probability that X takes the value x. The PMF satisfies the Kolmogorov axioms, ensuring that the total probability across all possible outcomes is unity

$$\sum_x p_X(x) = \sum_x P(X = x) = 1\,. \tag{3.18}$$

In contrast, for a continuous random variable Y, the probability density function $f(y)$ describes the relative likelihood of Y taking a particular value. Unlike the PMF, the PDF does not assign a direct probability to any specific value of Y. Instead, the probability that Y lies within an interval $[a, b]$ is the integral of the PDF over that interval. Consider a discrete random variable with values y_i and a small interval δy around each y_i. Defining a density function $f(y)$ such that

$$P(a \leq Y \leq b) \approx \sum_{y_i \in [a,b]} p_i \approx \sum_{y_i \in [a,b]} f(y_i)\delta y\,, \tag{3.19}$$

we take the limit as $\Delta y \to 0$, replacing the summation with an integral:

$$P(a \leq Y \leq b) = \int_a^b f(y)\, dy, \tag{3.20}$$

where $f(y)$ is the probability density function (PDF) and must satisfy the same properties as the discrete case in the limit of $\delta y_i \to 0$, such that

$$\begin{aligned} f(y) &\geq 0 \quad \text{for all } y, \\ \int_{-\infty}^{\infty} f(y)\,\mathrm{d}y &= 1\,, \end{aligned} \tag{3.21}$$

ensuring that the area under the $f(y)$ curve equals one, reflecting the total probability.

The distinction between PMFs and PDFs is critical in statistical analysis as it guides the appropriate methods for computing probabilities and deriving inferential statistics. This section will cover some fundamental concepts related to the PMF and PDF functions and define a selection of distributions encountered in collider statistics relevant to later chapters.

3.2.1. Expectation values and moments

Given a probability density function (PDF) defined for a random variable x, the expectation value of some function $g(x)$ is given by

$$\langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) f(x) \, \mathrm{d}x \,, \tag{3.22}$$

where $f(x)$ is the PDF of the random variable x and $\langle g(x) \rangle$ represents the expected value or mean of the function $g(x)$ under the distribution defined by $f(x)$. Statistical moments, specifically the central moments about the mean, are essential tools for characterising distributions. The central moments are a common applications of Equation (3.22) where the n-th central moment of a random variable x is defined as [124]

$$\mu_n = \langle x^n \rangle = \int_{-\infty}^{\infty} x^n f(x) \, \mathrm{d}x \,. \tag{3.23}$$

The first moment, μ_1, represents the mean of the distribution and provides a measure of the central location of the data. The mean is fundamental in descriptive and inferential statistics, as it is a point of reference for other moments [125]. The second moment μ_2, commonly expressed as σ^2, is the variance of the distribution, defined as

$$\sigma^2 = \langle (x - \mu_1)^2 \rangle = \int_{-\infty}^{\infty} (x - \mu_1)^2 f(x) \, \mathrm{d}x \,. \tag{3.24}$$

The variance quantifies the dispersion of the data around the mean, offering insight into the variability within the dataset. The square root of the variance, σ, is the standard deviation often referred to as the scale or width of the distribution.

Higher-order moments provide additional information about the distribution's shape. The third moment, normalised by the cube of the standard deviation, is known as the skewness:

$$\gamma_1 = \frac{\langle (x - \mu_1)^3 \rangle}{\sigma^3} \,. \tag{3.25}$$

Skewness measures the asymmetry of the distribution around its mean. A positive skewness indicates a distribution with a longer right tail, while a negative skewness indicates a longer left tail. A skewness of zero corresponds to a symmetric distribution. A non-zero skewness (and higher moments) will also result in divergence between the mean, mode and median values. The fourth moment, normalised by the fourth power of

the standard deviation, the kurtosis

$$\gamma_2 = \frac{\langle (x - \mu_1)^4 \rangle}{\sigma^4}\,, \tag{3.26}$$

Kurtosis measures the "tailedness" of the distribution, indicating the presence of outliers. A high kurtosis implies heavy tails, suggesting that extreme values are more likely than in a normal distribution, whereas a low kurtosis indicates lighter tails. In practice, excess kurtosis, defined as $\gamma_2 - 3$, is often used to compare a distribution's kurtosis to that of a normal distribution, which has a kurtosis of 3.

Together, the mean, variance, skewness, and kurtosis form the basis for understanding and describing the shape and characteristics of a probability distribution. The moments beyond the fourth are rarely used in practice but provide further detail if necessary. Statistical moments are powerful tools in both theoretical and applied statistics. They enable the characterisation of distributions beyond mere central tendency.

3.2.2. The binomial distributions

To understand the binomial distribution, let's return to the coin-toss example raised when discussing the frequentist interpretation. Figure 3.2 shows the frequency distribution of the sum k of n binary trials, specifically, taking the sum of the outcomes from a repeated coin-toss experiment. Each histogram illustrates the results of N repetitions, with the number of binary trials n increasing from left to right. The histogram in the upper left-hand plot of Figure 3.2, corresponding to $n = 1$, shows $N = 10,000$ repeats of a single binary event, resulting in an approximately equal probability of obtaining either $k = 0$ or $k = 1$. This reflects the inherent symmetry of a fair coin-toss, where the probabilities of heads (1) and tails (0) are both 0.5. Moving to the right, the value of n increases, expanding the range of possible outcomes. The mean and variance of the distribution are np and $np(1-p)$, where p represents the probability of success in each binary trial ($p = 0.5$). These are the expectation value and variance of the binomial distribution with parameters k, n and p,

$$P(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}\,. \tag{3.27}$$

The red points correspond to the expected frequency value of each bin according to the binomial PMF, while the solid red line demonstrates the convergence to the

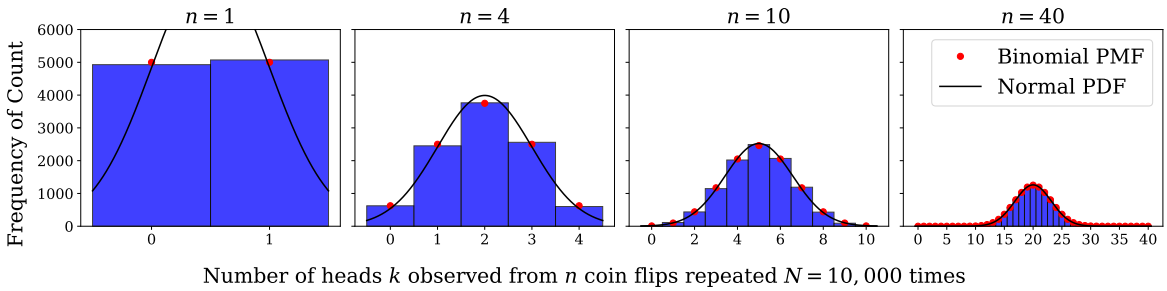


Figure 3.2.: Histograms showing the frequencies of values obtained by taking the sum of n binary trials (i.e. a coin-toss) repeated multiple times (N). The left-hand plots show 10,000 repeats of a single binary event ($n = 1$) with a near-equal split between 0 and 1. Moving to the right, the number of events increases, and thus, the range of possible values increases, with the mean value being $n/2 = np$ where $p = 0.5$. The red points indicate the expected results using the Binomial distribution and the black solid line illustrates how the distribution approaches the normal distribution at the limit of large n.

normal distribution at large n (this is covered in greater detail later in the chapter, see Section 3.2.4)

3.2.3. The Poisson distribution

The Poisson distribution can be understood as a limiting case of the binomial distribution (Equation (3.27)). This relationship emerges when the number of trials n becomes infinitely large, while the product $\lambda = np$, representing the expected number of successes, remains finite. To elaborate, the binomial distribution describes the probability of obtaining a fixed number of successes in a series of n independent trials, each with a success probability p. Redefining p as λ/n Equation (3.27) can be rewritten as

$$P(X = k, n, \lambda) = \frac{\lambda^k}{k!} \frac{n!}{(n-k)!n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} . \tag{3.28}$$

In this limits of $n \to \infty$ the terms containing n become

$$\begin{aligned} \lim_{n\to\infty} \frac{n!}{(n-k)!n^k} &= \frac{n(n-1)...(n-k+1)}{n^k} = 1 \\ \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1\,, \end{aligned} \tag{3.29}$$

giving the Poisson distribution

$$P(X = k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{Z}^+ . \tag{3.30}$$

The binomial distribution converges to the Poisson distribution as n increases without bound and p correspondingly decreases such that the product $\lambda = np$ remains fixed. Mathematically, Equation (3.29) is a straightforward exercise in limits; however, the physical interpretation of this limit can be hard to visualise; One way to do this is to consider that the Poisson distribution expresses the probability of a given number of events occurring in a fixed interval. Looking back to the previous Binomial example, consider the coin-toss shown in Figure 3.2; one might ask the question, "Given N repeated trials of counting the number heads in n coin-tosses, where n is an even number, what is the distribution of k being equal to the expectation value?" Looking at Figure 3.2, it is clear that as n increases, the probability of k from a single trial being equal to any given value between 0 and n reduces. The probability that any single trial is equal to the expectation is given by:

$$P(K, N, P_n) = \binom{N}{K} P_n^K (1 - P_n)^{N-K} , \tag{3.31}$$

where P_n is the Binomial PMF function (Equation (3.27)) evaluated at the expectation value np and K is the the successful trail defined as $k = np$. The expectation of value of Equation (3.31) is simply NP_n with a variance of $NP_n(1 - P_n)$. Following the Poisson limit of P_n going to zero while N tends to infinity is equivalent to the limit of $N \to \infty$ and $n \to \infty$. For example, in a single trial, a coin is flipped twice ($n = 2$), and the number of heads is counted, giving three possible outcomes: 0, 1, and 2, with an expectation value of $np = n/2 = 1$. If this experiment is repeated one thousand times ($N = 1,000$), one would expect around 500 instances of a trial outcome equal to 1. Repeating this experiment many times produces the distribution shown in the upper left-hand plot of Figure 3.3, and as expected, the central mean value sits at 500. Moving to the right, the number of flips per trial n increases, increasing the possible outcomes. With more possible outcomes, the probability that a single trial will equal the expectation value (P_n) decreases. Each plot shows the binned results (with integer spacing) of the trails, along with the expected frequency according to both the binomial and Poisson distributions (Equations 3.31 and 3.30). As P_n decreases, so does the difference between the binomial and Poisson distributions shown in the residuals below each plot. The convergence of the two distributions is quantified by the Kullback–Leibler divergence (D_{KL}) using the

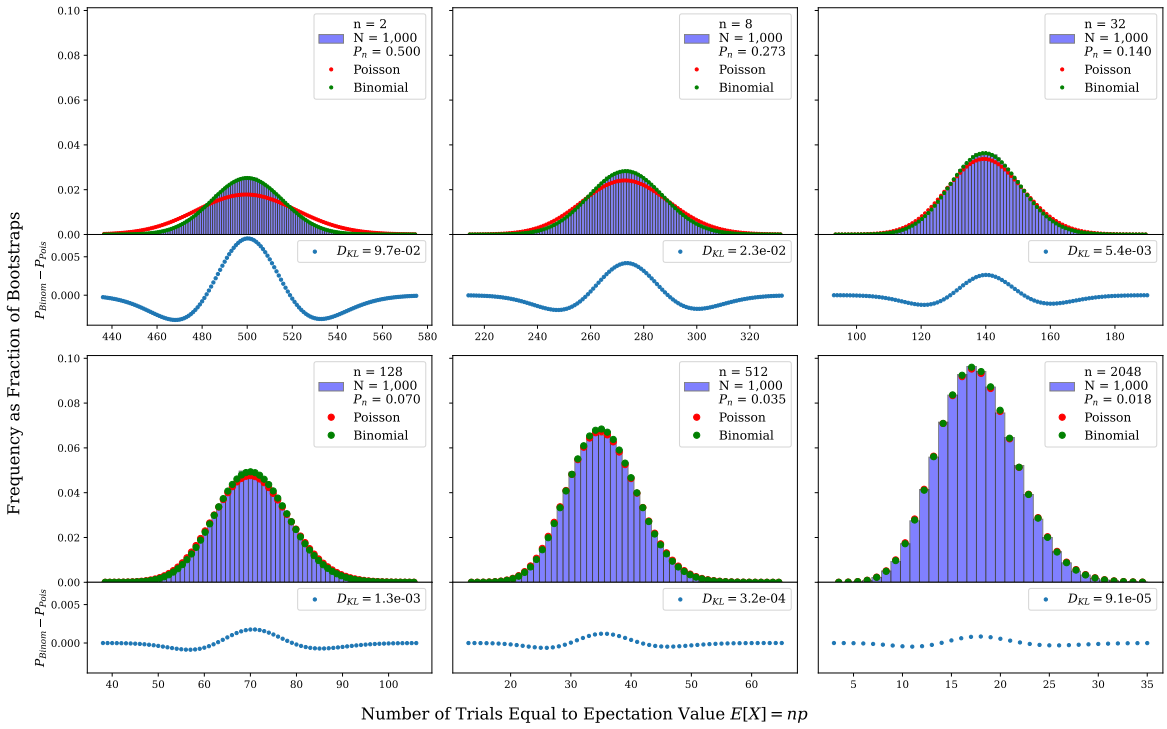


Figure 3.3.: Histograms displaying the results of 100,000 bootstrap, each consisting of $N = 1,000$ trials. In each trial, a coin is flipped n times, and a trial is considered successful if the number of heads, k, matches the expected value np. For example, if a coin is tossed twice ($n = 2$) with a fair probability of heads ($p = 0.5$), a successful trial occurs when exactly one head is observed, as the expectation is $np = 2 \times 0.5 = 1$. Thus for 1,000 trials with $n = 2$, one would expect to see 500 ± 16 heads. As n increases, the probability P_n of observing k trials equal to the expectation value np approaches zero, this pushes the distribution closer to the Poisson approximation (3.29), which is shown in the residuals along with the decreasing Kullback-Leibler divergence.

following equation [126]

$$D_{KL}(P_A \parallel P_B) = \sum_x P_A(x) \log\left(\frac{P_A(x)}{P_B(x)}\right) , \tag{3.32}$$

where P_A and P_B are the two probability distributions of interest. The Kullback-Leibler divergence is a measure of relative entropy or a measure of information variation [127]. Looking at the residuals in Figure 3.3, it is clear that as $P_n \to 0$, the two distributions converge and the D_{KL} "distance" reduces; this demonstrates how in the limit of large N and small P_n the binomial distribution can be approximated by the Poisson.

The Poisson distribution is widely used in collider physics due to the discrete and stochastic nature of particle-detection processes. This application is grounded in several

fundamental properties of the Poisson distribution that align well with the characteristics of particle events.

Firstly, the Poisson distribution describes events occurring in a fixed interval; this is an appropriate choice for modelling the number of events happening at a given rate. In collider physics, the rate is defined by the relationship between instantaneous luminosity $L(t)$, integrated luminosity L_{int}, cross-section σ and count N:

$$\begin{aligned} N = \sigma L_{\text{int}} &= \sigma \int_0^T L(t)\,\mathrm{d}t \\ \frac{\mathrm{d}N}{\mathrm{d}t} &= \sigma L(t)\,. \end{aligned} \tag{3.33}$$

Here, we can see that the integrated luminosity measures how much potential data has been taken in time T [7]. In practice, expressing the rate in terms of unit time fails to capture the relationship between particle process and beam energy. Thus, the rate is expressed as per unit integrated luminosity, a continuous measurement of the total number of collisions that occur over time and is usually expressed in inverse femtobarns (fb^{-1}). Therefore, the rate of such detections can be modelled as a Poisson process, where the mean number of events, denoted as λ where $\lambda = \sigma L_{\text{int}}$, represents the expected number of particles detected in a given interval.

The suitability of Poisson statistics in collider physics also stems from its applicability in scenarios where the mean event rate λ is low, but the number of trials is large. Given that some particle interactions are often rare events, the assumption of a small λ is realistic. This aligns with the conditions frequently encountered in experiments, where the total number of particles or events under observation is very large, while only a small fraction is detected or measured.

3.2.4. The normal distribution & the central limit theorem

The normal distribution, or Gaussian distribution, is one of statistics' most fundamental and widely used probability distributions. It is defined for a continuous random variable and is characterised by its bell-shaped curve, which is symmetric about the mean. The general form of the probability density function of a normal distribution is given by:

$$\mathcal{N}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{3.34}$$

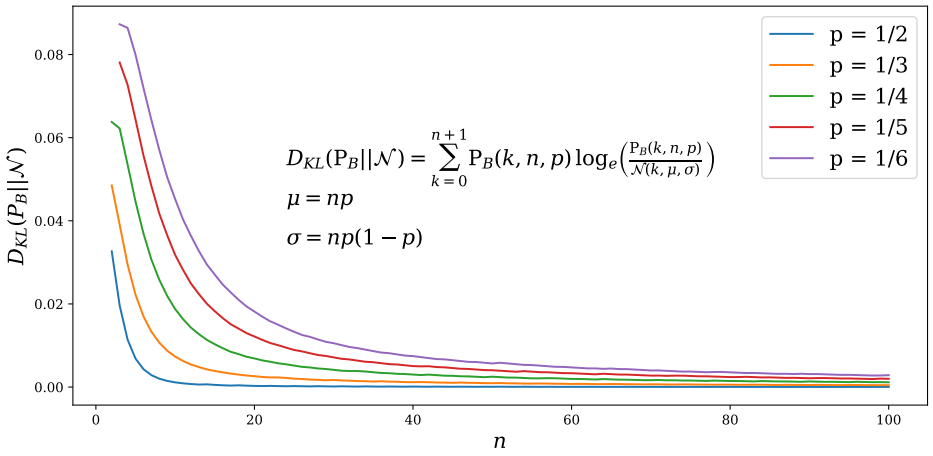


Figure 3.4.: The Kullback–Leibler divergence between the Binomial (P_B) and normal ($\mathcal{N}$) distributions as n increases. The compared values are taken as integers between the 0.01% and 99.99% extent of the binomial distribution evaluated at n and p. As n increases, the normal distribution becomes a better approximation to the Binomial distribution as per the Central Limit Theorem (CLT).

where μ is the mean of the distribution, and σ^2 is the variance. The mean μ represents the location parameter, indicating the centre of the distribution, while the variance σ^2 describes the spread or dispersion around the mean. The scale parameter or standard deviation, σ, is an important metric in experimental physics as it provides a standardised description of distance from the mean; this will be discussed further in Chapter 4.2 The normal distribution exhibits several key properties; first, as noted earlier, it is symmetric about its mean, implying that $P(X \leq \mu) = P(X \geq \mu) = 0.5$. The distribution is unimodal, meaning it has a single peak at the mean μ and is fully determined by its first two moments: the mean μ and the variance σ^2. The normal distribution is also known for its asymptotic behaviour; as x moves further away from the mean, the PDF approaches but never reaches zero. This implies that all possible outcomes, no matter how extreme, have a non-zero probability, although very large deviations from the mean are highly unlikely. A particular case of the normal distribution is the standard normal distribution ($\mathcal{N}(0,1)$), which has a mean of zero and a variance of one. The "standardised" or "reduced" form of the normal PDF is often written in terms of z

$$f(z) = \mathcal{N}(z,0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) , \tag{3.35}$$

where $z = \frac{x-\mu_x}{\sigma_x}$ is the standard score, or z-score. The standard normal distribution or normalised Gaussian is often used in statistical hypothesis testing, particularly in

z-tests, where sample means are compared to population means, assuming the underlying population follows a normal distribution.

A series of coin flips were used as an example of how the binomial distribution can be used to calculate the probability of successful trials, i.e. the number of heads (k) from n flips (Section 3.2.2). In the lower plots of Figure 3.2, it was shown as n continued to increase, the distribution began to approximate a normal distribution. Figure 3.4 shows this convergence can be seen using the Kullback–Leibler divergence (D_{KL}, Equation (3.32)). This convergence is consistent with the Central Limit Theorem (CLT), a fundamental result in probability theory and statistics. CLT states that, given a sufficiently large sample size n, the distribution of the sum (or equivalently, the sample mean) of independent and identically distributed (i.i.d.) random variables, each with finite mean $E[X_i] = \mu$ and variance $V[X_i] = \sigma^2$, will tend to approximate a normal distribution, regardless of the original distribution of the variables. Mathematically, if $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with mean μ and variance σ^2, the standardised sum Z_n is defined as:

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \,. \tag{3.36}$$

As n approaches infinity, the distribution of Z_n converges to a normal distribution $\mathcal{N}(0,1)$ [128]. Specifically, the convergence will satisfy the following equation:

$$\lim_{n\to\infty} P\big\{Z_n \leq x\big\} = P_X(x) \,, \tag{3.37}$$

where $P_X(x)$ is the cumulative distribution of the *standard* normal variable given by:

$$P_X(x) = \int_{-\infty}^{x} \mathcal{N}(0,1)\, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, \mathrm{d}x \,, \tag{3.38}$$

This result is crucial because it justifies using normal distribution approximations where n is very large, even if the underlying data is not normally distributed.

3.2.5. The χ^2 distribution

The χ^2 distribution is widely used in physics and statistics, particularly in the context of hypothesis testing and inferential statistics. It is a continuous probability distribution that is defined as a sum of the squares of k independent standard normal random variables.

Mathematically, if $Z_1, Z_2, \ldots, Z_k$ are independent and identically distributed standard normal variables, then the random variable X given by:

$$X = \sum_{i=1}^{k} Z_i^2 \,, \tag{3.39}$$

where X follows a χ^2 distribution with k degrees of freedom, denoted as $X \sim \chi^2(k)$. The parameter k, known as the degrees of freedom, is a positive integer that is important in determining the shape of the distribution. The probability density function (PDF) of the χ^2 distribution is given by:

$$f(x, k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} \quad \text{for } x > 0 \,, \tag{3.40}$$

where $\Gamma(\cdot)$ denotes the Gamma function, the commonly used extension of the factorial function, defined as:

$$\Gamma(k) = \int_0^\infty t^{k-1} e^{-t}\, dt \,. \tag{3.41}$$

It is important to note that the χ^2 distribution is a specific case of the gamma distribution, a distribution parameterised by a shape parameter κ and a scale parameter θ, has a PDF given by

$$f(x, \kappa, \theta) = \frac{1}{\Gamma(\kappa)\theta^\kappa} x^{\kappa-1} e^{-x/\theta}, \quad x > 0 \,. \tag{3.42}$$

The shape parameter κ governs the distribution's form, and the scale parameter θ stretches or compresses the distribution along the x-axis. When κ is an integer, the gamma distribution is known as the Erlang distribution, which describes the time until the n^{th} event of a Poisson process with a rate of λ [129]. The χ^2 distribution emerges as a special case of the gamma distribution when the shape parameter is $\kappa = \frac{k}{2}$ and the scale parameter is $\theta = 2$, where k represents the degrees of freedom.

Returning to the χ^2 distribution, the PDF is positively skewed, especially for small k. As k increases, the distribution becomes more symmetric and approaches a normal distribution in accordance with Central Limit Theorem. The mean and variance of the

χ^2 distribution are directly related to the degrees of freedom:

$$\begin{gathered} \mathrm{E}[X] = k \\ \text{and} \\ \mathrm{Var}[X] = 2k\,. \end{gathered} \tag{3.43}$$

These properties imply that as the degrees of freedom increase, the mean and variance both increase linearly, leading to a broader distribution. Additionally, the skewness of the distribution is given by $\sqrt{8/k}$, which indicates that the distribution becomes less skewed as k increases.

The χ^2 distribution is widely used in various statistical procedures. One of its primary applications is in the χ^2 test, which assesses the goodness-of-fit of an observed distribution to a theoretical one. In the context of a Poisson random variable where E_i is the variance, the χ^2 statistic is computed as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}\,, \tag{3.44}$$

where O_i and E_i are the observed and expected data, respectively. This statistic follows a χ^2 distribution under the null hypothesis, determining the significance of deviations between observed and expected values. A more generalised test can be constructed using the uncertainty on the observation σ_i

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{\sigma_i^2}\,. \tag{3.45}$$

In linear algebra, the χ^2 statistic can be expressed as a quadratic form involving the covariance matrix of the residuals Σ:

$$\chi^2 = \mathbf{X}^T \Sigma^{-1} \mathbf{X}\,, \tag{3.46}$$

where $\mathbf{X} = \mathbf{O} - \mathbf{E}$ is the vector of residuals (differences between observed and expected values), and Σ is the covariance matrix of the residuals. In many cases, Σ is a diagonal matrix with the expected values E_i on the diagonal, i.e., $\Sigma = \mathrm{diag}(E_i)$

$$\chi^2 = \mathbf{X}^T \mathrm{diag}\left(\frac{1}{E_i}\right) \mathbf{X} = \sum_i \frac{X_i^2}{E_i}\,. \tag{3.47}$$

This returns the traditional scalar expression of the χ^2 statistic shown in Equation (3.44).

3.2.6. The beta distribution

The beta distribution is a continuous probability distribution defined over the interval $[0, 1]$ and is characterised by two shape parameters, α and β. It is often used to model random variables that represent probabilities, proportions or rates that lie within the interval $[0, 1]$, which makes it particularly useful in Bayesian statistics and various applications in experimental physics [130, 131]. The probability density function (PDF) of the beta distribution is given by [132]:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \tag{3.48}$$

where $0 \leq x \leq 1$, $\alpha > 0$, $\beta > 0$, and $B(\alpha, \beta)$ is the beta function, defined as:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt\,. \tag{3.49}$$

The beta function serves as a normalising constant to ensure that the integral of the probability density function over $[0, 1]$ equals 1. For integers α and β, the beta function can be expressed in terms of the Gamma function as

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,. \tag{3.50}$$

This derivation links the beta distribution with the Gamma function (Equation (3.41)), connecting it to other statistical distributions, such as the Gamma distribution. The first two moments, mean and variance, are defined as

$$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\,. \tag{3.51}$$

The shape parameters α and β control the behaviour of the distribution. Specifically, α governs the distribution's behaviour near $x = 0$, while β affects the shape near $x = 1$. When $\alpha = \beta = 1$, the beta distribution reduces to the uniform distribution over $[0, 1]$. When $\alpha > \beta$, the distribution skews towards $x = 1$, and when $\alpha < \beta$, it skews towards $x = 0$ [133]. These shape parameters provide flexibility in modelling a wide range of probabilistic behaviours.

The beta distribution finds significant use in Bayesian statistics, where the prior probability $P(\theta)$ can be chosen such that the posterior probability $P(\theta \mid D)$ takes the same parametric form as the prior probability. Such a prior choice is called a conjugate prior for the likelihood $P(D \mid \theta)$ [134, 135]. The beta distribution often serves as the conjugate prior to the binomial distribution. This property simplifies the process of updating the distribution in light of new experimental data. Consider a scenario where the success probability of a particular outcome in an experiment, such as detecting a specific particle, is unknown. A beta distribution with prior parameters α_0 and β_0 can be chosen to represent prior beliefs about the probability of success. After observing n trials with k successes, Bayes' theorem is applied to update the prior distribution, resulting in a posterior beta distribution with updated parameters [136]

$$\alpha_{\text{posterior}} = \alpha_0 + k, \quad \beta_{\text{posterior}} = \beta_0 + (n - k)\,. \tag{3.52}$$

This updated distribution provides a refined estimate of the probability of success, accounting for both prior knowledge and the new data. The conjugacy of the beta distribution in this context is highly advantageous, as it allows for analytical tractability in Bayesian inference. This makes it popular for modelling uncertainties in detection probabilities, branching ratios, and other binomial processes.

Beyond its role in Bayesian inference, the beta distribution is also utilised in parameter estimation techniques such as maximum likelihood estimation (MLE) and the construction of credible intervals for proportions. Given its support on the unit interval, it is well-suited for modelling uncertainties in Monte Carlo simulations commonly used in particle physics, where random variables confined to a range between 0 and 1 are frequent. For instance, the beta distribution can model efficiency measurements, where the efficiency is constrained to lie between 0 and 1, and the beta distribution naturally describes uncertainties around these estimates.

The beta distribution is a versatile tool in particle physics, particularly in the context of Bayesian analysis and Monte Carlo simulations. Its ability to model probabilities and proportions makes it especially useful for handling binomial processes, parameter estimation, and the quantification of uncertainties. The connection between the beta distribution and the Gamma function further reinforces its importance in statistical modelling, allowing for flexible and analytically tractable applications across various probabilistic problems in the field.

3.3. The likelihood

In Section 3.1, the likelihood function was introduced within the framework of Bayesian statistics. More generally, the "likelihood" is central to statistical data analysis. While it is closely related to "probability," it is essential to understand the distinction between these two terms. In everyday language, "probability" and "likelihood" are often used interchangeably; however, they have distinct meanings within the formal context of statistics. This section will investigate the theoretical underpinnings of the likelihood function and its role in statistical inference. Subsequently, the discussion will transition to various applications of the likelihood function in collider physics. These applications include but are not limited to the estimation of parameters, the construction of confidence intervals, and the implementation of hypothesis testing across different scenarios.

3.3.1. The likelihood function

Fundamentally, the likelihood function, $\mathcal{L}(x|\theta)$, quantifies the probability of observing the data x given a set of parameters θ that define the theoretical model. It is defined as $\mathcal{L}(x|\theta) = P(\text{data}|\theta)$, where $P(\text{data}|\theta)$ is the probability of the observed data given the parameters θ. It is important to note that θ could be a known or unknown parameter or vector of parameters in the parameter space $\boldsymbol{\Theta}$. It is often the case that θ is a vector. If the parameters of interest are a subset of the available components, the remaining parameters are referred to as nuisance parameters [137].

Constructing a likelihood function that describes a physical process often involves complex processes, including the convolution of experimental influence with theoretical predictions. For example, in searching for a Higgs boson, the likelihood function must account for both the signal and background processes. The signal process is modelled by the expected distribution of events based on the hypothesised production and decay of the Higgs boson, parameterised by θ_s. The background processes, which can mimic the signal, are modelled by a separate set of parameters θ_b. The total likelihood function is then the product of the likelihoods for the signal and background components, reflecting the combined probability of observing the data under both hypotheses. Mathematically, if n independent events are observed, each with a probability $P(x_i|\theta)$, the likelihood

function is given by

$$\mathcal{L}(x|\theta) = \prod_{i=1}^{n} P(x_i|\theta). \tag{3.53}$$

The application and interpretation of the likelihood function $\mathcal{L}(x|\theta)$ falls into multiple schools of thought. In Section 3.1.1, the Bayesian interpretation of the likelihood was presented as a distinct part of Bayes' theorem. It is also worth considering three other interpretations, namely the Fisherian, Neyman-Pearson and Likelihoodist frameworks [121].

3.3.2. The Fisherian interpretation

The concept of likelihood, as introduced by Sir Ronald A. Fisher, plays an important role in the development of statistical theory. The Fisherian approach to likelihoods provides a framework through which the plausibility of different parameter values, given observed data, can be assessed. Fisher's development of the likelihood function and its related concepts, such as maximum likelihood estimation (MLE), has profoundly influenced modern statistical methods [138].

Although closely related, the likelihood function is distinct from the probability function in the Fisherian framework. Consider a random variable X with a probability density function $f(x|\theta)$, where θ is a parameter (or vector of parameters) characterising the distribution. Consistent with previous definitions, the probability density function, $f(x|\theta)$, is viewed as a function of x for a fixed θ. In contrast, the likelihood function is obtained by considering x as fixed, while θ is treated as the variable. Formally, the likelihood function is defined as:

$$\mathcal{L}(\theta|x) = f(x|\theta)\,. \tag{3.54}$$

This distinction is crucial because, in frequentist inference, the observed data x is treated as fixed, and the parameter θ is unknown[1]. The likelihood function, therefore, represents the plausibility of different values of θ given the observed data. Fisher emphasised that the likelihood function should not be interpreted as a probability distribution over θ; instead, it is a measure of the relative support provided by the data for various values of θ [139].

[1]Note here that the position of arguments x and θ can switch depending on the specific interpretation

One of Fisher's key contributions to statistical inference is the maximum likelihood estimation method. Given a sample of observed data $x = (x_1, x_2, \ldots, x_n)$, the maximum-likelihood estimate (MLE) of θ is the value that maximises the likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \mathcal{L}(\theta|x)\,. \tag{3.55}$$

In practice, it is often more convenient to work with the log-likelihood function, defined as:

$$l(\theta|x) = \log \mathcal{L}(\theta|x) = \sum_{i=1}^{n} \log f(x_i|\theta)\,. \tag{3.56}$$

The log-likelihood function retains the properties of the likelihood function and has two key advantages. Firstly, it is numerically stable when dealing with exponential distributions like the Gaussian or Poisson. Secondly, it is often easier to differentiate, making it easier to identify the MLE via gradient-based approaches such as the score equation $s(\theta|x)$, which is derived from the first derivative of the log-likelihood function wrt θ:

$$s(\theta|x) = \frac{\partial l(\theta|x)}{\partial \theta}\,. \tag{3.57}$$

The Fisher information, which quantifies the amount of information the data provides about the parameter θ, is related to the second derivative of the log-likelihood function:

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2 l(\theta|X)}{\partial \theta^2}\right]\,. \tag{3.58}$$

This quantity plays a central role in the asymptotic properties of the MLE, including the derivation of its variance and the construction of confidence intervals (see Section 4.2) [140].

The likelihood principle, which arises naturally from the Fisherian approach, states that the likelihood function captures all the information about the parameter θ contained in the data. More formally, if two likelihood functions are proportional to each other, they provide the same information for statistical inference. That is, if

$$\mathcal{L}(\theta|x) \propto \mathcal{L}(\theta|x')\,, \tag{3.59}$$

for all θ, where x and x' are two different data sets, then any inference about θ should be identical whether one uses x or x' [141]. This principle implies that the actual probability model generating the data is irrelevant for inference as long as the likelihood function remains unchanged. Consequently, Fisher's approach stands in contrast to methods that depend on the sampling distribution of estimators, such as those rooted in the Neyman-Pearson framework [142].

An important feature of the likelihood function is its invariance under reparameterisation. Suppose the parameter θ is reparameterised as $\phi = g(\theta)$, where g is a one-to-one function. The likelihood function in terms of ϕ is:

$$\mathcal{L}(\phi|x) = \mathcal{L}(g^{-1}(\phi)|x)\,. \tag{3.60}$$

The invariance property ensures that the MLE of ϕ is simply the transformation of the MLE of θ:

$$\hat{\phi}_{\text{MLE}} = g(\hat{\theta}_{\text{MLE}})\,. \tag{3.61}$$

This property is advantageous when dealing with complex models where a reparameterisation can simplify the likelihood function.

Despite its widespread use, the Fisherian approach to likelihoods is not without criticism. One limitation is that the MLE can exhibit significant bias, particularly in small samples. Various modifications and extensions to the Fisherian approach have been developed in response to such concerns. For example, bias-corrected estimators and penalised likelihood methods, such as L1 and L2 regularization, have been proposed to address this bias by adding penalty terms to overly complex models and handling situations where the number of parameters exceeds the number of observations [143].

A final interesting observation comes from interpreting the likelihood function itself. While Fisher asserted that the likelihood function provided a measure of the "support" for different parameter values, he did not formalise this concept as a probability distribution over parameters. This contrasts with the Bayesian interpretation, where the likelihood is combined with a prior distribution to form a posterior distribution, providing a probabilistic interpretation of parameter uncertainty [144].

References to important works on the Fisherian interpretation include Fisher's original papers and subsequent developments by later statisticians who extended or critiqued his

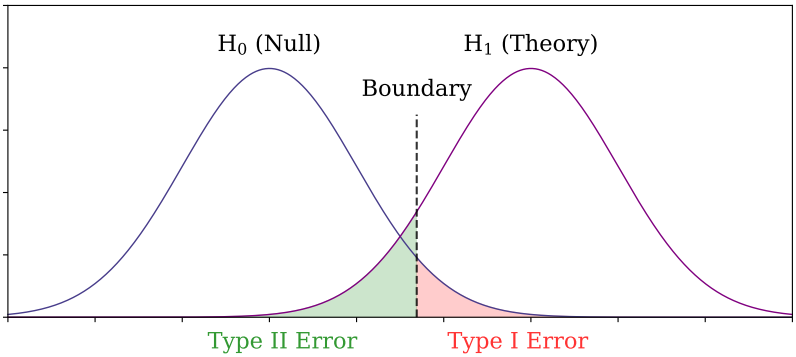


Figure 3.5.: Demonstration of Type I and II error. A Type I error (α) occurs when a null hypothesis (H_0)) is incorrectly rejected, whereas a Type II error (β) is the failure to reject a false null hypothesis when the alternative is true.

methods. Further discussion on the relative merits of the Fisherian approach, particularly in comparison to the Bayesian and frequentist approaches, can be found in [145–147].

3.3.3. The Neyman-Pearson framework

The Neyman-Pearson (NP) framework provides a formal method for evaluating hypotheses based on data, establishing a systematic procedure to decide between competing statistical models. This method is rooted in the work of Jerzy Neyman and Egon Pearson in the early 20th century [148], where they addressed the limitations of previous approaches, such as those proposed by Fisher [138–140]. The NP framework is distinguished by its emphasis on power and its formalisation of Type I and Type II errors.

The Neyman-Pearson approach begins with two competing hypotheses: the null hypothesis, denoted by H_0, and the alternative hypothesis, denoted by H_1. The null hypothesis H_0 typically represents the status quo or a baseline assumption, while H_1 represents a different, often more complex, scenario. A Type I error (α) occurs when a null hypothesis is incorrectly rejected, whereas a Type II error (β) is the failure to reject a false null hypothesis. An illustration of the error types is presented in Figure 3.5. The primary objective in the NP framework is to design a test that maximises the probability of correctly rejecting H_0 while controlling the probability of a Type I error. The Neyman-Pearson lemma provides a key result in this framework, asserting that for simple hypotheses (where both H_0 and H_1 are fully specified), the most powerful test at a given significance level α is the likelihood ratio test. Formally, the lemma states

that the test which rejects H_0 in favour of H_1 when the likelihood ratio exceeds a critical value k is the most powerful test for detecting H_1 at level α [148]:

$$\Lambda(x) = \frac{\mathcal{L}(x; H_1)}{\mathcal{L}(x; H_0)} \,. \tag{3.62}$$

Here, $\mathcal{L}(x; H)$ denotes the likelihood of observing the data x under hypothesis H. The critical value k is determined by the condition that the probability of Type I error equals α:

$$P(\Lambda(X) > k \mid H_0) = \alpha \,. \tag{3.63}$$

The likelihood ratio test (LRT) derived from the NP-lemma is a general tool applicable beyond the context of simple hypotheses. For composite hypotheses, where either H_0 or H_1 (or both) are not fully specified, the LRT remains an effective method, although additional complexities arise. In such cases, the likelihood functions involve maximisation over the parameter spaces under H_0 and H_1, leading to the generalised likelihood ratio [145, 149]:

$$\Lambda(x) = \frac{\max_{\theta \in \Theta_1} \mathcal{L}(x; \theta)}{\max_{\theta \in \Theta_0} \mathcal{L}(x; \theta)} \,, \tag{3.64}$$

where Θ_0 and Θ_1 are the parameter spaces associated with H_0 and H_1, respectively. The NP framework then prescribes rejecting H_0 if $\Lambda(x) > k$, with k determined to maintain the desired significance level α. This is similar in form to the Bayes factor from Equation (3.16), except for the Bayesian case the parameter space is marginalised and not maximised and weighted by the prior, e.g. a ratio of marginal likelihoods. This point will be further explored in Section 3.3.5.

The properties of LRTs, particularly their asymptotic behaviour, make them valuable in practice. Under regularity conditions, the distribution of the test statistic $-2 \ln \Lambda(x)$ asymptotically follows a chi-square distribution with degrees of freedom equal to the difference in dimensionality between Θ_0 and Θ_1 [150]. This result, known as Wilks' theorem (covered formally in Section 3.3.4), facilitates the determination of k and enables the use of LRTs in various applications.

In the NP framework, the concept of power plays a crucial role. The power of a test is defined as the probability of correctly rejecting the null hypothesis when the alternative

hypothesis is true:

$$\text{Power} = 1 - \beta = P(\text{Reject } H_0 \mid H_1 \text{ is true})\,. \tag{3.65}$$

Given the significance level α, the NP framework aims to maximise power. This is achieved through the likelihood ratio test, which, according to the NP-lemma, is the most powerful test for simple hypotheses. The situation is more complex for composite hypotheses, but the likelihood ratio remains a central tool for developing tests with desirable power properties.

The trade-off between α and β is a fundamental aspect of the NP approach. Lowering α reduces the risk of Type I error but generally increases β, thereby reducing power. Conversely, increasing α raises power but at the cost of a higher risk of Type I error. This trade-off must be carefully managed depending on the context, balancing errors' costs against detection's benefits.

While the NP framework provides a rigorous basis for hypothesis testing, it is not without criticism. One common critique is its strict binary decision-making process, which does not naturally accommodate situations where evidence is ambiguous or a continuous measure of support for hypotheses would be more appropriate [151]. Additionally, the NP approach assumes that hypotheses are fixed and that repeated testing is performed, which may not align with the practices of exploratory data analysis.

Extensions to the Neyman-Pearson framework have been proposed to address these limitations. For example, the Bayesian approach incorporates prior information and updates beliefs probabilistically, providing a more nuanced perspective on hypothesis testing. Furthermore, the development of false discovery rate (FDR) control techniques allows for more flexible error management in settings involving multiple comparisons, mitigating the rigid dichotomy imposed by traditional NP tests.

The Neyman-Pearson approach to likelihoods is a cornerstone of modern statistical hypothesis testing. By formalising the concepts of Type I and Type II errors and introducing the notion of the most powerful test, the NP framework provides a robust method for making decisions based on data. As derived from the NP lemma, the likelihood ratio test remains widely used, particularly valued for its optimal properties in simple and composite hypothesis testing scenarios.

Despite its limitations and the emergence of alternative frameworks, the Neyman-Pearson approach continues to influence statistical practice and theory. Its emphasis on

balancing error probabilities and optimising test power ensures that it remains relevant in various applications, from scientific research to industrial quality control. Future developments in statistical methodology will likely continue to build on the foundational principles established by Neyman and Pearson, refining and extending their approach to meet the evolving needs of data-driven decision-making.

The likelihoodist Framework

An additional and contrasting framework to consider is the likelihoodist approach. It offers a distinctive perspective within statistical inference by interpreting the likelihood function as a direct measure of evidence for different parameter values. Central to this approach is the likelihood principle, which asserts that all information relevant to the inference about a parameter is encapsulated within the likelihood function. This principle implies that once the data have been observed, the likelihood function should be the basis for making inferences about the parameter θ. The likelihood function, therefore, serves as the foundation upon which likelihoodists build their analyses, distinguishing this framework from others, particularly the frequentist and Bayesian approaches [152].

Like NP-lemma, the key tool in the likelihoodist framework is the likelihood ratio, which allows for comparing different parameter values by assessing the relative plausibility of one value of θ compared to another. Unlike other statistical frameworks, the likelihood approach does not require the specification of prior distributions (as in Bayesian inference) or the long-run frequency properties of estimators (as in frequentist inference). Instead, it focuses on the likelihood function's role as a measure of evidence, allowing for inductive reasoning based on the observed data alone [153].

Following Fisher's influence, Maximum Likelihood Estimation (MLE) is an important tool within this framework also (see Equation (3.55), Section 3.3.2). What distinguishes the likelihoodist use of MLE is the interpretation of the resulting estimate as the parameter value with the strongest evidential support rather than as an estimator with desirable long-run properties [147].

One of the main critiques of the likelihoodist framework is its lack of a formal decision rule for hypothesis testing or interval estimation. In contrast to the frequentist approach, which provides a framework for making decisions based on p-values or confidence intervals (see Sections 4.1 and 4.2), the likelihoodist perspective emphasises the interpretation of likelihoods as relative evidence without dictating specific actions. Understandably, this

has led to debates regarding the practical applicability of the likelihoodist approach in scenarios where decision-making is required.

Moreover, the likelihoodist framework faces challenges when dealing with nuisance parameters—parameters that are not of direct interest but must be accounted for in the analysis. In such cases, the likelihood function may be maximised over the nuisance parameters, but this process can lead to overfitting or misinterpreting the evidence. Additionally, when the likelihood function is not well-behaved (e.g. when it is multimodal or flat over a range of values), the inference drawn from it can be ambiguous or unreliable, limiting the robustness of this approach.

While the likelihoodist framework provides a compelling interpretation of statistical evidence through the lens of the likelihood function, it is accompanied by limitations that restrict its applicability in specific contexts. Its reliance on the likelihood principle as the basis for inference underscores a commitment to evidence-based reasoning. Yet, this commitment necessitates carefully considering the framework's boundaries and the contexts in which it is most effective [152]. The likelihoodist framework is not used in the work presented here; however, it does demonstrate that the interpretation of the likelihood function has a degree of flexibility that should not be overlooked.

3.3.4. Wilks' theorem

So far, in this section, the likelihood-ratio test (LRT) has been mentioned several times. This is because the LRT is a powerful tool for comparing the fit of two competing models. Specifically, evaluating the ratio of the maximum likelihoods of these models provides a mechanism for determining which model better explains the observed data.

Wilks' theorem provides a critical result regarding the distribution of this statistic. According to Wilks, under certain regularity conditions, when the null hypothesis H_0 is true and in the large-sample limit i.e. the number of observations n tends to infinity ($n \to \infty$), the distribution of $-2\log\lambda$ asymptotically follows a chi-square distribution with degrees of freedom equal to the difference in dimensionality between Θ and Θ_0. Specifically, if $\dim(\Theta) = k$ and $\dim(\Theta_0) = m$, then [150]:

$$-2\log\lambda \xrightarrow{d} \chi^2_{k-m} \quad \text{as} \quad n \to \infty\,, \tag{3.66}$$

where n is the sample size and $\xrightarrow{d}$ denotes convergence in distribution.

To derive Wilks' theorem, consider the general framework of the likelihood ratio test (LRT) from the Neyman-Pearson framework in Section 3.3.3. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ denote a sample of independent and identically distributed (i.i.d.) observations from a probability distribution with a density function $f(x_i; \theta)$, where θ is a vector of unknown parameters. The parameter space is denoted by Θ, and we test the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative hypothesis $H_1 : \theta \in \Theta$, where $\Theta_0 \subseteq \Theta$. To clarify the previous statement, consider a parameter vector θ that belongs to both Θ_0 (denoted as θ_0) and Θ_1 (denoted as θ). This implies that θ_0 and θ share the same dimensionality. A fundamental aspect of Wilks' theorem is that, given $\Theta_0 \subseteq \Theta$, there must exist a set of component values that are constrained to be null within the larger space Θ but satisfy the conditions imposed by Θ_0. This relationship is central to the asymptotic distribution of likelihood ratio test statistics, where the difference in dimensionality between Θ_0 and Θ determines the degrees of freedom in the limiting chi-square distribution of the test statistic [150]

The generalised likelihood ratio statistic $\Lambda(x)$, defined in Equation (3.64), can be used to define a new test statistic $-2\log\lambda$, which can be expressed as [149]:

$$-2\log\Lambda = 2\left[\log\mathcal{L}(\hat{\theta}|x) - \log\mathcal{L}(\hat{\theta}_0|x)\right], \tag{3.67}$$

where $\hat{\theta}$ is the usual maximum likelihood estimate (MLE) of θ under the alternative hypothesis, and $\hat{\theta}_0$ is the MLE under the null hypothesis. To analyse the asymptotic distribution of $-2\log\Lambda$, it is useful to Taylor-expand the log-likelihood function $\log\mathcal{L}(\theta|x)$ around $\hat{\theta}_0$. The log-likelihood function can be expanded as follows:

$$\log\mathcal{L}(\theta|x) \approx \log\mathcal{L}(\hat{\theta}_0) + (\theta-\hat{\theta}_0)^T \frac{\partial\log\mathcal{L}(\theta|x)}{\partial\theta}\bigg|_{\theta=\hat{\theta}_0} + \frac{1}{2}(\theta-\hat{\theta}_0)^T \frac{\partial^2\log\mathcal{L}(\theta|x)}{\partial\theta\,\partial\theta^T}\bigg|_{\theta=\hat{\theta}_0}(\theta-\hat{\theta}_0)\,. \tag{3.68}$$

Given that $\hat{\theta}_0$ maximises $\log\mathcal{L}(\theta_0|x)$ under the null hypothesis, the first derivative term vanishes:

$$\frac{\partial\log\mathcal{L}(\theta|x)}{\partial\theta}\bigg|_{\theta=\hat{\theta}_0} = 0\,. \tag{3.69}$$

Thus, the expansion simplifies to:

$$\log\mathcal{L}(\theta|x) \approx \log\mathcal{L}(\hat{\theta}_0) + \frac{1}{2}(\theta-\hat{\theta}_0)^T \frac{\partial^2\log\mathcal{L}(\theta|x)}{\partial\theta\,\partial\theta^T}\bigg|_{\theta=\hat{\theta}_0}(\theta-\hat{\theta}_0)\,. \tag{3.70}$$

The second derivative of the log-likelihood function with respect to θ is the observed Fisher information matrix, denoted as $\mathbf{I}_n(\hat{\theta}_0)$. The matrix is defined as

$$\mathbf{I}_n(\hat{\theta}_0) = -\frac{\partial^2 \log \mathcal{L}(\theta|x)}{\partial\theta\, \partial\theta^T}\bigg|_{\theta=\hat{\theta}_0} . \tag{3.71}$$

Under standard regularity conditions, the MLE $\hat{\theta}$ is considered to be is asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, \mathbf{I}^{-1}(\theta_0)) , \tag{3.72}$$

where $\mathbf{I}(\theta_0)$ is the expected Fisher information matrix. Since $\hat{\theta}$ maximises $\log \mathcal{L}(x|\theta)$ over Θ, the difference in log-likelihoods can be written as:

$$\log \mathcal{L}(\hat{\theta}) - \log \mathcal{L}(\hat{\theta}_0) \approx \frac{1}{2}(\hat{\theta}_0 - \hat{\theta})^T \mathbf{I}_n(\hat{\theta}_0)(\hat{\theta}_0 - \hat{\theta}) . \tag{3.73}$$

Substituting this into the expression for $-2\log\Lambda$ gives:

$$-2\log\Lambda \approx (\hat{\theta} - \hat{\theta}_0)^T \mathbf{I}_n(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0) . \tag{3.74}$$

Under the null hypothesis, $\hat{\theta} \in \Theta_0$. Therefore, the difference $\hat{\theta} - \hat{\theta}_0$ lies approximately in a subspace of $\mathbb{R}^k$, where k is the number of free parameters under H_1. The dimensionality of this subspace is $d = \dim(\Theta) - \dim(\Theta_0)$. The quadratic form $(\hat{\theta} - \hat{\theta}_0)^T \mathbf{I}_n(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0)$ converges on Equation (3.46) and thus follows a chi-square distribution with d degrees of freedom [154].

Wilks' theorem holds under several regularity conditions, which include the assumptions that the true parameter lies within the interior of the parameter space, the models are correctly specified, and the likelihood function satisfies certain smoothness conditions (e.g., differentiability). Moreover, the parameter estimates should be consistent, and the information matrix must be positive definite.

The practical implication of Wilks' theorem is significant. It enables the use of the chi-square distribution as an approximation for the distribution of the likelihood ratio test statistic in large samples, simplifying hypothesis testing. This allows researchers to compute p-values and make decisions about the null hypothesis using the chi-square distribution without relying on the exact distribution of the test statistic, which may be complex or unknown.

3.3.5. The profiled and marginalised likelihood

In statistical analysis, profiling and marginalising a likelihood are two different methods of handling nuisance parameters when estimating parameters of interest. Both methods simplify the likelihood function but do so in distinct ways, each with specific implications for the estimation process. Profiling a likelihood involves focusing on the parameter of interest while maximising the likelihood function with respect to the nuisance parameters. Specifically, if we have a likelihood function $\mathcal{L}(\theta, \psi)$, where θ is the parameter of interest and ψ represents nuisance parameters, profiling is achieved by finding the value of ψ that maximises the likelihood for each fixed value of θ. The profiled likelihood is then defined as

$$\mathcal{L}_p(\theta) = \max_{\psi} \big\{ \mathcal{L}(\theta, \psi) \big\} \,. \tag{3.75}$$

This method reduces the problem's dimensionality by removing the nuisance parameters through maximisation. The resulting likelihood function $L_p(\theta)$ depends solely on the parameter of interest θ. Profiling is particularly useful in frequentist inference, where confidence intervals (see Section 4.2) for θ can be derived from the profiled likelihood, especially when using asymptotic approximations.

Marginalising a likelihood, on the other hand, involves integrating out the nuisance parameters from the joint likelihood function. Given the same likelihood function $\mathcal{L}(\theta, \psi)$, the marginalised likelihood for θ is obtained by integrating over the nuisance parameters ψ wrt their probability distribution $\pi(\psi)$:

$$\mathcal{L}_m(\theta) = \int \mathcal{L}(\theta, \psi) \pi(\psi) \, d\psi \,. \tag{3.76}$$

This approach is commonly employed in Bayesian inference, where integration accounts for prior beliefs regarding the nuisance parameters. By marginalising over these parameters rather than optimising them, the resulting marginal likelihood $L_m(\theta)$ reflects the uncertainty associated with ψ in the estimation of θ. However, the choice of prior distribution for ψ plays a crucial role in this formulation, as it directly influences the resulting marginal likelihood. Different prior specifications can lead to substantially different inferences, particularly when the prior carries significant weight relative to the available data. This sensitivity highlights a fundamental challenge in Bayesian model selection, where the integration over nuisance parameters introduces an additional layer of subjectivity that must be carefully considered. The critical difference between profiling

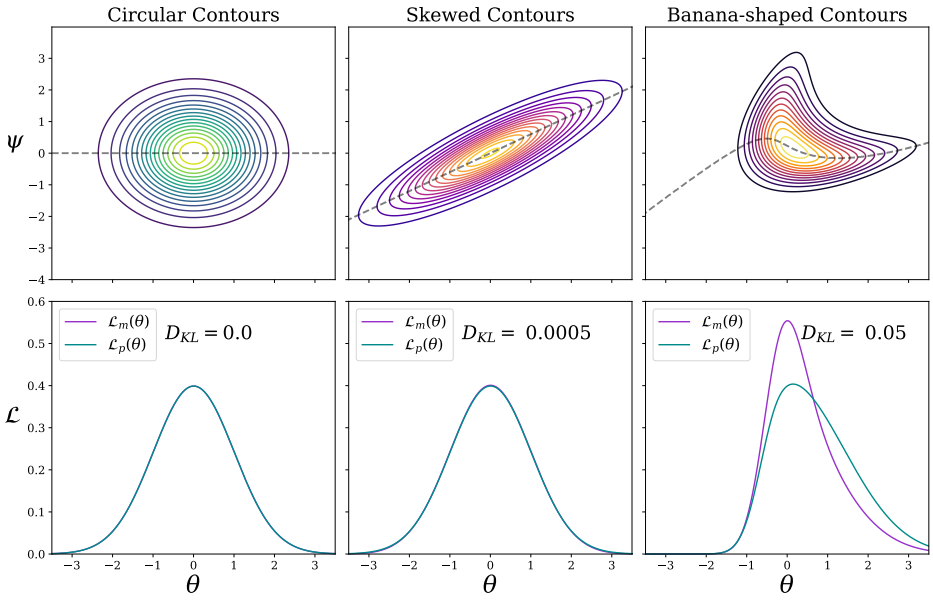


Figure 3.6.: Upper plots show three examples of multivariate likelihoods $\mathcal{L}(\theta, \psi)$, the lower plots show the results of marginalising and profiling over ψ for each of the likelihoods. The black line in the upper plot traces the value of Ψ that maximises the profiled distribution. Moving from left to right, the upper distributions become more complex, causing $\mathcal{L}_m(\theta)$ and $\mathcal{L}_p(\theta)$ to diverge. This is quantified using the Kullback-Leibler divergence D_{KL}

and marginalising lies in how the nuisance parameters are treated; profiling involves maximisation, which selects the most favourable value of the nuisance parameter for each value of θ, effectively considering only the best-case scenario for the nuisance parameter Marginalising averages over all possible values of the nuisance parameter according to a specified distribution, thereby incorporating the uncertainty associated with ψ into the estimation process. In terms of their applications, profiling is more aligned with frequentist methods and is particularly useful when the primary interest is in the point estimate and the related confidence intervals. Marginalising is intrinsic to Bayesian methods, where the goal is often to account for all sources of uncertainty, including those associated with nuisance parameters. Figure 3.6 illustrates how these differences can significantly affect the resulting estimates and their interpretations. Like many of the methods discussed so far, the choice between profiling and marginalising should be based on the statistical framework and the specific objectives of the analysis.

Chapter 4.

Hypothesis Testing

In particle physics, hypothesis testing plays a central role in determining the validity of theoretical models against experimental data. This process allows researchers to evaluate competing hypotheses, typically structured as a null hypothesis representing a well-established theory and an alternative hypothesis proposing a deviation from the expected model. The statistical framework employed in these analyses is critical to distinguishing signal from noise, especially when observed phenomena are rare or subtle. For our purposes, the null hypothesis, denoted as H_0, typically represents the SM prediction, while the alternative hypothesis, H_1, often suggests a BSM theory.

When comparing hypotheses, it is often convenient to construct a test statistic $t(X)$ derived from experimental data, X, that quantifies the difference between H_0 and H_1. Various methods are employed to compute this statistic, yielding insights into whether deviations from H_0 are statistically significant. The corresponding p-value, which measures the probability of observing data at least as extreme as the test statistic under H_0, is often used to assess the evidence against the null hypothesis. Furthermore, physicists often rely on confidence intervals or regions to estimate the parameters of interest, incorporating concepts such as Type I and Type II errors to minimise incorrect conclusions.

Hypothesis testing in particle physics is a mathematical tool and a methodological cornerstone, ensuring rigorous and systematic comparisons between theoretical predictions and empirical observations. This chapter will consider how effectively a chosen theory explains the data. A framework of hypothesis tests must exist to assess whether a theory is consistent with data, encompassing tests of a single hypothesis, comparisons between different hypotheses, the optimisation of parameters and statements of confidence in the conclusions reached.

4.1. *p*-values

The p-value provides a robust measure to quantify the strength of the evidence against the null hypothesis, H_0. The p-value is defined as the probability of obtaining an observation at least as extreme as the one observed, assuming that H_0 is true. Mathematically, if t represents a test statistic calculated from the data–a good example would be $t = -2\ln\Lambda(x)$, from Section 3.3.3–the p-value is given by [155]

$$p = P(t \geq t_{\text{obs}} \mid H_0)\,, \tag{4.1}$$

where t_{obs} is the observed value of the test statistic, and the probability is calculated under the null hypothesis distribution. When the p-value is small, it suggests that the observed data is unlikely under H_0, which suggests rejecting H_0 in favour of the alternative hypothesis H_1. However, it is essential to note that the p-value does not provide the probability that H_0 is true or false, nor does it measure the size of any effect. Instead, it quantifies how compatible the observed data is with the assumption that H_0 holds [153]. Figure 4.1 shows the standard normal distribution divided into segments corresponding to standard deviations (σ) from the mean. One can extract three quantities of interest for each unit of standard deviation taken from the mean. First is the area under the curve within the band, i.e. how much of the total probability is contained within a single band; this is given as a percentage under each segment. The second quantity is the total probability under the curve between equivalent bands, i.e. $\pm n\sigma$; this is shown as a percentage in the upper section of the Figure and roughly corresponds to the confidence interval (CI). For the case of the standard normal distribution, the CI's correspond to the empirical rule, often referred to as the 68-95-99.7 rule, which states that approximately 95% of the data points in a normal distribution fall within two standard deviations of the mean. The third quantity in Figure 4.1 is the p-value, which is given as p following Equation (4.1), it is clear that moving from left to right, the value decreases. For the case of a continuous probability distribution, the p-value is defined as

$$p = \int_{t_{\text{obs}}}^{\infty} P(t \mid H_0)\,\mathrm{d}t \tag{4.2}$$

where the integral ranges from t_{obs} to ∞ covering the proportion greater than t_{obs} following Equation (4.1).

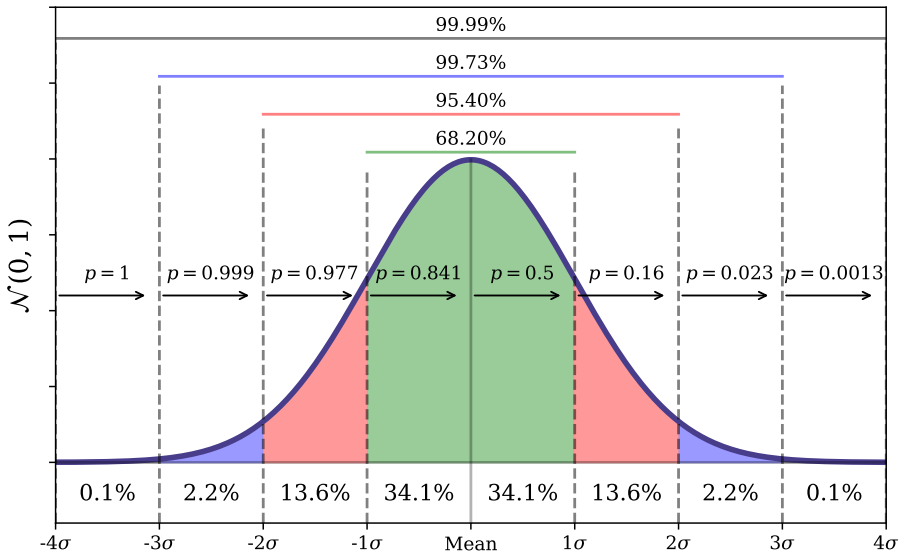


Figure 4.1.: The standard normal distribution shown in terms standard deviations (σ) from the mean. starting from the upper section of the plot, the values correspond to first, the approximate confidence intervals (CIs), then the p-values taken from σ (LHS of the arrow) to ∞ and finally the percentage of probability within a single σ band

In experimental physics, p-values are often used in the context of searching for new particles or interactions. To claim a discovery, a p-value threshold corresponding to a "5-sigma" (5σ) significance level is typically used. The threshold is a convention in particle physics used to define the level of statistical significance required to claim a discovery. It corresponds to a p-value of approximately 3×10^{-7}, or a 1 in 3.5 million chance that the observed effect is due to random fluctuations under the null hypothesis. This stringent criterion was established to minimise the frequency of type I errors, which could lead to incorrect claims of new physics. The threshold is set at 5-sigma because experience in high-energy physics has shown that lower significance levels often result in false discoveries due to unaccounted-for systematic uncertainties, background fluctuations, or other experimental errors. By requiring such a high level of statistical significance, the field aims to ensure that discoveries are both statistically and scientifically credible [156].

The calculation of the p-value often involves generating distributions of the test statistic under the null hypothesis using Monte Carlo simulations. These simulations model the expected background processes without any signal. The p-value is then obtained by integrating the tail of the distribution beyond the observed value of the

test statistic. In cases where the test statistic follows a known distribution (e.g., a chi-squared distribution), the p-value can be computed analytically. This method will be used extensively in Chapter 8 where p-values are extracted from an unknown distribution.

The uniformity of p-values under the null hypothesis is a useful property in hypothesis testing. This uniformity arises from the cumulative distribution function (CDF) properties associated with the test statistic when the null hypothesis H_0 is true. To understand this, consider the p-value as the area under the probability distribution curve of t to the right of t_{obs}. When H_0 is true, the test statistic t distribution is correctly specified, meaning that t behaves according to its known probability distribution. Now, consider the CDF of the test statistic, $F(t)$, which gives the probability that t is less than or equal to some value t_{obs}:

$$F(t_{\text{obs}}) = P(t \leq t_{\text{obs}} \mid H_0)\,. \tag{4.3}$$

The p-value can be expressed in terms of the CDF as

$$p = 1 - F(t_{\text{obs}})\,. \tag{4.4}$$

If t is a continuous random variable, then under H_0, the CDF $F(t)$ is uniformly distributed over the interval $[0, 1]$. This uniformity means that for any p-value p computed under the null hypothesis, the probability of observing a p-value within a specific interval $[a, b]$ is proportional to the length of that interval:

$$P(a \leq p \leq b) = b - a\,. \tag{4.5}$$

This relationship directly follows from the properties of the CDF and the fact that it maps the values of t onto the unit interval $[0, 1]$ in a linear manner. Consequently, when H_0 is true, the distribution of p-values across many repeated samples should be uniform. Thus, each possible p-value is equally likely, and the distribution of these p-values will be flat across the interval $[0, 1]$ [157]. This uniformity is fundamental to the interpretation of p-values in hypothesis testing. It ensures that under H_0, the likelihood of observing a p-value at any particular level is consistent, which allows significance thresholds (e.g., $\alpha = 0.05$) with a clear understanding of the associated Type I error rate. If p-values were not uniformly distributed under the null hypothesis, it would imply a bias in the test statistic or an incorrect specification of the null distribution [158].

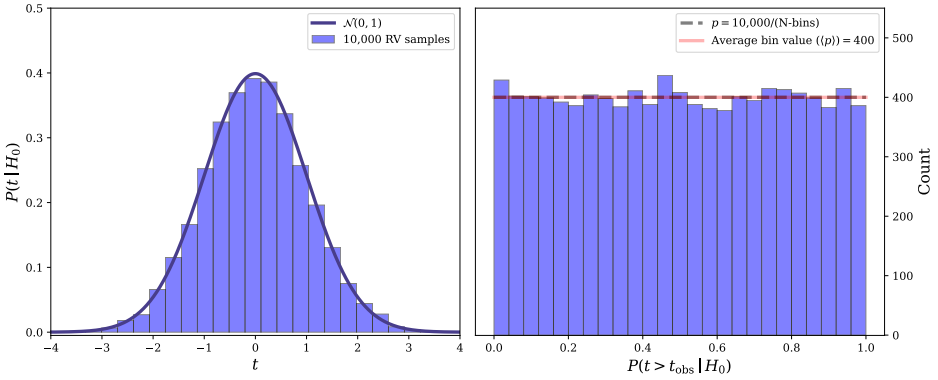


Figure 4.2.: Plot showing the uniformity of the p-value distribution. The left-hand plot shows a histogram of 10,000 randomly generated, normally distributed samples of test statistic t. The right-hand plot shows a histogram of the corresponding p-values.

Despite its widespread use, the interpretation of the p-value requires caution. A small p-value does not imply that the null hypothesis is definitively false, nor does it quantify the probability of a false positive. Consider a case where we want to use Wilks' theorem (Section 3.3.4) to compare H_0 and H_1. Knowing that $H_0 \in H_1$ and $\Delta\text{dof} = k$ we can use the χ^2 distribution with k dof to calculate the p-value. However, we can now consider the same comparison using the Bayes factor, the ratio of Equation (3.11), or

$$\frac{P(H_0 \mid D)}{P(H_1 \mid D)} = \frac{\mathcal{L}(D \mid H_0)}{\mathcal{L}(D \mid H_1)} \times \frac{\pi(H_0)}{\pi(H_1)}\,. \tag{4.6}$$

Here, instead of the frequentist likelihood ratio, a ratio of probabilities is used, the usual likelihood ratio multiplied by the ratio of the prior probabilities. Hence, in this approach, even if the likelihood ratio favours H_1, we would still prefer H_0 if our prior belief in H_1 was very low. This reweighing of the likelihood with a prior belief provides a layer of protection at the expense of the asymptotic behaviour predicted by Wilks. Therefore, it is common to complement p-values with other metrics, such as the Bayes factor or confidence intervals, to provide a more robust assessment of the evidence [156].

4.2. Confidence limits

In frequentist statistical inference, a confidence interval (CI) is a range of values derived from a dataset that is believed to contain the true value of an unknown population parameter with a specified probability, often referred to as the confidence level. The CI is closely related to the significance level denoted α, covered in Section 3.3.3, defined as

the type I error, the probability of incorrectly rejecting the null hypothesis. Formally, if θ represents the parameter of interest and $\hat{\theta}$ denotes its estimator, then the confidence interval for θ can be expressed as [155]

$$\left[\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}}\,, \hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}\right]\,, \tag{4.7}$$

where $\sigma_{\hat{\theta}}$ is the standard error of the estimator $\hat{\theta}$, and $z_{\alpha/2}$ corresponds to the critical value from the standard normal distribution associated with a confidence level of $1-\alpha$. The confidence level, often expressed as a percentage (e.g., 95% or 99%), indicates the frequency with which the true parameter would fall within the interval if the experiment were repeated multiple times. For example, a 95% confidence interval implies that the interval would capture the true parameter value in 95 out of 100 samples drawn from the population; this interval is approximated as the 95.4% line in Figure 4.1.

It is crucial to distinguish that the CI does not provide a probability statement about the parameter itself; instead, it reflects the long-term reliability of the estimation process. The interpretation of the CI must align with the frequentist framework: it is incorrect to state that there is a 95% probability that the true parameter lies within a specific calculated interval. The correct interpretation is that 95% of similarly constructed intervals would contain the parameter [137].

The bounds of the CI are called confidence limits (CL). Specifically, the lower confidence limit (LCL) and upper confidence limit (UCL) denote the smallest and largest values of the interval, respectively. For the two-sided CI described above, the CLs can be expressed as

$$\begin{aligned} \text{LCL} &= \hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}}, \\ \text{UCL} &= \hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}\,. \end{aligned} \tag{4.8}$$

Confidence limits serve as practical markers in hypothesis testing and the evaluation of statistical significance. For example, in a hypothesis test for a population mean, if the null hypothesis value is outside the CL, it can be rejected at the corresponding significance level α. Conversely, if the null hypothesis value lies within the CL, it cannot be dismissed [159, 160].

The LHC has a well-defined convention for specifying confidence levels for the background and signal-plus-background hypotheses [161, 162]. Here, CL_{sb} and CL_{b} represent the 95% confidence limits for the signal-plus-background and background hypotheses,

respectively. The confidence level for the signal $\mathrm{CL_s}$ is then defined as:

$$\mathrm{CL_s} = \frac{\mathrm{CL_{sb}}}{\mathrm{CL_b}} \equiv \frac{p_{\mathrm{sb}}}{1 - p_{\mathrm{b}}} < \alpha \tag{4.9}$$

where $\alpha = 0.05$ and p_{sb}, p_{b} defined as:

$$\begin{aligned} p_{\mathrm{sb}} &= P(t \geq t_{\mathrm{obs}} \mid \mathrm{s} + \mathrm{b}) = \int_{t_{\mathrm{obs}}}^{\infty} f(t \mid \mathrm{s} + \mathrm{b})\, \mathrm{d}t\,, \\ p_{\mathrm{b}} &= P(t \leq t_{\mathrm{obs}} \mid \mathrm{b}) = \int_{-\infty}^{t_{\mathrm{obs}}} f(t \mid \mathrm{b})\, \mathrm{d}t\,. \end{aligned} \tag{4.10}$$

Despite their widespread use, confidence intervals are not without limitations. The accuracy of a CI depends on the validity of the underlying assumptions, such as the normality of the estimator distribution or the correct specification of the model. Violations of these assumptions can lead to misleading intervals. Additionally, the width of a CI is inversely related to sample size; small samples yield wider intervals, reflecting greater uncertainty. These factors necessitate careful consideration when interpreting CIs in practical applications.

Confidence intervals and their corresponding limits are effective tools in statistical analysis, offering a structured approach to quantifying and communicating uncertainty. Their proper interpretation and use, however, require careful attention to the underlying assumptions and the context of the data analysis. As with all statistical methods, they should be applied thoughtfully, with an awareness of their strengths and limitations.

4.3. The χ^2 test

The χ^2 test is a non-parametric statistical method commonly employed in hypothesis testing to assess the relationship between categorical variables. This test is particularly useful in situations where the data do not meet the assumptions of parametric tests, such as normality or homoscedasticity (also known as the homogeneity of variance). The χ^2 test, through its relation to the χ^2 distribution introduced in section 3.2.5, facilitates the evaluation of observed versus expected frequencies across different categories. This approach is used to determine whether there is a significant association between variables or whether the distribution of a categorical variable differs from a hypothesised distribution.

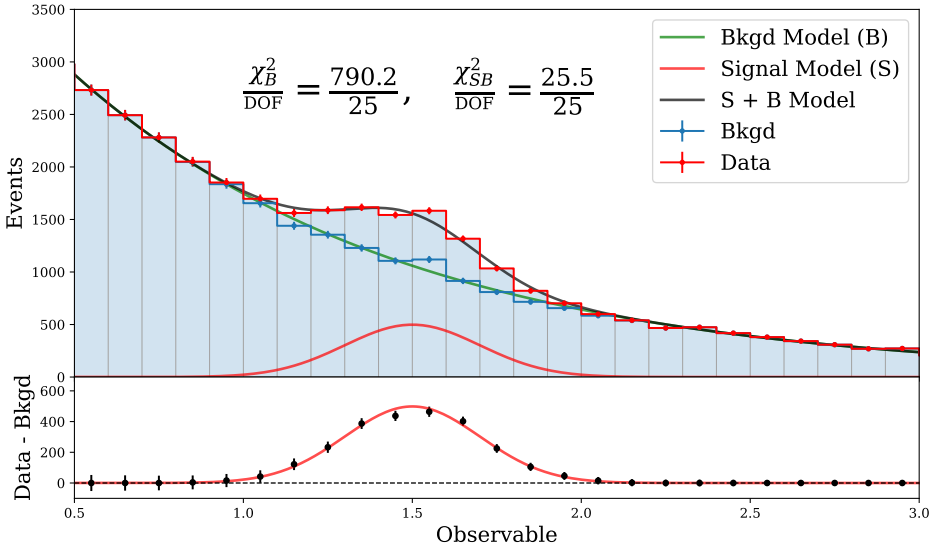


Figure 4.3.: Demonstration of the χ^2 test statistic using a toy data set. The data simulates the comparison between a background (B) and a background-plus -signal (B + S) model. The histogram and red error bar show the simulated S + B data, while the blue error bar shows the background-only hypothesis. The ratio of the χ^2_{SB}/DOF is approximately one, suggesting that the data is a good fit to the S + B model.

The χ^2 test has two primary forms: the χ^2 goodness-of-fit test and the χ^2 test of independence. The goodness-of-fit test determines if a sample distribution fits a specific theoretical distribution. For example, it can be used to test whether a sample's frequency of observed events aligns with the expected frequencies derived from a uniform distribution or any other specified distribution. The test statistic for the goodness-of-fit test is calculated Equation (3.44) [155]. Where, for this case, O_i is the observed frequency for category i with E_i as the expected frequency for that category, assuming Poisson statistics with n categories. This results in a test statistic that follows a χ^2 distribution with degrees of freedom (DOF) given by $k = n - 1$, where the subtraction of one accounts for the constraint that the total observed frequency $\sum O_i$ must equal the total expected frequency $\sum E_i$. The computed χ^2 statistic is then compared to the critical value from the χ^2 distribution with $k = n - 1$ DOF, as discussed in Section 3.2.5.

Figure 4.3 shows an example of the goodness of fit test using background (B) and signal-plus-background (S + B) models. The filled histogram and red error bars show the S + B model, while the blue error bars show the B-only hypothesis; both data sets are assumed to follow Poissonian statistics and thus have a per-bin error of $\sqrt{N}$. The χ^2 goodness of fit has been evaluated for both the B and S + B models using 26 bins, giving a

DOF of 25. Looking at the figure, it is clear that the S + B model is the better fit, and this is corroborated by the test statistic. A statistically significant result indicates that the observed frequencies deviate from the expected frequencies more than would be expected by chance. In the case of Figure 4.3, if we were to define exclusion of the background-only hypothesis as a p-value < 0.05, this would correspond to a $\chi^2_{\mathrm{B}} > 38$. Thus, an observed value of 875 would strongly favour excluding the background hypothesis. It is common practice to present the χ^2 score as a fraction in terms of the DOF. This convention allows for quick evaluation of the fit, as a value close to unity indicates a preference towards the chosen model and a correctly calibrated uncertainty.

An alternative χ^2 test is the test of independence used, unsurprisingly, to evaluate whether two categorical variables are independent. This test is commonly applied to contingency tables where the relationship between two variables is of interest. The test statistic is calculated like the goodness-of-fit test [163]

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}\,, \tag{4.11}$$

where O_{ij} is the observed frequency corresponding to the ith row and jth column of a matrix, E_{ij} is the expected frequency under the assumption of independence, r is the number of rows, and c is the number of columns. The expected frequency E_{ij} is calculated as:

$$E_{ij} = \frac{R_i \cdot C_j}{N}\,, \tag{4.12}$$

where R_i is the total frequency for the ith row, C_j is the total frequency for the jth column, and N is the overall sample size. The degrees of freedom for this test are given by $(r-1) \times (c-1)$.

Both forms of the χ^2 test hinge on the assumption that the expected frequency in each category is sufficiently large, typically at least 5, to ensure the validity of the test results. When this assumption is violated, the χ^2 statistic may not closely follow the χ^2 distribution, leading to potentially inaccurate conclusions. In such cases, alternative methods, such as Fisher's exact test, may be more appropriate [164].

With very large samples, even small deviations from the expected frequencies can produce a large χ^2 statistic, potentially leading to the rejection of the null hypothesis even when the practical significance of the deviation is negligible. When errors are inflated

or overly conservative, each term in the sum is reduced, meaning the total χ^2 statistic decreases. This results in a χ^2 value that is artificially closer to the expected range for a good fit, making it more likely that the model appears to fit the data well, even if there are significant discrepancies. This increases the risk of Type II errors because the inflated errors make the model look better than it is. With these conditions in mind, the results of a χ^2 test should be interpreted with caution, considering both the statistical significance and the effect size.

4.4. Parameter estimation

Parameter estimation is the process of determining the values of unknown parameters within a model that best describes the observed data. A model's accuracy and reliability are heavily dependent on the correct estimation of these parameters. Therefore, effective parameter estimation is crucial for making valid inferences and predictions from the model.

Parameter estimation can be approached through various methods, with three of the most widely used being Maximum Likelihood Estimation (MLE), Bayesian inference and χ^2 minimisation. Each approach provides a framework for estimating parameters based on the available data, but they differ significantly in their underlying philosophies and implementation.

It is important to note here that the mathematics of MLE and Bayesian inference are practically identical to that of profiling and marginalising in Section 3.3. This is also true for the χ^2 method, where the derivations were covered in Sections 4.3 and 3.2.5. However, in the context of parameter estimation, the implementation and interpretations are slightly different, so we will take a quantitative approach to their application and focus on assessing uncertainty in the estimate.

In Section 4.3, a toy model was used to demonstrate the use case of the χ^2 test for model comparison. The toy model comprised an exponentially decaying background with a Gaussian signal, giving a model with four free parameters, the signal mean (μ), the signal standard deviation (σ), the background decay rate λ and the signal fraction f.

$$P(x; f, \mu, \sigma, \lambda) = f \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) + (1-f) \times \lambda \exp(-\lambda x)\,. \tag{4.13}$$

We can assume that the background is well understood and fix the decay rate parameter, which leaves three free parameters of interest. Maintaining continuity with Equation (4.13) and Figure 4.3, the parameters μ, σ and λ are 1.5, 0.5 and 0.05 respectively. Using this toy model, we can now apply the parameter estimation methods, compare the results and discuss the relative pros and cons.

Maximum-likelihood estimation (MLE)

Maximum Likelihood Estimation is used to estimate the parameters of a model by maximising the likelihood function. The likelihood function, denoted as $\mathcal{L}(\theta)$, represents the probability of observing the given data X as a function of the parameters θ. The goal of MLE is to find the parameter values $\hat{\theta}$ that maximise this likelihood function:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\theta)\,. \tag{4.14}$$

In practice, the logarithm of the likelihood function, known as the log-likelihood, is often used instead of the likelihood itself due to its computational simplicity and numerical stability. The log-likelihood function is given by:

$$\ell(\theta)) = \log \mathcal{L}(\theta))\,, \tag{4.15}$$

and the MLE is then obtained by solving:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta)\,. \tag{4.16}$$

In some instances, this equation can be solved analytically by taking the derivative of $\ell(\theta)$ with respect to θ, setting the derivative equal to zero, and solving for θ. Analytically solving for the MLE becomes impractical when dealing with high-dimensional spaces, complex models, or non-linear relationships. For multi-parameter models, the likelihood function depends on several parameters simultaneously, necessitating numerical optimisation techniques. In addition to high dimensionality, the likelihood function $\mathcal{L}(\theta_1, \theta_2, \ldots, \theta_k)$ is often constructed as a product of probability density functions $\mathcal{L} = \prod_i P(\theta)_i$ Thus, maximising this function's logarithm is easier to handle due to its addictive properties.

To numerically compute the MLE in such cases, algorithms such as the Nelder-Mead [165] method, the Expectation-Maximization (EM) algorithm [166], and gradient-based techniques like Broyden-Fletcher-Goldfarb-Shanno (BFGS) [167] are commonly employed.

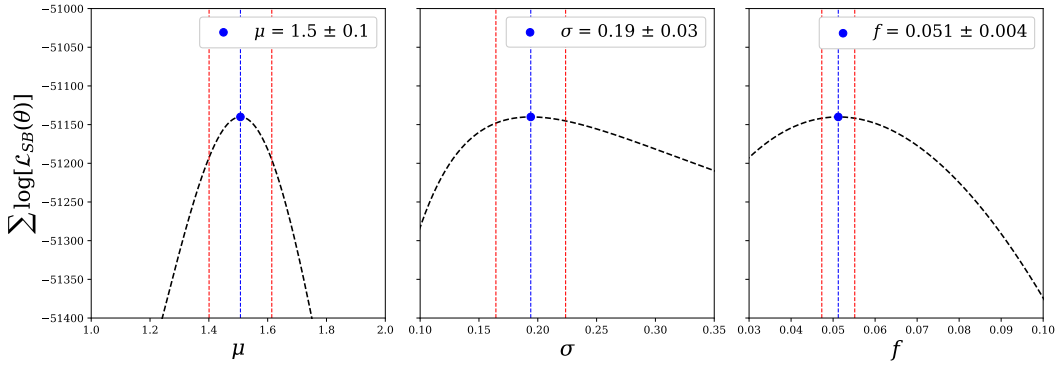


Figure 4.4.: Application of the MLE method to a three-parameter toy model, where each plot shows the maximum likelihood $l(\theta)$ estimate for each parameter (blue dashed line) with an associated one sigma (σ) error band (red dashed line) estimated using the inverse hessian matrix. For this simple model, the MLE method was executed using a BFGS maximisation algorithm, which correctly estimated each parameter to within one standard deviation

These methods iteratively search for the parameter values that satisfy the condition

$$\nabla_\theta \log \mathcal{L}(\theta_1, \theta_2, \ldots, \theta_k) = 0\,, \tag{4.17}$$

where ∇_θ is the gradient with respect to the parameters. In multi-parameter models, each step simultaneously updates all parameters, often requiring the computation of the Hessian matrix or its approximations to capture second-order information. The Hessian matrix is often used to understand the curvature of the log-likelihood function and the uncertainty of parameter estimates. When the log-likelihood function is maximised, the Hessian matrix $\mathbf{H}(\theta)$ is the matrix of second-order partial derivatives of the log-likelihood with respect to the parameters θ_i. Specifically, the Hessian for the log-likelihood function is defined as:

$$\mathbf{H}(\theta) = \begin{bmatrix} \frac{\partial^2 l(\theta)}{\partial \theta_1^2} & \frac{\partial^2 l(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta)}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 l(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l(\theta)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l(\theta)}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\theta)}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 l(\theta)}{\partial \theta_k \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta)}{\partial \theta_k^2} \end{bmatrix}. \tag{4.18}$$

The Hessian matrix is closely related to the Fisher information $I(\theta)$ from Equation (3.58) where $I(\theta) = -\mathrm{E}[\mathbf{H}(\theta)]$. Near the MLE solution $\hat{\theta}$, the negative inverse of the Hessian matrix provides an approximation for the covariance matrix of the estimated

parameters. To prove this, we first consider a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}^*$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, so its joint probability density function (PDF) is given by

$$P(\boldsymbol{\theta}) = (2\pi)^{-\frac{N_{\boldsymbol{\theta}}}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\right], \tag{4.19}$$

where $N_{\boldsymbol{\theta}}$ is the dimension of the vector. Taking the natural logarithm of $P(\boldsymbol{\theta})$ gives:

$$l(\boldsymbol{\theta}) = \ln P(\boldsymbol{\theta}) = -\frac{N_{\boldsymbol{\theta}}}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^*), \tag{4.20}$$

the $\boldsymbol{\theta}$ dependence is now isolated into the final term, allowing us to evaluate the Hessian matrix as

$$\mathbf{H}(\boldsymbol{\theta}) = \left.\frac{\partial^2 \log l(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = -\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}. \tag{4.21}$$

According to CLT, Equation (4.19) is a generalised limit of $l(\theta)$; thus, it can be stated that $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \geq -\mathbf{H}(\boldsymbol{\theta})^{-1}$, meaning that the Hessian provides a lower limit on the covariance matrix that, provided high statistics, can be used to approximate the uncertainty in the MLE. Figure 4.4 shows the application of the MLE method to the toy model, where each plot shows the maximum likelihood $l(\theta)$ (blue dashed line) with an associated error band (red dashed line) for each free parameter. For this simple model, the MLE method was executed using a BFGS maximisation algorithm, which estimated each parameter to within one σ.

One of the challenges in using MLE for more complex models is ensuring convergence, especially when the likelihood surface has flat regions or multiple local maxima. Regularisation techniques or suitable constraints on parameter space are sometimes applied to improve the stability and convergence of the optimisation process. Furthermore, care must be taken in interpreting the covariance matrix derived from the inverse of the Fisher information matrix, as it provides insights into the uncertainties of the estimated parameters.

Bayesian inference

Bayesian inference, grounded in Bayes' theorem, offers an alternative approach to parameter estimation. Unlike MLE, which provides point estimates of parameters, Bayesian

inference generates a posterior distribution over the parameters, incorporating prior beliefs about the parameters and the observed data.

The posterior distribution $p(\theta|X)$ encapsulates all the information about the parameters after observing the data and allows for the computation of various statistics, such as the mean or credible intervals, which can be used to estimate the parameters. Bayesian methods are particularly useful when prior information is available or when dealing with complex models where MLE may be difficult to apply. For a set of parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2 \dots \theta_N\}$, given a set of observed data D and assuming that the prior distribution has no correlation between the parameters and is separable, Bayes' theorem can be written as

$$p(\boldsymbol{\theta}|D) = \frac{\mathcal{L}(D|\boldsymbol{\theta})}{\mathcal{Z}(D)} \prod_i \pi(\theta_i) = \frac{\mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}(D)}\,, \tag{4.22}$$

where $p(\boldsymbol{\theta}|D)$ is the posterior distribution, $\mathcal{L}(D|\boldsymbol{\theta})$ is the combined likelihood, $\mathcal{Z}(D)$ is the evidence, and $\pi(\theta_i)$ is the prior distribution representing initial beliefs about each component of $\boldsymbol{\theta}$. The posterior distribution, $p(\boldsymbol{\theta}|D)$, combines the likelihood of the observed data with the prior distribution, representing initial beliefs about $\boldsymbol{\theta}$. Bayes' theorem shows that the posterior is proportional to the product of the likelihood and the prior since the marginal likelihood $p(D)$ is independent of θ,

$$p(\boldsymbol{\theta}|D) \propto \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,. \tag{4.23}$$

This proportionality highlights the role of the prior distribution in shaping the posterior. In physics, prior distributions often represent previous experimental results or theoretical expectations. For example, the prior might incorporate existing knowledge from earlier experiments through well-defined distributions that may provide bounds or expectation values on expected measurements. The marginal likelihood $p(D)$, also known as the evidence $\mathcal{Z}(D)$, which serves as a normalising constant, is obtained by integrating the product of the likelihood and the prior over all possible values of θ

$$\mathcal{Z}(D) = \int \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}\,. \tag{4.24}$$

Calculating this integral analytically is difficult in practical applications, particularly for high-dimensional parameter spaces or complex models. In these cases, numerical methods such as Markov Chain Monte Carlo (MCMC) are often employed to approximate $\mathcal{Z}(D)$.

The parameter θ can be estimated by leveraging the proportionality relationship of equation (4.23), depending on the goal of the analysis. One common approach is maximum a posteriori (MAP) estimation, which involves finding the value of θ that maximises the posterior distribution. Mathematically, the MAP estimate is the solution to [120]

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid D)\,. \tag{4.25}$$

This method is closely related to maximum likelihood estimation. Still, it differs in that it incorporates the prior distribution, which can significantly affect the estimate, especially in cases where the data are sparse or noisy. Another widely used approach is to compute the posterior mean, which estimates $\boldsymbol{\theta}$ by taking the expected value of the posterior distribution. The posterior mean is given by

$$\mathrm{E}[\boldsymbol{\theta} \mid D] = \int \boldsymbol{\theta} p(\boldsymbol{\theta} \mid D)\, d\boldsymbol{\theta}\,. \tag{4.26}$$

This estimate is useful because it provides the average value of $\boldsymbol{\theta}$ given the data and the prior. The posterior mean often produces estimates that are more robust to outliers than the MAP estimate. It is also important to quantify the uncertainty in parameter estimates. This is commonly done through credible intervals [121], representing a range of values within which the parameter is believed to lie with a specified probability. For example, a 95% credible interval for θ is defined as:

$$P(\theta_{\text{lower}} \leq \theta \leq \theta_{\text{upper}} \mid D) = 0.95\,. \tag{4.27}$$

Unlike confidence intervals seen in frequentist statistics, credible intervals have a direct probabilistic interpretation: the true parameter value lies within the interval with a specified probability.

Applying Bayesian inference techniques to the toy model from Equation (4.13) first required selecting appropriate priors for the three parameters. A Gaussian prior was selected for the mean of the signal μ as it was reasonable to assume that an expectation could be obtained from theory or previous results. A half-normal distribution was chosen for the signal width σ to ensure a weakly informative, positive-definite prior [168]. For the signal fraction prior, the beta distribution was used as it is defined over the interval $[0, 1]$. For this situation, a numerical MCMC integration approach was used to evaluate the posterior mean (Equation (4.26)). Figure 4.5 shows the analysis results, which used

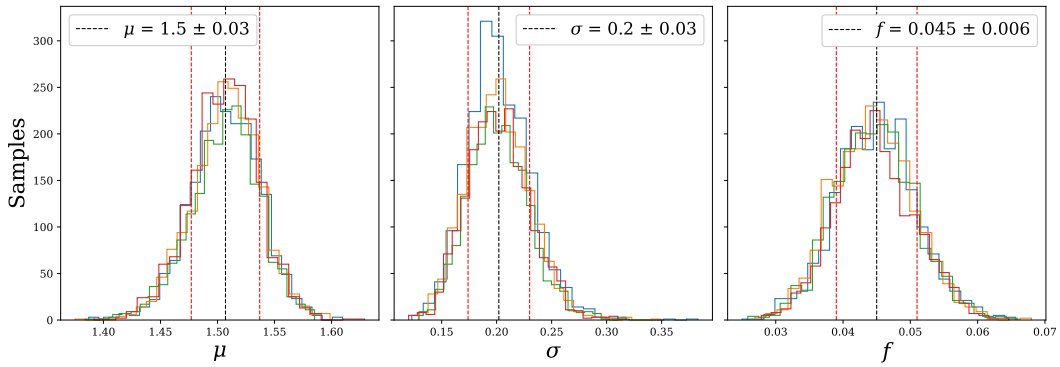


Figure 4.5.: Results from Bayesian inference used to fit the parameters of a toy model. The analysis used 40,000 samples split over four MCMC trials, producing four sets of posterior samples per parameter. Each parameter's final value and error are taken from the mean and standard deviation of the combined samples. When compared to MLE estimates, the Bayesian method provided a more robust error estimation but at the cost of computational complexity

40,000 samples split over four MCMC trials, producing four sets of posterior samples per parameter. Each parameter's final value and error are taken from the mean and standard deviation of the combined samples. The results are compatible with the MLE method. However, the errors are improved quantitatively and qualitatively by sampling the posterior distribution and producing a robust estimate of the uncertainty that does not rely on the approximation. The downside to this technique is the computational complexity due to the sampling requirements of the MCMC.

χ^2 minimisation

The final method of parameter estimation to compare is χ^2 minimisation. This technique follows from Section 4.3, where we used the χ^2 goodness of fit test to compare hypotheses. Implementing the χ^2 test statistic to parameter estimation effectively performs an MLE for a multivariate Gaussian likelihood. As such, the algorithm follows a similar gradient-based approach.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \chi^2(\theta)\,. \tag{4.28}$$

However, the implementation has one significant difference: the χ^2 method requires a direct comparison between observation and expectation. Thus, in this instance, the evaluation is made on a per-bin basis rather than per event. This raises an important

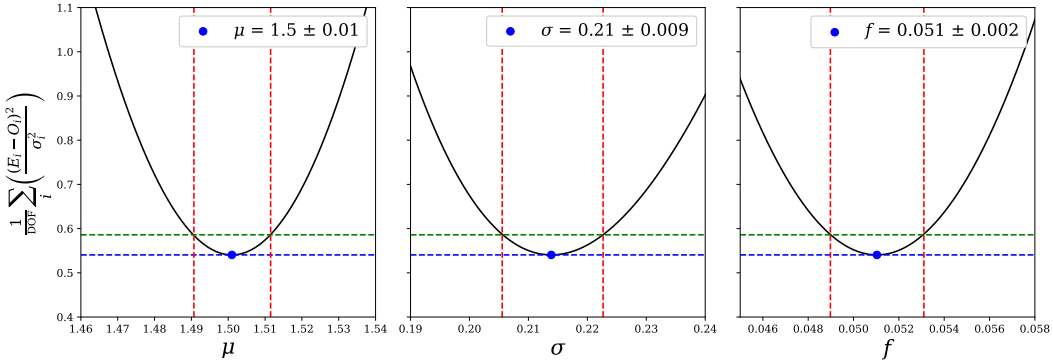


Figure 4.6.: χ^2 minimisation results for the three-parameter toy model where the best-fit parameter lies on the lower blue horizontal line. The parameter errors were calculated using the profile-likelihood approach [169], which calculates the 1-σ confidence interval at the $(\chi^2_{min}+1)/\text{DOF}$ intersection. The intersection corresponds to the upper green and vertical red dashed lines.

consideration, the number of bins, or equivalently, the bin-width, as it affects the trade-off between bias and variance; decreasing the bin-width reduces bias, allowing for a more detailed data representation. In other words, histograms with narrower bins can better approximate the underlying distribution. However, this also increases variance, as fewer data points are available to estimate the height of each bin. Consequently, histograms with narrower bins become more sensitive to random fluctuations in the data, leading to increased variability across datasets sampled from the same population.

In collider physics, data is typically presented as counts per bin, significantly reducing the complexity of describing a dataset. This simplification works well when the expected counts per bin are sufficiently high since statistical fluctuations scale with the square root of the expected number of events. However, issues arise when statistics are low and the number of events in a given bin approaches zero. This is particularly problematic when searching for BSM signals, which may manifest as only a handful of events through rare processes. While this challenge will be explored in later chapters, for our purposes, we assume that the event statistics are high enough to justify analysis on a per-bin basis.

Figure 4.6 shows the χ^2-minimisation results for the three-parameter toy model. The parameter errors were calculated using the profile-likelihood approach [169], which examines how χ^2 changes when one parameter is varied while others are fixed at their best-fit values. For a good fit, χ^2 increases by 1 when the parameter is varied by one standard deviation. The value of θ_i where χ^2 increases by one from its minimum is a good approximation of the one-sigma confidence interval for θ_i. The assumption of

	$\mu = 1.5$	$\sigma = 0.2$	$f = 0.05$
MLE	1.5 ± 0.1	0.19 ± 0.03	0.051 ± 0.004
Bayesian inference	1.5 ± 0.03	0.2 ± 0.03	0.045 ± 0.006
χ^2 minimisation	1.5 ± 0.1	0.21 ± 0.009	0.051 ± 0.002

Table 4.1.: Comparison of results of parameter estimation using three distinct methods to estimate three free parameters of a toy model

a symmetric error relies on the observation that $\chi^2(\theta)$ is approximately quadratic at its minimum, where the function is dominated by the second-order term of its Taylor expansion. Much like the estimation of uncertainty in the MLE method, this method offers an approximation of 1σ bound. There are instances where this approximation may fail; a good example would be for bounded or discrete parameters that constrain the minima breaking the quadratic statement.

The y-axis in Figure 4.6 is scaled in terms of the degrees of freedom $\text{DOF} = n - p - 1$, where n is the number of data points and p is the number of fitted parameters. This reduction was briefly discussed in Section 4.3, where the proximity to unity was used to assess the goodness of fit intuitively. In parameter estimation, the reduced chi-squared, χ^2_{red}, serves as an indicator of the fit quality. A χ^2_{red} value significantly greater than one suggests that the model does not adequately describe the data, which may indicate either a poor model or inconsistencies in the data sample. Conversely, a χ^2_{red} value much smaller than one can signal overfitting, often due to overly conservative error estimates. For example, if the variance in each bin is assumed to be Poisson, such that $\sigma_i^2 = E_i$, but the errors are artificially inflated so that $\sigma_i > \sqrt{E_i}$, the term $(E_i - O_i)^2/\sigma_i^2$ will frequently be less than one, leading to an underestimated χ^2_{red}.

Table 4.1 shows the results from all three methods. The χ^2 minimisation results are within the approximation of one sigma from the known value and are compatible with the MLE and Bayesian inference techniques in terms of accuracy and uncertainty. The toy model used in this example was a simple combination of two distinct processes; actual experimental data and theoretical models are rarely so simple. However, the purpose of a toy model is to test the application and compare the results of different methods, ensuring consistent and well-understood results. By starting with idealised conditions, one can refine techniques and identify potential problems in a controlled environment. In later chapters, toy models will be used to understand complex analysis chains and unknown emergent distributions.

Chapter 5.

Reinterpretation

In particle physics, data reinterpretation plays a crucial role in maximising the scientific return from LHC experiments. By applying different theoretical models and hypotheses to existing data, researchers can extract new insights without needing to perform additional experiments. This approach is especially useful when exploring the vast parameter spaces of "beyond the SM" (BSM) theories, where direct searches might not be feasible. The reinterpretation of LHC data requires dedicated tools and methods, including Monte Carlo event generators and specialised analysis frameworks, to simulate, analyse, and compare theoretical predictions with experimental observations.

Monte Carlo event generators, including MadGraph5 [170], Sherpa [171], and Pythia 8 [172], play a crucial role in generating theoretical predictions for particle collisions at the Large Hadron Collider (LHC). These generators effectively simulate various components of the collision process, encompassing hard scattering, parton showering, hadronization, and the underlying event, thereby offering a comprehensive understanding of the final state particles detected.

The simulated events generated by these tools are crucial for designing and interpreting LHC searches. To reinterpret existing experimental results under different theoretical frameworks, researchers utilize analysis tools like MadAnalysis 5 [173], Rivet [174], Contur [175] and SModelS [76].

For example, researchers can simulate theoretical models with Monte Carlo event generators, apply experimental selections using MadAnalysis 5 or Rivet, and then reinterpret the results using tools like SModelS. This workflow allows for the exploration of a broad spectrum of new physics scenarios, providing critical tests of the Standard Model's robustness and identifying potential signals of new phenomena. Importantly,

data reinterpretation efforts benefit from the ongoing development of open-source software and the collaborative nature of the high-energy physics community, which ensures that the tools and methodologies remain state-of-the-art and accessible.

As LHC experiments continue to collect data, the role of data reinterpretation will only grow in significance. With the ongoing improvements in event generators and analysis tools, the community will be better equipped to explore the limits of our understanding of particle physics. This approach not only enhances the scientific value of existing datasets but also guides future experimental efforts by identifying promising areas for further investigation.

5.1. Reinterpretation tools

Reinterpretation has been greatly facilitated by the development of dedicated tools such as Rivet, MadAnalysis 5, and SModelS, each offering unique features and methodologies that contribute to the comprehensive analysis of collider data. This section will look at these tools individually and highlight their distinct capabilities, underlying procedures, and the specific roles they play in the reinterpretation of experimental results.

5.1.1. Rivet

Rivet [174] (Robust Independent Validation of Experiment and Theory) is a software tool used to validate and compare theoretical predictions from particle-physics models against experimental data. It provides a framework for performing these comparisons in a consistent and reproducible manner. Recent updates to Rivet, particularly since version 4 [176], have introduced modifications intended to accommodate evolving research needs, such as extending the range and precision of comparisons and enabling the development of new analysis techniques.

The primary function of Rivet is to facilitate comparisons between theoretical models, such as those generated by Monte Carlo event generators and experimental data. It provides a standardised interface for assessing simulated events against measurements, incorporating options for detector simulation and event smearing. The workflow is structured around explicit runs of simulated collider events, produced using external event generator tools and recorded in the HepMC 3 event format [177]. The software

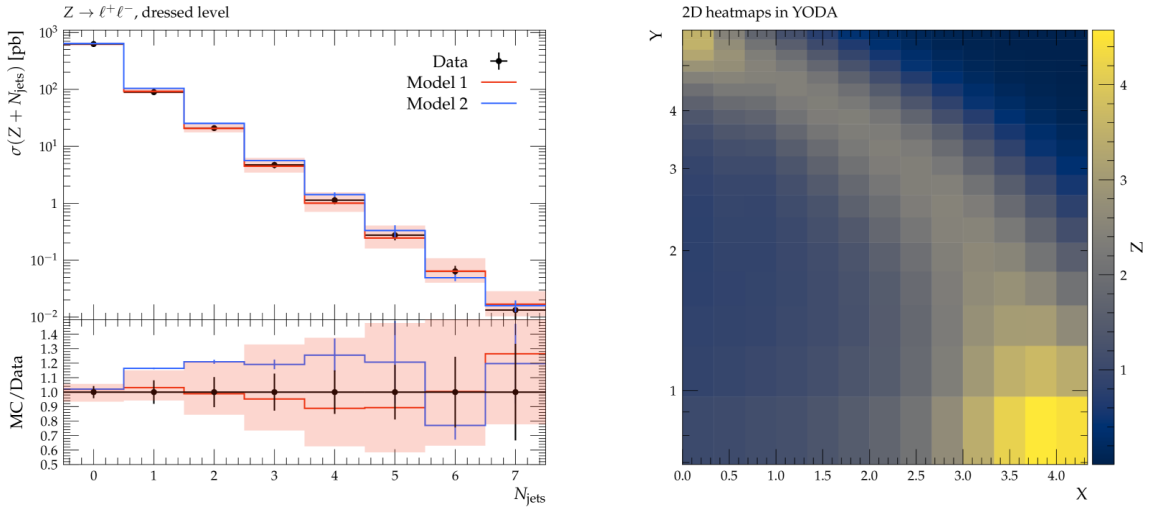
consists of a computational core designed for consistent processing of experimental data, along with a configurable system for smearing-based detector emulation and a library of over 1,500 preserved analyses.

RIVET is integrated with the HepData [178] database, which archives published data from high-energy physics experiments. This integration allows researchers to access experimental results directly, facilitating the validation of theoretical models. The modular structure of RIVET supports applications in beyond the Standard Model (BSM) scenarios. For instance, CONTUR ("Constraints On New Theories Using Rivet" [175]) utilises experimental measurements available through RIVET to investigate BSM theories. This approach takes advantage of the model independence of particle-level differential measurements within fiducial regions of scattering phase space, enabling systematic comparisons with BSM scenarios simulated using Monte Carlo generators.

Version 4 introduced several new features, including more flexible histogram handling, aligned with updates to the YODA 2 [179] statistical data-analysis library. Additional capabilities include support for storing and loading analysis-specific data in HDF5 format and the ability to import machine learning (ML) models, particularly deep neural networks and graph neural networks, from ONNX serialisation files [180].

A central aspect of RIVET 's development is the provision of reliable, reproducible comparisons between theory and experiment. Version 4 facilitates this by ensuring consistent handling of experimental measurements and supporting independent validation through cross-checking methods. Given the diversity of theoretical models employed in collider physics, this validation process is an important component of the analysis workflow.

YODA serves as the backend for managing histogram objects within RIVET [179] and providing the main output data format. YODA provides a system for handling these objects with improved memory management and scalability, including dynamic binning and multi-weight options. The framework is integrated with RIVET to ensure that Monte Carlo-generated data can be compared with experimental measurements in a reproducible manner. Through YODA, RIVET stores analysis results in a standardised format, facilitating data visualisation and comparison. The latest version, YODA 2, includes support for generating PYTHON visualisation scripts using the MATPLOTLIB package [181]. In collaboration with the YODA team, contributions were made to visualisation updates, particularly in the development of 2D histogram outputs. Figure 5.1 presents examples of 1D and 2D plot outputs from the YODA 2 plotting interface.



(a) **(b)**

Figure 5.1.: Example 1D and 2D plot outputs from the YODA 2 plotting interface, illustrating a mix of uniform and irregular bin sizes and automatic ratio plotting for 1D data/prediction comparisons. Plots were taken from [179].

The release of RIVET 4 expanded the number of available analyses to over 1,500, incorporating recent experimental data. Furthermore, improvements to the underlying framework have been implemented to enhance performance and scalability, allowing for larger datasets and more complex analyses while maintaining efficiency. These updates align with the increasing data volumes generated in high-energy physics experiments, ensuring that RIVET remains a useful tool for theoretical model validation.

5.1.2. MadAnalysis 5

MADANALYSIS 5 [173] is a comprehensive framework that integrates a wide range of functionalities to analyse both simulated and experimental data. Similar to RIVET, MADANALYSIS 5 enables users to perform highly flexible event-level analyses. MADANALYSIS 5 offers a platform for beginner and expert users to conduct physics reinterpretations and analyses using event files, such as those generated by a large class of Monte Carlo event generators. The tool is optimised for phenomenological investigations, making it easy to analyse, simulate, and reconstruct events. The framework provides two operating modes: one tailored for simplicity (a beginner-friendly mode) and another for experts needing advanced control over their analyses.

MadAnalysis 5 includes support for multiple event formats (parton-level, hadron-level, and reconstructed-level events) and the ability to handle complex simulations such as parton showering, hadronization, and fragmentation. It integrates with Monte Carlo generators like Pythia 8 [172] and Herwig [182] to simulate particle interactions and supports fast detector simulation using external programs like Delphes 3 [183], allowing researchers to test the response of a collider detector to the particle. While full detector simulations used by large experimental collaborations are time-consuming and complex, fast simulations offer a streamlined approach, providing realistic estimates of particle interactions, smearing effects, and detector resolutions. This capability enables users to quickly evaluate the feasibility of analyses, study signatures of new physics, and estimate backgrounds without the computational overhead of full-scale simulations. It is particularly beneficial for early-stage phenomenological studies or when testing numerous models.

MadAnalysis 5 also supports a user-friendly interface for constructing analysis routines through a command-line interface or by writing Python scripts, offering a lower entry barrier for new users. Furthermore, the tool incorporates advanced statistical analysis capabilities, allowing users to compute exclusion limits, p-values, and confidence intervals based on the observed data. This makes MadAnalysis 5 particularly suited for reinterpreting experimental results in the context of various BSM scenarios, as it provides a streamlined workflow from event generation to statistical interpretation.

5.1.3. SModelS

SModelS provides a framework for the reinterpretation of beyond the Standard Model (BSM) scenarios by decomposing them into simplified model spectra (SMS) [76]. This approach enables a systematic comparison of theoretical predictions with existing experimental constraints from the Large Hadron Collider (LHC), facilitating the identification of viable parameter spaces for new physics.

The methodology employed in SModelS involves the decomposition of supersymmetric (susy) theories into SMS, which correspond to distinct experimental signatures (see Section 2.3.3). These SMS are then compared against LHC data to assess their consistency or exclusion. The automation of this decomposition and comparison process allows for an efficient evaluation of a broad class of BSM models without necessitating a full phenomenological simulation of each scenario. This approach is particularly useful when the complexity of a model makes direct comparison with experimental searches

challenging. By focusing on SMS components, SModelS provides a structured reinterpretation of experimental results, indicating which aspects of a model are subject to constraints from current data.

The second version of SModelS introduced significant modifications aimed at enhancing its applicability to a wider range of BSM scenarios [81]. Among these modifications was the extension of its capabilities to accommodate more complex topologies beyond the minimal SMS approach, thereby expanding the theoretical search space. Additionally, the version incorporated an enlarged database of experimental results from ATLAS and CMS, including over 100 new searches for BSM phenomena spanning various topologies and final states. These results impose constraints on different new physics scenarios by setting exclusion limits based on observed data, thereby refining the viable parameter space of theoretical models.

The most recent iteration, SModelS v3 [184], introduced further developments to the framework. Earlier versions primarily focused on supersymmetric models with $\mathcal{Z}_2$ symmetries, which enforce specific particle decay patterns. These topologies are relevant in SUSY models as they characterize stable particles and decay chains leading to the lightest supersymmetric particle (LSP), which is often considered a dark matter candidate. However, the assumption of $\mathcal{Z}_2$ symmetry imposes limitations when analyzing more general BSM scenarios. SModelS v3 extends beyond these constraints by incorporating topologies with multi-step decay chains, metastable intermediate states, and non-minimal final states. These features align with the complexity of realistic BSM models that cannot be reduced to simple two-state $\mathcal{Z}_2$ structures. The extended framework enables an analysis of scenarios where particles undergo multiple intermediate decays, thereby broadening the range of possible new physics signals that can be tested against LHC data. SModelS v3 also includes an expanded database of experimental constraints from recent LHC runs, refining the exclusion limits applied to BSM models. Additionally, improvements in computational efficiency facilitate the analysis of more intricate scenarios, enabling broader model scans.

A technical development in SModelS v3 is the implementation of graph-based analysis techniques to address the increasing complexity of particle decay chains and topologies. Within this framework, decay processes are represented as directed graphs, where vertices correspond to particles and edges denote decay transitions. Figure 5.2 illustrates a simplified model topology using this representation. The graph-based formalism provides a systematic approach for encoding and analyzing complex decay chains.

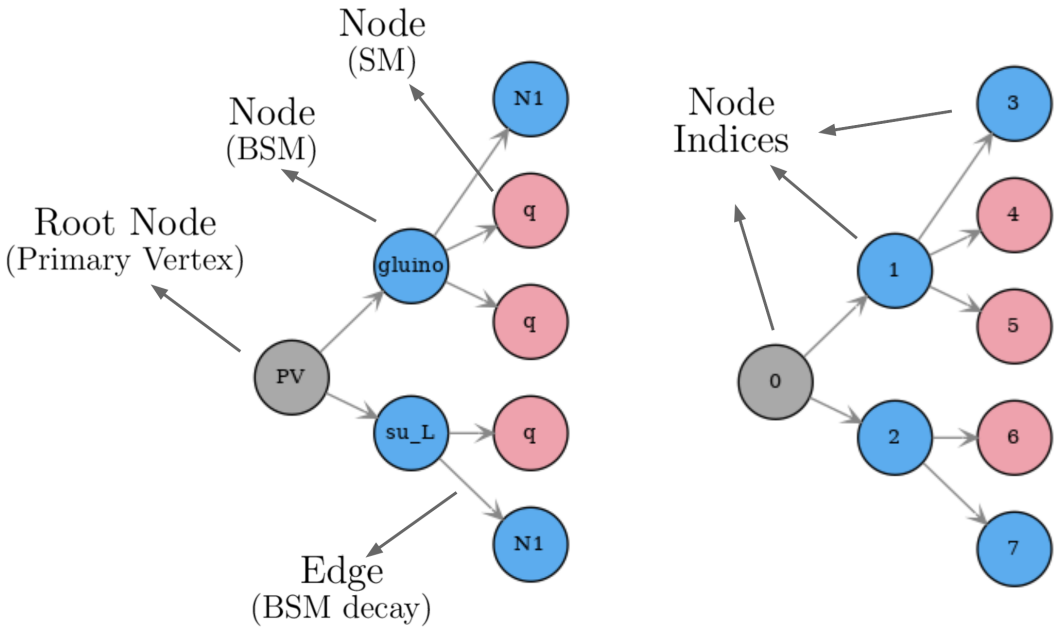


Figure 5.2.: Graph representation of a simplified model topology and its elements: root node, SM and BSM nodes, edges, and node indices. Plot taken from the paper "SModelS v3: Going Beyond $\mathcal{Z}_2$ Topologies" [184].

By utilizing graph structures, SModelS v3 facilitates an efficient decomposition and analysis of intricate topologies, including those with loops or branching points in decay chains. Each topology is treated as a graph, allowing the tool to explore a diverse set of BSM signatures beyond simple tree-like decay patterns. Graph traversal algorithms enable the identification of possible decay paths, which are then compared against LHC data. This approach generalizes the decomposition of models, enabling the detection and analysis of new physics scenarios that involve complex final states.

With these extensions, SModelS v3 provides a structured framework for testing a broad range of BSM models. The inclusion of more general decay topologies and an expanded database of experimental constraints enhances its capability to assess theoretical scenarios against LHC data, offering a systematic approach for confronting new physics models with experimental observations.

5.2. Limitations of Current Reinterpretation Tools

Tools such as Rivet, MadAnalysis 5, and SModelS have facilitated the reinterpretation of experimental results in particle physics. However, several limitations remain that

affect their applicability and precision. A key challenge is the inherent model dependence of these tools. The results of reinterpretation studies are often sensitive to the specific theoretical assumptions and input parameters, such as the choice of parton distribution functions (PDFs) and the treatment of systematic uncertainties. This is particularly relevant in the context of the subsequent chapters, where the absence of uncertainty propagation to BSM reinterpretation statistics is considered. Mitigation strategies include event re-weighting or re-sampling of the PDFs, though these approaches do not eliminate all sources of bias.

Another limitation concerns the coverage of phase space. While Rivet and MadAnalysis 5 provide extensive libraries of analysis routines, these are generally derived from published experimental analyses, which may not encompass all possible final states or kinematic regimes. Consequently, certain new physics scenarios, particularly those involving rare or unconventional signatures, may not be adequately tested. Similarly, SModelS, which rely on simplified model spectra (SMS) for reinterpretation, may not fully capture the complexity of more intricate BSM scenarios, potentially neglecting important correlations or decay chains.

Computational complexity presents an additional constraint, particularly in high-dimensional parameter spaces or when combining multiple datasets. Large-scale Monte Carlo simulations and detailed statistical analyses can be computationally expensive, limiting the feasibility of certain studies. Furthermore, the accuracy of reinterpretation outcomes is influenced by approximations in detector simulation and event reconstruction, which may not fully account for the complexities of real experimental conditions.

Finally, the reliance on published data remains a significant challenge, as data may not always be available in a format suitable for reinterpretation. Although initiatives have sought to enhance data preservation and accessibility (see [178, 180]), gaps persist, particularly for older experiments or results lacking sufficient documentation. This limitation constrains the scope of possible reinterpretations and introduces biases based on the selection of available analyses.

In summary, while reinterpretation tools such as Rivet, MadAnalysis 5, and SModelS provide valuable frameworks for testing new physics models against existing data, limitations related to model dependence, phase-space coverage, computational demands, detector simulation accuracy, and data availability necessitate continued efforts to enhance their flexibility, accuracy, and accessibility within the particle physics community.

5.3. Combining results

The ATLAS and CMS experiments at the LHC are designed to probe the Standard Model (SM) through precise measurements and searches for phenomena beyond established theoretical frameworks. These searches typically involve testing multiple hypotheses over a broad parameter space, with each experiment reporting its findings independently. To fully exploit the statistical power of both experiments, it is often necessary to combine their results. However, this process presents a number of challenges arising from differences in experimental methodologies, systematic uncertainties, and the statistical treatment of correlated and uncorrelated uncertainties. This section examines the challenges associated with combining results from ATLAS and CMS, with a focus on detector sensitivities, systematic uncertainties, and their correlations.

Detector sensitivities, systematic uncertainties, and correlations

One of the main challenges in combining results from ATLAS and CMS stems from differences in detector design and operational parameters. Despite both being general-purpose detectors, they differ in magnet configurations, calorimeter technologies, and tracking systems, leading to variations in energy resolution, particle identification, and event reconstruction. These differences introduce experiment-specific systematic uncertainties that must be carefully treated when performing combined analyses.

The ATLAS and CMS collaborations conduct analyses independently, each using distinct calibration procedures, data reconstruction algorithms, and statistical methodologies. This independence mitigates common systematic biases, lending robustness to results when both experiments observe the same physical phenomenon. However, while the results are largely independent, they are not entirely uncorrelated. Some systematic uncertainties, such as those arising from detector calibration, particle identification efficiencies, and background estimation procedures, have experiment-specific components, while others, such as luminosity uncertainties, are partially correlated between the two experiments.

Formally, let θ_i represent a nuisance parameter associated with a systematic uncertainty that is correlated between ATLAS and CMS. The joint likelihood function for a parameter of interest m_X, $\mathcal{L}_{\text{joint}}(m_X, \theta_i)$, can be expressed as the product $\mathcal{L}_{\text{ATLAS}}(m_X, \theta_i) \cdot \mathcal{L}_{\text{CMS}}(m_X, \theta_i)$. In this expression, consistent treatment of θ_i across both likelihoods is required, necessitating detailed communication and data sharing between

the collaborations. Additionally, external inputs, such as theoretical uncertainties on cross-section calculations, contribute to systematic uncertainties common to both experiments. The propagation of these uncertainties in combined analyses often relies on advanced statistical techniques, including Markov Chain Monte Carlo (MCMC) methods or numerical integration approaches.

Although the distinct designs and methodologies of ATLAS and CMS reduce direct correlations, shared theoretical models and environmental conditions can introduce dependencies. Therefore, while the results of these experiments can be considered largely independent, they are not completely uncorrelated, and careful treatment of systematic uncertainties is required when performing combined analyses.

Theoretical Model Dependencies

Theoretical uncertainties, such as those associated with parton distribution functions (PDFs) and higher-order QCD corrections, introduce additional challenges when combining search results from different experiments. These uncertainties influence both the predicted signal and background rates and are often correlated across different analyses. A consistent combination of results necessitates the systematic treatment of such theoretical uncertainties, which are typically obtained from external groups or derived from global fits to experimental data.

Discrepancies may arise when combining results obtained under different theoretical assumptions. For instance, one experiment may employ next-to-leading-order (NLO) QCD calculations for a signal process, while another may use leading-order (LO) predictions. To ensure consistency, either a common theoretical framework must be adopted, or the discrepancies must be accounted for through additional uncertainty terms.

Dependence on specific theoretical models further complicates the combination of results across experiments. LHC searches are often interpreted within particular frameworks, such as the Standard Model (SM) or beyond the Standard Model (BSM) scenarios, including supersymmetry (SUSY) or models with extra dimensions. The optimisation of search strategies may vary between experiments due to differing assumptions about model parameters, such as the mass spectrum of SUSY particles or the coupling strengths in extra-dimensional models.

Consequently, combining results requires careful consideration of the treatment of theoretical uncertainties and model parameters. If one experiment is primarily sensitive to

low-mass SUSY particles while another is optimised for high-mass regions, the combined result must appropriately incorporate these different sensitivities. This issue becomes particularly relevant in models with a large number of free parameters, where the coverage of parameter space by individual experiments may be non-uniform.

Signal Model Dependence

In searches for new particles or interactions, the signal model defines the expected manifestation of a hypothesised process in experimental data. This model is typically parameterised by quantities such as particle masses, coupling constants, and decay channels. Differences in these assumptions between experiments can lead to variations in reported exclusion limits and significance levels.

A consistent combination of results requires that both experiments test the same hypothesis. This can be challenging if different signal models or theoretical frameworks are used in the interpretations. In SUSY searches, for example, exclusion limits may depend on the choice of a specific SUSY-breaking scenario. A meaningful combination of results necessitates the application of a consistent theoretical scenario across datasets.

One approach to addressing this issue is the reinterpretation of one experiment's results within the signal model used by the other. This requires access to detailed information on signal acceptance and efficiency, which may not always be available. Alternatively, a more model-independent combination can be pursued by focusing on simplified models or benchmark scenarios common to both collaborations. In either case, residual differences arising from model dependence must be systematically evaluated and incorporated into the final combination.

Differences in event selection criteria and the definition of signal regions present additional challenges. Each experiment applies distinct selection criteria to enhance sensitivity while suppressing backgrounds, but these selections are not always directly comparable. For instance, variations may exist in the transverse-momentum thresholds for jets or leptons, as well as in isolation criteria for electrons and muons. Addressing these discrepancies requires either re-analysis with standardised selection criteria or the application of efficiency and acceptance corrections, both of which introduce additional uncertainties. Furthermore, the treatment of background processes, such as multijet production or top-quark pair production, may differ between experiments, as these backgrounds are often estimated using different methodologies and control regions.

Beyond the combination of results, the interpretation of combined data in the context of global fits introduces further complexities. Global fits aim to determine the best-fit values of model parameters by incorporating data from multiple experiments and decay channels. This process requires careful treatment of correlations between measurements and the associated systematic uncertainties to ensure a consistent and robust interpretation of the results.

Data combination techniques

The process of data combination introduces a range of complexities. A common approach involves constructing a combined likelihood function that incorporates the likelihoods from multiple experiments while accounting for both correlated and uncorrelated systematic uncertainties. This method requires careful formulation to ensure compatibility across different datasets, which can be non-trivial given the variations in detector performance, event selection criteria, and background modelling.

An alternative method employs simplified template cross sections (STXS) [185] or signal strength modifiers ($\hat{\mu}$), facilitating a standardised representation of results across multiple experiments and decay channels. The use of STXS aims to minimise theoretical uncertainties directly embedded in the measurements while allowing for a consistent combination of results across different decay channels and experimental setups.

The signal strength modifier $\hat{\mu}$, defined as $\sigma_{\text{obs}}/\sigma_{\text{SM}}$, provides a normalised measure relative to the Standard Model (SM) expectation, aiding in the comparison and combination of results across production modes and decay channels (e.g., $H \to \gamma\gamma$, $H \to ZZ$). This normalisation is particularly relevant given the distinct systematic uncertainties, backgrounds, and signal efficiencies associated with each channel. However, differences in the treatment of systematic uncertainties and signal modelling across experiments may introduce additional sources of uncertainty. Furthermore, the applicability of STXS or $\hat{\mu}$ values relies on the assumption that the theoretical model remains valid across the full parameter space of interest, which may not always be the case.

Practical aspects of combination

In addition to statistical and theoretical considerations, practical challenges arise in the combination of ATLAS and CMS search results. Differences in data formats, analysis frameworks, and software tools can complicate this process. While efforts have been made

to establish standardised procedures, residual discrepancies can impede the efficiency of result combination.

A fundamental aspect of the combination process is the validation of the combined results. This involves ensuring consistency with individual ATLAS and CMS results and verifying that all sources of uncertainty are appropriately accounted for. Tools such as RooStats [186] and PYHF [187] (the PYTHON implementation of HistFactory [188]) provide functionality for constructing probability models and digitally publishing analysis results. These tools are widely adopted within the high-energy physics community and contribute to the validation process. However, given the complexity of these analyses and the large number of systematic uncertainties involved, additional cross-checks and robustness studies are often required.

The scale of LHC data and the complexity of the associated analyses present significant computational challenges. The combination of experimental results frequently necessitates access to substantial computing resources, particularly when handling high-dimensional likelihood functions or conducting global fits. The computational demands are further increased by the need for extensive systematic uncertainty studies, which often involve large numbers of pseudo-experiments or toy Monte Carlo simulations.

Moreover, coordination between different experimental collaborations is a complex task that requires significant organisational effort. Effective coordination is necessary to ensure that results are combined in a scientifically rigorous manner, with proper treatment of relevant uncertainties and correlations. Cross-experiment communication plays a crucial role in this process, as different experiments may have distinct priorities, timelines, and resource constraints.

The combination of search results from LHC experiments requires consideration of multiple factors, including differences in detector sensitivities, statistical methodologies, theoretical model dependencies, and event selection criteria. The process itself introduces additional challenges, particularly in the construction of combined likelihood functions and the interpretation of results within global fits. These challenges are further compounded by computational and resource requirements. Despite these complexities, combining results from different LHC experiments remains an essential task, enabling more comprehensive tests of the Standard Model and potential deviations from it.

Chapter 6.

Optimal Combinations

Combining ATLAS and CMS analyses poses multiple challenges. However, as is often the case in physics, the problem can be broken down into smaller parts and considered one step at a time. This chapter will consider the combinatorial challenge of selecting the optimum set of analyses and/or signal regions–each defined by different kinematic cuts, particle identification requirements, and background suppression techniques–for a given BSM model. One of the key insights to this problem was the realisation that the process is analogous to feature-selection in machine learning, particularly in handling issues of redundancy, correlation, and optimality in high-dimensional data spaces. In both cases, the objective is to select an optimal subset from a large pool of potential features or regions, where choosing poorly can introduce bias, lead to overfitting, or result in the double-counting of information. In machine learning, feature-selection aims to identify minimally correlated features that best represent the dataset without introducing redundancy. Similarly, analyses and/or signal regions must be selected to maximise the sensitivity of searches for new physics while avoiding those that record the same underlying physical processes. This parallels the feature-selection process, where including highly correlated features leads to biased models, just as selecting overlapping signal regions can inflate statistical significance and introduce bias into the results.

A key similarity lies in the combinatorial nature of both problems. In feature-selection, the search for an optimal subset of features from a large feature space grows, as will be shown, as 2^n with the number of features (n), making brute-force searches impractical as n increases. This is analogous to the situation in high-energy physics, where the number of possible signal regions grows exponentially with the dimensionality of the phase space. The challenge is compounded by the need to account for correlations between regions, as multiple regions may be sensitive to or contain the same underlying physical process.

Using combinatorial optimisation techniques, such as genetic algorithms [189, 190] or simulated annealing [191], explores the space of possible feature sets or signal regions while avoiding local optima that might arise from greedy search strategies. However, such searches are computationally expensive and often struggle with many features.

Another parallel is the treatment of correlations between selected features or regions. In machine learning, pairwise correlations between features are often measured to prevent the selection of redundant features. Similarly, in ATLAS and CMS analyses, the combination of signal regions must carefully consider correlations between the events in each region. This is typically done using statistical techniques like profile-likelihood fits, accounting for the uncertainties and correlations between signal regions. When such correlation information is available, these methods ensure that the combination of signal regions maximises the statistical power of the analysis while preventing the over-counting of events common to multiple regions.

The first simplifying assumption we can make is that it is possible to make the statement that the two analyses can be combined. In many cases, this is already the case. For instance, if the combination process requires the construction of a test statistic based on event yields, then correlations would be arguably negligible if results from different experimental locations or ones performed at different energies were combined. So, assuming correlations can be estimated, the question becomes how to choose a minimally correlated set from hundreds or thousands of options.

6.1. Compatible sets as path-finding

Without prior information about overlaps, correlations or statistical significances, finding a preferred subset of data features l from a set of size n is a combinatorial challenge with 2^n possible solutions. Specifically, the total number N of subsets with $r > 1$ elements is

$$N = \sum_{r=2}^{n} \binom{n}{r} \equiv \sum_{r=2}^{n} \frac{n!}{r!(n-r)!} = 2^n - (n+1)\,. \tag{6.1}$$

Exhaustively generating and evaluating each potential combination exhibits exponential time complexity, making it computationally infeasible even for relatively small values of n, especially for the $n \sim \mathcal{O}(1000)$ required in practical applications. However, we can reduce this number by acknowledging that there will exist *forbidden* pair-wise combinations due to excessive correlation or shared information–a concise description of this metric is

required; thus, this study will refer to the pair-wise relation as *overlap*–We can assume that a binary pair-wise matrix exists that determines whether or not any given pair of features $\{l_i, l_j\}$ can be combined such that B_{ij} is either true or false; this is defined as the binary acceptance matrix (BAM). When accounting for exclusivities, B_{ij}, over an entire set, most of these N combinations will be *forbidden*. This is because, for large r, it becomes increasingly likely that naïve subsets will include at least one pair of overlapping features. Given prior knowledge of B_{ij}, the pertinent question is whether it is possible to evaluate all *allowed* combinations more efficiently than through exhaustive generation followed by overlap checking.

This section demonstrates that the answer to this question is yes and that the solution can be considered the optimal path within a directed acyclic graph (DAG). I will introduce a new algorithm that improves the asymptotic time complexity by efficiently selecting path elements through the recursive application of the BAM. This approach reduces the combinatorial nature of the problem, making it computationally feasible while significantly enhancing efficiency.

The critical optimisation is to avoid generating invalid combinations in the first place; this can be achieved by evaluating the combinations directly from the BAM. Hence, we must restrict the generated subsets to those with $B_{ij} = 1$ for all the set's distinct elements i, j. This condition requires that if a subset of all possible features is built up iteratively, its jth element must have no significant overlap with all the previously selected elements $0...j-1$.

Figure 6.1 illustrates a specific instance of a binary adjacency matrix (B) consisting of ten features. To create this example, a symmetric matrix of random values with dimensions 10×10 was generated and subsequently transformed into the BAM by applying a threshold of $T = 0.5$. The matrix ρ contains random numbers where $\rho_{ij} \sim U(0, 1)$; further discussion will demonstrate alternative methods for constructing this matrix, which include using correlation metrics. In this example, the elements of ρ_{ij} that are below the specified threshold are deemed combinable and are depicted in white, whereas those that exceed the threshold are shaded in black. For reasons that will be covered in more detail later in this chapter, the construction of combinations is limited to subsets where features are added in a strictly increasing index order. Beginning at the top left corner of Figure 6.1, corresponding to the element ρ_{00} (or (l_0, l_0)), the features available for combination with l_0 are constrained to those associated with white elements in the first column, i.e., $B_{i0} = 1$. We define A_i as the ordered set of all permissible, minimally

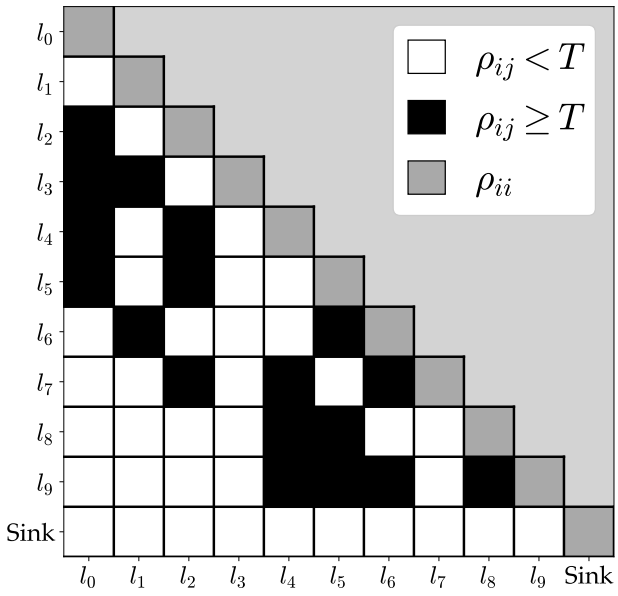


Figure 6.1.: Binary acceptance matrix of 10 data features ($l_0 - l_9$) with values masked according to some threshold T e.g. correlation or mutual-information below a certain value. The final *Sink* index has been inserted to provide a target for the path-finding algorithm.

overlapping feature indices relative to element l_i, such that

$$A_i \equiv \{j : \rho_{ij} < T,\ i < j < n\}\,. \tag{6.2}$$

Using Figure 6.1 as an example, A_0 would be $\{1, 7, 8, 9\}$. We now define a set of indexed variables to represent the sets and subsets of features: the single-index version, K_i, is a set of all allowed paths with initial elements l_i such that

$$K_i \equiv \{\{l_i, \ldots, l_{\text{final}}\}, \ldots\}\,. \tag{6.3}$$

Following this construction, $K_{i,j}$ is the jth path within K_i, and by extension $K_{i,j,k}$ would refer to the kth element of $K_{i,j}$. Applying this formalism to Figure 6.1 and initiating a subset $K_{0,0,0}$ with l_0, the available options for the second element are given by indices in A_0 (Equation 6.2). It follows that $K_{0,0,1} = l_1$ as this is the first index in A_0, and thus $K_{0,0,2} = l_7$ as this is the first available index that is allowed by the intersection of A_0 and

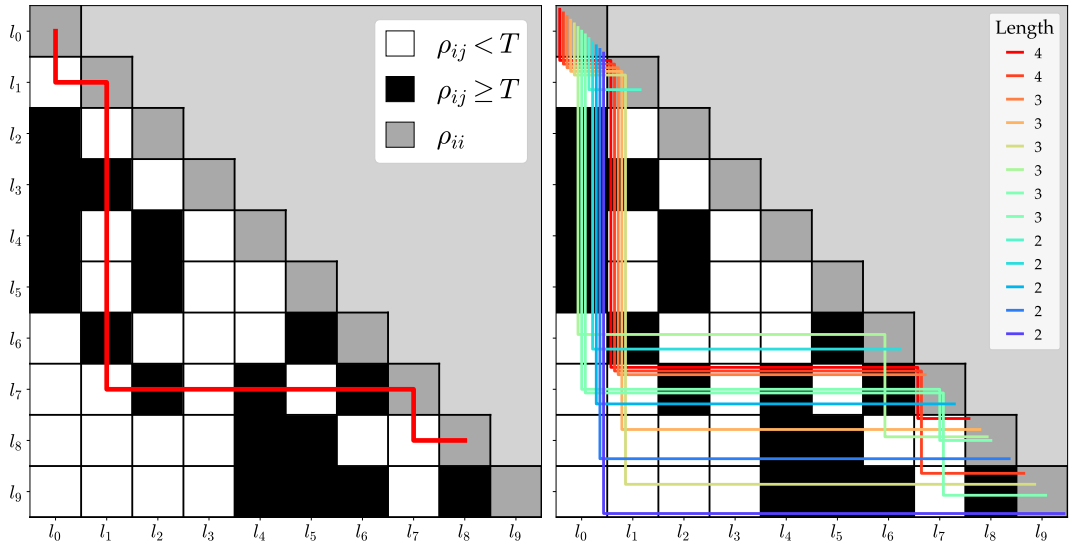


Figure 6.2.: Binary acceptance matrix of 10 data features (l_0−l_9) with values masked according to some threshold T e.g. correlation or mutual-information below a certain value. The left-hand plot is the first allowed (legal) path evaluated from l_0 following the HDFS algorithm. The right-hand plot shows all available (legal) paths originating at l_0.

A_1. Repeating the procedure and taking the intersection of A_0, A_1 and A_7 gives $\{l_8, l_9\}$, but now we see that $B_{8,9}$=0 and is not allowed, meaning that $K_{0,0}$ is a complete subset of four features containing $\{l_0, l_1, l_7, l_8\}$, all with overlaps below T. The next combination, $K_{0,1}$, is the first allowed alternative to the final element of $K_{0,0}$: $K_{0,1} = \{l_0, l_1, l_7, l_9\}$. The left-hand plot of Figure 6.2 $K_{0,0}$ as a single red line originating from l_0.

The method employed for constructing paths closely resembles that of a depth-first search in an unweighted, directed acyclic graph (DAG), where the nodes represent features and the edges signify the allowed pairwise combinations between them. The directed and acyclic structure of the graph is maintained by imposing an ordering on the features, ensuring that the edges consistently point from lower to higher indices. Nevertheless, an important distinction arises: the choice of each feature is dependent on those allowed by all previous elements in the path, or in other words, the allowed vertices would be inherited. This *hereditary condition*, however, can be seamlessly incorporated into well-established DAG "simple path" algorithms [192, 193].

Recasting the problem as an optimum-path search requires minor changes to the definitions covered. Firstly, each path has to be defined between two points: a source and a sink. As previously stated, each combination within the subset K_i has a defined source, however, the final element in each set will depend on the path taken. A convenient way

to deal with this condition is to define a universally allowed nth feature so that every possible path terminates at index n. This can be done by appending an nth "sink" feature to ρ, this is shown in Figure 6.1 but can also be expressed as

$$\rho_{n,i} = \rho_{i,n} = 0.0 \;:\; 0 \leq i \leq n\,. \tag{6.4}$$

This modification of ρ necessitates that the definition of A_i also be modified to include the nth term:

$$A_i \equiv \{j : \rho_{i,j} < T,\; i \leq j \leq n\}\,. \tag{6.5}$$

With A_i defined in terms inclusive of n, we can define a modified hereditary depth-first search (HDFS) algorithm that generates all the available paths starting from an initial feature. This algorithm proceeds by recursively appending diminishing subsets of allowed elements $\mathcal{S}$, with the current subset $\mathcal{S}_c$ defined as the intersection of A_c with the previous subset such that

$$\mathcal{S}_c \equiv A_c \cap \mathcal{S}_{c-1}\,. \tag{6.6}$$

The total intersection of the compatible sets for the feature elements already in the path gives the remaining compatible features. As this is constructed iteratively, each stage of completion-refinement needs only to be compared against the set of completions for the current final element, $\mathcal{S}_{c-1}$. The HDFS algorithm uses this condition to exclude overlapping feature combinations from consideration efficiently. Figure 6.3 shows the PathFinder algorithm used to find the longest path for two examples of the BAM with ten (upper) and fifteen (lower) features. The right-hand plots show a graph representation of the hereditary condition with the longest path given by tracing the red arrow. The HDFS algorithm uses this condition to efficiently exclude overlapping feature combinations from consideration.

In summary, initiated from a source l_i, with the first element of $\mathcal{S}$ being A_i, the set of all allowed paths K_i can be built by recursively evaluating the subsets of $\mathcal{S}$. Once the current iteration has reached the "sink" l_n, a full path is defined by the steps taken. Upon each path completion, the set $\mathcal{S}$ is recalculated from the next element of the previous set ($\mathcal{S}_{c-1}$). This process is repeated until the "sink" (l_n) is reached within the first subset of $\mathcal{S}$. Figure 6.2 shows the results from running this algorithm using the BAM from Figure 6.1 for paths starting from l_0. The left-hand figure shows the first allowed path starting from l_0 $K_{0,0}$. The right-hand plot shows all paths that are allowed from source

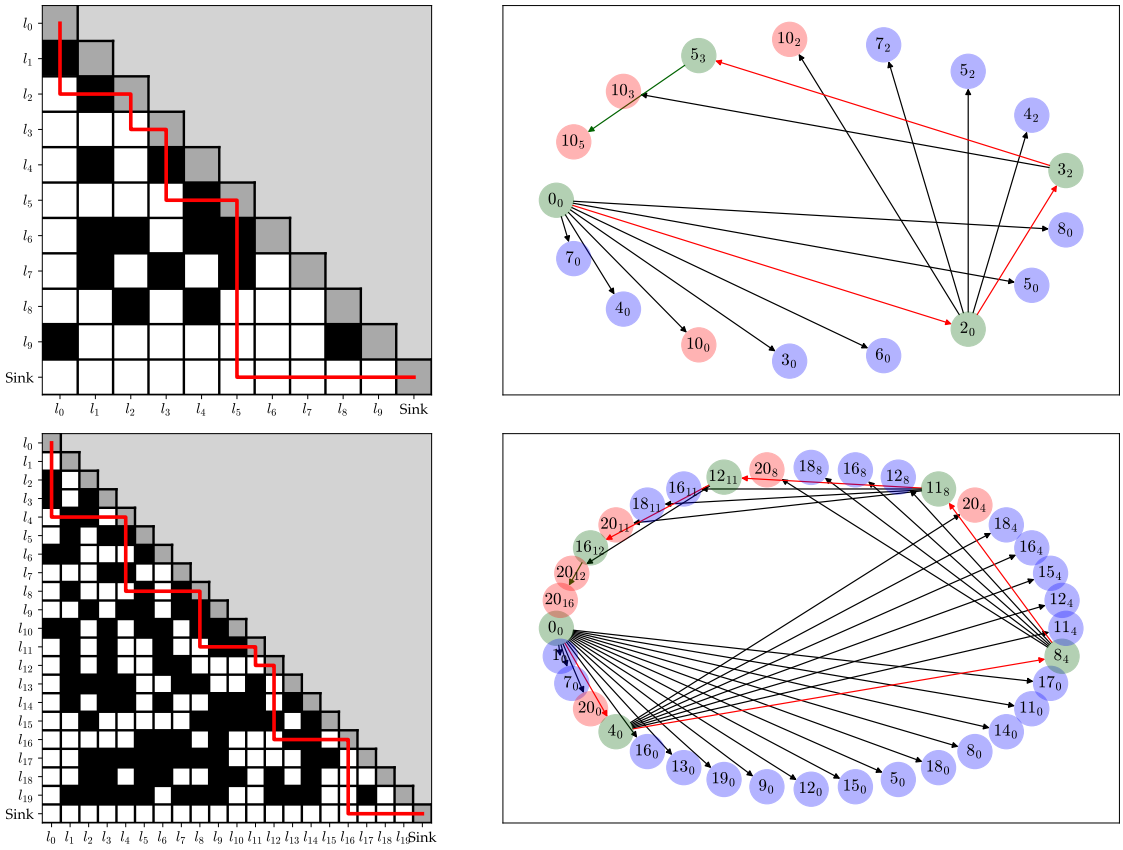


Figure 6.3.: Graph representation of the longest path identified from two examples BAM's using ten (upper) and fifteen (lower) features. The left-hand plots show the combinations as a path traced on the BAM, while the right-hand plots show the hereditary condition as a graph. The nodes of the graph evolve as with the selection such that the subscript of the label refers to the previous node. The sink node is shaded red, with the red arrows tracing the longest path to the sink. The nodes within the combination are shaded green, with the alternatives shaded blue.

l_0. The complete HDFS algorithm is given in pseudocode as Algorithm 1 in Appendix A. The PATHFINDER code is available at https://github.com/J-Yellen/PathFinder.

6.2. Weighted edges for sensitivity optimisation

So far, in generating the set of allowed paths, we have been concerned solely with feature-exclusivity and treated all graph edges (and hence feature) as of equivalent value within the fixed DAG ordering provided. However, this is most often not the case in practice: for each specific model, certain features will be more sensitive than others. Thus, the

edge weight will determine the optimum set in most use cases, which is any that isn't directly performing a longest-path search. Such weights should be motivated by the statistical goal being tested and, ideally, should be additive so standard longest-path optimisation can be used to identify the most sensitive allowed feature combination.

In general, the optimal path can be found in a reasonable time by evaluating the overall sensitivity metric for every allowed element combination identified by the HDFS algorithm of the previous section. However, in the case of additive weights, further algorithmic optimisations are possible by a) ordering the elements in decreasing order of individual sensitivity and b) exiting early from the generation of allowed-path subsets for which there is no possibility of exceeding the metric obtained for the current maximum-sensitivity path. The first of these conditions is implemented by *a priori* ordering the elements according to decreasing weight, such that paths dominant features are evaluated first — this opens the possibility of evaluating only the sets of paths starting with the first $\mathcal{O}(10)$ elements. The second makes such a manual cut-off largely redundant by maintaining records of the highest complete-path weight and the sum of weights over all remaining elements in $\mathcal{S}_c$ as the allowed paths are generated. Should the sum of the current path's weight and its maximum possible completion become smaller than the current best complete path, there is no point in continuing to evaluate that set of completions. This condition is given by

$$\sum_{i \in \mathcal{S}_c} \omega_i + \sum_{j \in A_{c+1}} \omega_i > \sum_{k \in K_{\text{best}}} \omega_k \,, \tag{6.7}$$

where ω_i is the weight of feature i and K_{best} is the set of features with the highest combined weight evaluated so far. This condition acts as a "short-circuit", reducing the path-finding algorithmic complexity. However, this final optimisation step has the consequence of no longer being able to track all of the possible paths available. This optimised version of the HDFS algorithm for weighted graph edges defines an alternative weighted HDFS algorithm (WHDFS). This algorithm is the final method for efficiently addressing the problem of finding the combination of statistically non-overlapping elements that maximises their additive combination of sensitivities in a given model. While the WHDFS outperforms the HDFS in terms of efficiently finding the best combination, the HDFS algorithm does not discard the information. Thus, the algorithms answer two distinct questions:

- What are the possible combinations? → HDFS
- What are the best combinations? → WHDFS

6.3. Optimisation and performance

Ordering

Introducing weights introduces the possibility of arranging the BAM to improve the run time. Finding an optimal set of features depends on which metric defines optimality, e.g. maximum information, p-value, or similar. By ordering the BAM according to decreasing single-feature contribution to the optimality metric, the optimum set will probably be initiated from an "early" l_i with $i \ll n$ (for large n). In practice, this means that the PATHFINDER algorithms only need to be run over the top few indices. Ordering the BAM and weights can be done by calling the `sort_bam_by_weights` method of the BAM class after defining the BAM object:

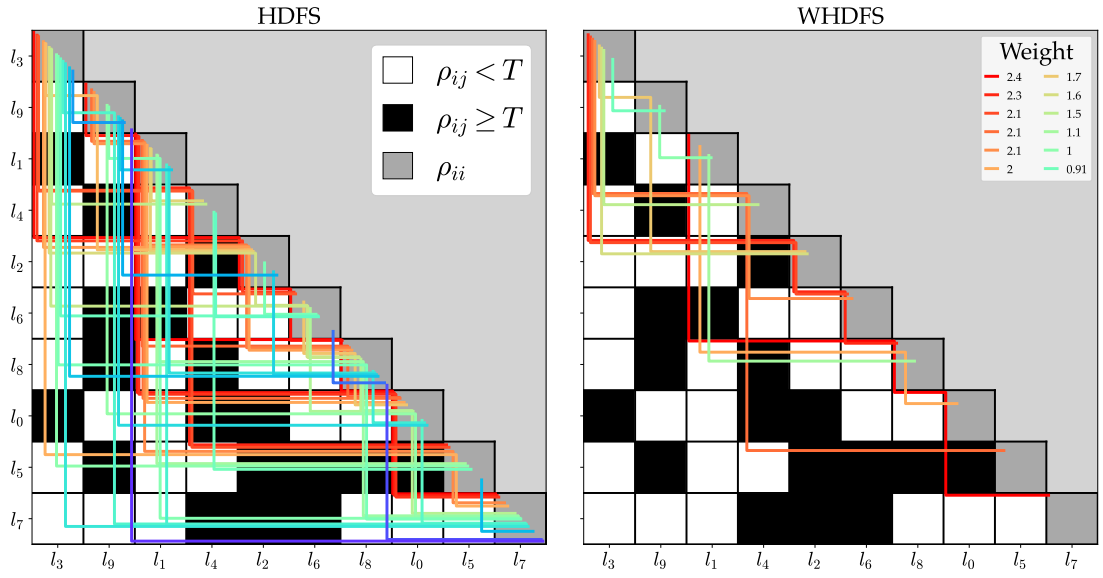
```
bam = pathfinder.BinaryAcceptance(input_matrix, weights=weights)
index_map = bam.sort_bam_by_weight()
```

This returns an array of indices that map back to the original configuration. After calculating the best path, the result can be converted back to the original indices via the `remap_path` method of the Results object.

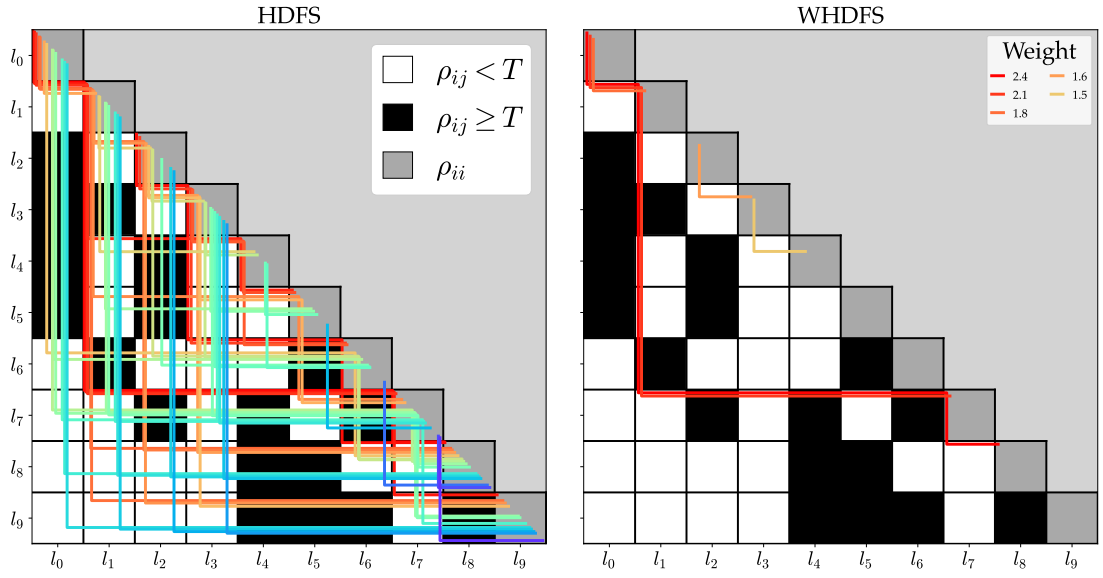
```
hdfs = pathfinder.HDFS(bam)
hdfs.find_paths()
results = hdfs.remap_path(index_map)
```

Here, we call an instance of the HDFS class and then calculate the results using the `find_path` method. The results are stored on a `Results` class–from which the HDFS and WHDFS classes inherit–which handles all of the paths and corresponding weights using ordered PYTHON data classes for each result. This structure allows for quick sorting of each path according to its weight.

To understand the benefit of introducing ordering, Figure 6.4 compares the internal path evaluations performed by each algorithm when calculating the best path for ten features. The upper plots (a) show the evaluations made while running on unordered weights; the HDFS performs 52 path evaluations while the WHDFS performs 12 to reach an identical conclusion. The lower plots (b) show the evaluation made using ordered weights. For the HDFS, the number of evaluations stays the same as no stopping criterion exists. However, the WHDFS algorithm only takes five evaluations, less than half of what was required in the unordered case. The best path, in this case, was $\{l_0, l_1, l_7, l_8\}$, for the unordered mode, the best path was evaluated on the third "run" starting at l_1 (the third



(a) Unordered



(b) Ordered

Figure 6.4.: Comparison between the HDFS and WHDFS algorithms running with 10 features. The plots demonstrate every path evaluation performed by the two algorithms running in weight-ordered (lower) and unordered (upper) modes. The HDFS evaluated 52 paths for both modes, evaluating each allowed combination before returning the best result. The WHDFS reduced the number to 12 paths in the unordered mode and 5 when ordered. For consistency, the axis labels have been presented in weight order for all plots.

feature). Both ordered runs evaluated the best path within the first iterations of the first run, demonstrating the benefits of weight ordering for both algorithms.

6.3.1. Subsets

A logical inconsistency was discovered early in developing the HDFS algorithm. If the BAM were constructed with all features combinable, one would expect a single result of length n containing all the features. However, the HDFS returned $\approx 2^n$ results, which made sense as that is the number of all possible paths shown in Equation (6.1). This exposed a subtle difference between the expected and designed behaviour because, as it turned out, the answer to the question "How many available paths there are?" depends on whether or not subsets are allowed in the combination. To achieve the behaviour of generating one path for a fully "legal" or "allowed" BAM, the definition of the superset K must be altered such that:

$$\forall K_{\alpha,i}, K_{\alpha,j} \in K_\alpha, \quad K_{\alpha,i} \not\subseteq K_{\alpha,j} \quad \text{and} \quad K_{\alpha,j} \not\subseteq K_{\alpha,i}\,, \tag{6.8}$$

where the number of combinations entering K depends on the fraction of the elements in BAM that are "allowed". The allowed fraction (f_A) is defined as:

$$f_\mathrm{A} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathrm{B}_{i,j}\,, \tag{6.9}$$

where the elements of the BAM (B_{ij}) are treated as binary; thus, the sum of elements must be less than or equal to $n(n-1)$. Figure 6.5 shows the number of paths available for twenty features for a given f_A. The BAM was randomly generated 50 times for each, and the HDFS algorithm was used to calculate the total number of paths under subsets-allowed and subsets-not-allowed configurations. Both configurations show the expected behaviour with the "allowed" results tending to 2^n and the "not-allowed" results tending to unity.

This modification becomes particularly important when dealing with negative weights. To demonstrate this issue, let's consider the following scenario:

$$W = \sum_{i \in K_{best}}^{l} \omega_i = 5 + 4 + 3 + 2 + 1 + (-1)\,. \tag{6.10}$$

The addition of the final weight in Equation (6.10) depends on the condition of allowing subsets. If allowed, the negative value is dropped in favour of a larger sum. There are examples of use cases where this is the desired behaviour, i.e., when data reduction is the goal and selecting the datasets most sensitive to a target is the primary task. However, there are also use cases where selection biases, such as cherry-picking or the

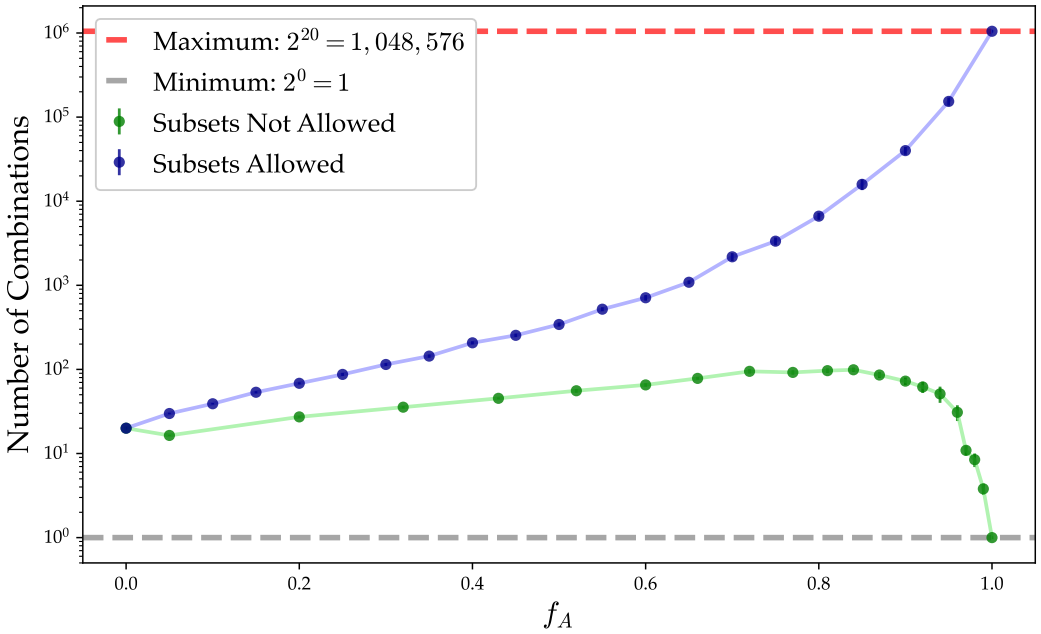


Figure 6.5.: Difference in the number of combinations when allowing and not allowing subsets.

"look-elsewhere effect", are of concern. In these cases, negative weight can signify data in favour of the null hypothesis, and thus, their inclusion reduces the chance of type I errors.

By default, the PathFinder module does not allow negative weights. When calling an instance of the `BinaryAcceptance` object with negative weights, a warning is raised, providing the user with two suggestions. The first is to rescale the weights to positive values; this option must be used if subsets are not allowed in the calling of HDFS or WHDFS. The second option is to set the flag `allow_negative_weights=True` when calling the `BinaryAcceptance` object. If the first option is chosen and the weights are rescaled, the original weights can be recovered via the `remap_path` method of the HDFS or WHDFS object:

```
constant_shift = abs(min(weights)) + 1
weights = weights + constant_shift
bam = pathfinder.BinaryAcceptance(input_matrix, weights=weights)
hdfs = pathfinder.HDFS(bam)
hdfs.find_paths()
results = hdfs.remap_path(weight_offset=constant_shift)
```

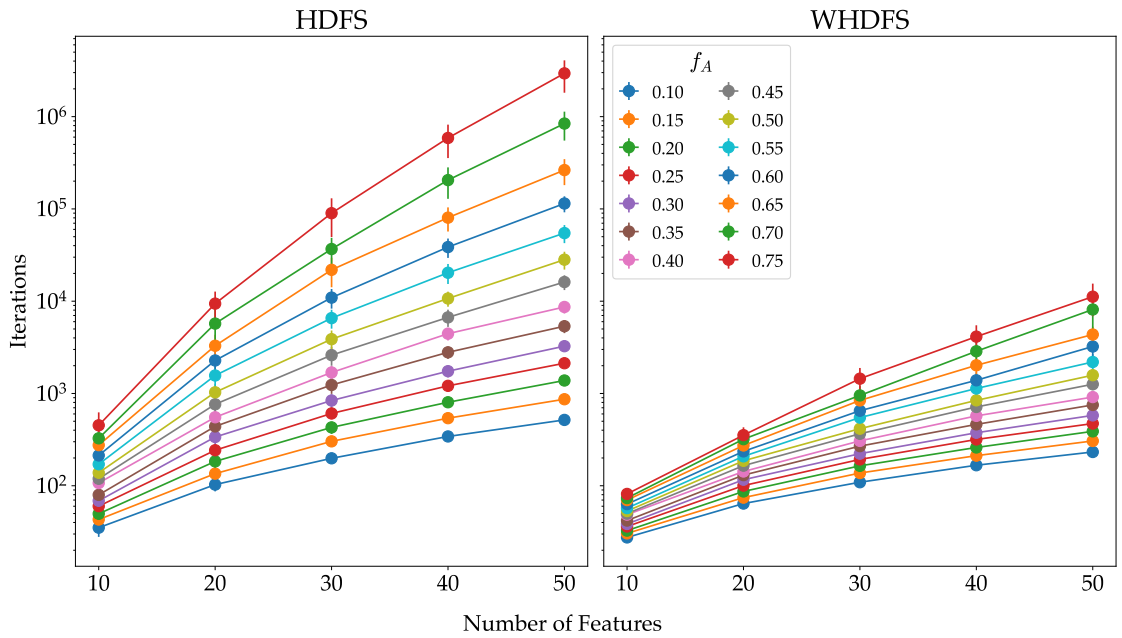


Figure 6.6.: Comparison of the number of iterations, or "steps" taken by the HDFS and WHDFS algorithms. The analysis was performed over five BAM sizes indicated by the "number of features", each generated with different values of f_A. In the most extreme case, the WHDFS outperformed the HDFS algorithm by three orders of magnitude.

6.3.2. Performance

The performance of the PATHFINDER algorithms is dependent on two factors. First is the dimensionality (N), the number of features entering the BAM. The second is the fraction of allowed combinations f_A. The algorithm's performance can be measured in two ways: completion time, i.e., how long it takes to find the optimum combination. Alternatively, one could measure the number of iterations required, which is another way of saying the number of "steps" taken. Figure 6.6 compares the HDFS and WHDFS algorithms applied to five BAM matrices ranging in size from $n = 10$ to $n = 50$ features with f_A ranging from 0.1 to 0.75. For each configuration, the same BAM object was provided to both algorithms, calling the PATHFINDER, HDFS and WHDFS classes via:

```
hdfs = pathfinder.HDFS(bam, top=1))
hdfs.find_paths()
wdfs = pathfinder.WHDFS(bam, top=1)
wdfs.find_paths()
```

Where "top" refers to the number of best-performing results to be tracked, i.e. keep the top n results. Both algorithms essentially loop over the available nodes; thus, they can be arbitrarily modified to keep track of the number of loops or "steps" performed. Looking at the right-hand plot of Figure 6.6, the number of iterations performed by the HDFS scales quickly with the BAM size, less than exponential but greater than logarithmic. Moving to the left-hand WHDFS plot, the scaling is reduced to near $n\log(n)$ for $f_A = 0.1$. The red (upper) line corresponding to $f_A = 0.75$ shows a difference of three orders of magnitude in the number of iterations. It is reasonable to assume that both algorithms take the same time per iteration. Thus, the gain in efficiency for the WHDFS algorithm reduces the time complexity close to $\mathcal{O}(n \ln n)$[1]. This is a valuable result as it means that a combinatorial problem that once required approximations or large amounts of computing time now has a solution for situations that involve potentially thousands of features.

[1] A comparison between the HDFS, WHDFS and a brute force approach is provided in Section 7.4.5

Chapter 7.

Exclusion: Testing Analysis Combinations

7.1. The TACO project

Multiple theories of physics beyond the standard model (BSM) have been proposed to address the questions left by the Standard Model. Over the last decade, the ATLAS and CMS experiments at the Large Hadron Collider (LHC) have played a large part in excluding the most obvious models beyond the Standard Model (BSM) and, as such, more complex models are entering the game. These bring increasing complexity in both the dimensionality of their parameter spaces and the range of phenomenology possible within them. This increases the probability that new physics will not be discoverable via one powerful experimental signature but will disperse across many signatures at a level below direct exclusion in any search analysis. Thus, to view what the LHC tells us about physics beyond the Standard Model, it is crucial that different BSM-sensitive analyses can be combined.

The current construction of search analyses means that comprehensive combinations require knowledge of how the same events co-populate multiple analyses' signal regions. The TACO (Testing Analysis Combinations) project presented a novel, stochastic method to determine this degree of overlap. It combines this with the WHDFS algorithm to find the optimum combination of signal regions (see Chapter 6). The project published the paper "Strength in numbers: Optimal and scalable combination of LHC new-physics searches" in April 2023 [194]; much of the research presented in the following chapter is taken from the TACO paper with additional content taking a retrospective view of the project. My

contributions to the TACO project included developing and implementing the WHDFS (see Chapter 6), which I used to generate the results presented in Section 7.4. While the procedures for calculating overlaps between signal regions (SRs) were already well established before my involvement, I contributed to the implementation and fine-tuning of the final procedure as presented in the TACO paper [194].

Before we get into the main content, we must carefully define the term "optimum combination". For the TACO project, this was taken to mean a set of signal regions with no mutual overlap in event co-population that simultaneously optimises expected upper limits on BSM-model cross-sections. With this definition, the project demonstrates the gain in exclusion power relative to single-analysis limits using models with varying degrees of complexity, ranging from simplified models to the 19-dimensional pMSSM.

The project primarily focused on using the MADANALYSIS 5 analysis toolkit with SMODELS to investigate how to best combine analyses for optimal statistical power. The motivation for the project came from the idea that definitive statements about any dispersed signature would require the combination of as many analyses as possible. As previously stated, not all analyses *can* be combined because simply combining the test statistics of every signal region (SR) from every analysis would certainly double-count physics effects since the same events could pass multiple analyses' event-selection cuts and observable binnings. As observables used for selection-cut purposes can be highly correlated, with a complex dependence on the rest of the selection phase-space, it is impossible to reliably identify these degrees of overlap directly from a list of cut observables and values. And crucially, even when the analysis correlations *are* known, there remains the problem of identifying which compatible subset will place the optimal constraint on any given BSM model.

7.2. Overlap estimation

To investigate how analyses could be combined to provide the most stringent constraints on a BSM model point, the selection of analyses available in both SMODELS and MADANALYSIS 5 was chosen as the database. At the time of TACO project in 2020, this included 19 analyses: ATLAS-SUSY-2013-02 [195], ATLAS-SUSY-2013-04 [196], ATLAS-SUSY-2013-05 [197], ATLAS-SUSY-2013-11 [198], ATLAS-SUSY-2013-21 [199], ATLAS-SUSY-2015-06 [200], ATLAS-SUSY-2016-07 [201], ATLAS-SUSY-2018-04 [202], ATLAS-SUSY-2018-06 [203], ATLAS-SUSY-2018-31 [204], ATLAS-SUSY-2018-32 [205],

ATLAS-SUSY-2019-08 [206]; CMS-SUS-13-011 [207], CMS-SUS-13-012 [208], CMS-SUS-16-033 [209], CMS-SUS-16-039 [210], CMS-SUS-16-048 [211]. CMS-SUS-17-001 [212], and CMS-SUS-19-006 [213].

The cascade decays, or *topologies*, addressed by these analyses were simplified to focus on the production of two massive BSM states, each decaying into at most 2–3 final-state particles. For example, CMS-SUS-19-006 [213] covered six simplified models with 2–3 final-state particles, four with direct gluino pair production: T1tttt, T1bbbb, T1qqq, and three with direct squark pair production: T2tt, T2bb and T2qq. The topologies covered by all the named analyses, following the SModelS naming convention [214], were: T1, T1bbbb, T1btbt, T1tttt(-off), T2, T2bb, T2tt(-off), T2bbWW(-off), T2bt, T2cc, T3GQ, T5, T5bbbb, T5tctc, T5tttt, T5GQ, T5WW(-off), T5WZh, T5ZZ, T6bbhh, T6bbWW(-off), T6WW(-off), T6WZh, TChiChipmSlepL, TChiChipmSlepStau, TChiChipmStauStau, TChiChipmSlepSlep, TChipChimSlepSnu, TSlepSlep, TChiZZTChiWH, TChiWW, TChiWZ(-off), TChiZoff, TGQ, TSlepSlep, and TStauStau.

7.2.1. Model-space sampling and event generation

Efficiently estimating the degree of overlap between different SRs required an efficient sampling strategy. One of the first issues encountered by the project was how to define a sample space that contained all the relevant analyses in terms of a given simplified model. To achieve this, we considered the minimal volume in the mass-parameter space in terms of a convex hull. In mathematics, the convex hull of a set of points is the smallest convex set that contains all the points. Intuitively, you can think of it as the shape you would get if you stretched a rubber band around the outermost points of a set; the rubber band would enclose all the points and form a convex polygon or polyhedron. In more formal terms, a convex set is one in which, for any two points within the set, the line segment connecting them also lies entirely within the set. As such, the convex hull is the intersection of all convex sets that contain the given set of points. It is the smallest such convex set [215].

In 2D, the convex hull forms a polygon, while in 3D, it forms a polyhedron. For example, if you have a set of scattered points in the plane, their convex hull would be the boundary of the minimal convex polygon that encloses all the points. In 3D space, the convex hull would be a polyhedron where the outer faces form the smallest convex shape that encloses all points.

In N dimensions, the concept of a convex hull generalises similarly to how it works in 2D or 3D. The convex hull of a set of points in N-dimensional space is the smallest convex set that contains all the points. Given a set of points, $X = \{x_1, x_2, \ldots, x_k\}$ in N-dimensional space $\mathbb{R}^N$, the convex hull is the set of all convex combinations of these points. A convex combination is any point that can be written as:

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \,, \tag{7.1}$$

where the coefficients λ_i satisfy:

1. $\lambda_i \geq 0$ for all i (non-negative).
2. $\sum_{i=1}^{k} \lambda_i = 1$ (the sum of the weights) .

The resulting convex hull is the smallest convex "shape" that encloses all the points in X. As such, in ND, the convex hull forms a convex polytope [216], which is the N-dimensional analogue of a polygon (in 2D) or polyhedron (in 3D).

With an understanding of how to define a minimal volume in a high-dimensional space, the following procedure was followed to estimate signal overlaps for arbitrary scenarios. For each analysis, a convex hull was constructed in each simplified model's parameter space accessed by a given topology, using the efficiency maps implemented in SModelS [217]. The efficiency maps provided upper limits on the production cross-sections of the two relevant BSM states, depending on the masses in the simplified decay chains. For each simplified model, a convex hull existed for each analysis that included results for that specific model. The focus was on the joint set of convex hulls corresponding to each simplified model. A contour was then constructed around the mass-parameter space beyond which the expected event yield from all corresponding analyses was zero. The union of these regions was populated with events without redundantly populating areas shared between analyses. Events were uniformly generated within this joint set of convex hulls to introduce minimally informative flat prior to the procedure.

The MC events were generated at leading order (LO) with MadGraph5 (v2.6.5) [218] at the partonic level using the NNPDF 2.3 LO [219] set of parton distribution functions via the Lhapdf library [220], with parton-showering and hadronisation simulated by Pythia 8 [221] through the MadGraph5 interface. Detector-level events were obtained using Delphes 3 and FastJet [183, 222], executed through MadAnalysis 5 with analysis-specific configurations interleaved with the event-selection logic. The input for the generation pipeline was a corresponding SLHA-format [223] data file for each

topology, with the masses of the produced, final, and (in some cases) intermediate BSM states defined as free parameters. The initial partonic processes in the generation chain involved direct production of the topology's massive BSM states, with decay chains implemented via PYTHIA 8's decay mechanism.

The required output of the MC generation procedure was an *shared-event matrix* Θ of shape $N_{\text{evt}} \times N_{\text{SR}}$, where $\Theta_{e,s} = 1$ meant that event e populated SR-bin s, and *vice versa* $\Theta_{e,s} = 0$ indicated that event e did not pass the cuts for SR s. This matrix was produced using a command in MADANALYSIS 5,

```
set main.recast.TACO_output = <file-name>,
```

that was added to the framework for this purpose. The *shared-event matrices* were saved as text files with each event corresponding to a pair of lines encoding first the list of floating-point event weights [224] (in this study, we use only the nominal weight), and then a list of `0` and `1` characters corresponding to the N_{SR} signal regions. These files were written separately for each DELPHES 3 configuration to the location:

```
<Output>/SAF/defaultset/<delphes-card-name>.<file-name>.
```

In the output directory of the recasting process, the minimal number of Monte-Carlo events needed for a reliable estimation of the binary acceptance matrix was determined by starting with 100 events for 1000 random parameter points sampled from the union of the convex hulls of the signal regions, resulting in an initial $N = 100,000$ events. For any pair of signal regions SR_1 and SR_2 populated with n_1 and n_2 events, respectively, the number k of shared-events was determined. If $k > 100$, enough statistics had been accumulated, and the bootstrapping procedure proceeded (see Section 7.2.2).

For $k \leq 100$, with $n \equiv n_1 + n_2$, the confidence interval construction by Clopper and Pearson [225] of the binomial distribution was used.

$$\text{B}(n, p) = \binom{n}{k} p^k (1-p)^{n-k}\,, \tag{7.2}$$

where p is a free parameter defined as the *probability of overlap*. To guarantee enough events for the case of a negligible overlap, we had to obtain a one-sided (upper) confidence interval for p at confidence level $\text{CL} \equiv 1 - \alpha = 0.95$ and guarantee that it was below a certain threshold. From the Clopper–Pearson construction, this was computed as the

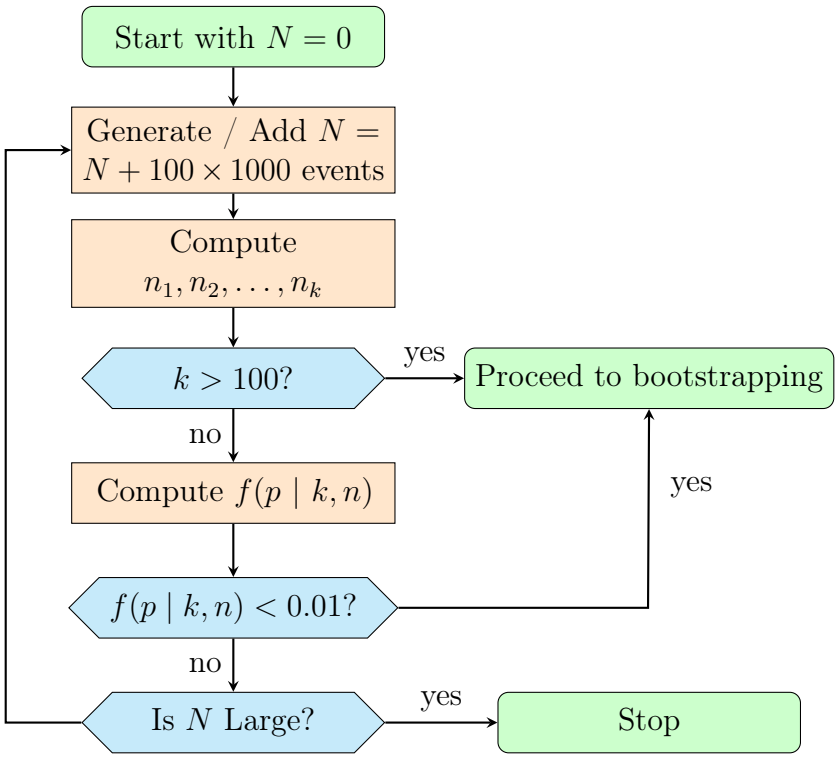


Figure 7.1.: Flowchart for determining the number of Monte Carlo events needed to estimate the overlap matrix.

$1 - \alpha$ quantile of the β distribution

$$f(p \mid k, n) = \beta_{1-\alpha;k+1;n-k}\,. \tag{7.3}$$

Suppose this upper bound is below the arbitrarily chosen threshold, $f(p \mid k, n) < 0.01$. In that case, we assumed that we had accumulated enough statistics to safely infer the potential absence of significant overlap, and we confidently proceeded to the bootstrapping procedure. Using these criteria, we employed the logic of Figure 7.1, as it guaranteed that enough statistics were available to robustly and reliably determine a significant or negligible overlap between a given pair of signal regions.

7.2.2. Overlap-matrix estimation

With a set of sufficiently populated SRs, the project moved to determine whether or not such SRs were approximately orthogonal or disjoint with respect to one another. The Pearson correlation can be estimated from the acceptance matrix via the event-averaged

covariance,

$$\begin{aligned} \mathrm{cov}_{ij} &= \langle \Theta_i \Theta_j \rangle - \langle \Theta_i \rangle \langle \Theta_j \rangle \,, \\ &\equiv \frac{\sum_e \Theta_{e,i} \Theta_{e,j}}{N_{\mathrm{evt}}} - \frac{\sum_{e'} \Theta_{e',i} \;\cdot\; \sum_{e''} \Theta_{e'',j}}{N_{\mathrm{evt}}^2} \,, \end{aligned} \tag{7.4}$$

where N_{evt} is the number of events in the estimation sample and, as made explicit in the second line, i and j are SR indices. This method was possible because the entire event-wise shared-event matrix was available; hence, overlaps can be estimated by averaging over the event axis of the matrix.

An alternative approach, which was employed in this project, involved bootstrap sampling from a unit Poisson distribution. In this method, each event was assigned N_{boot} independent "bootstrap weights" $w_{e,b} \sim \mathrm{Pois}(\lambda = 1)$, which were subsequently aggregated to construct N_{boot} replicas of the signal region (SR) yield estimates. This procedure resulted in an $N_{\mathrm{SR}} \times N_{\mathrm{boot}}$ *yield matrix* Y, where each entry represents the sum of event weights falling into the SRs for a given bootstrap replica. The matrix encapsulates N_{boot} alternative realisations of the yields, generated from a single set of input events.

The degree of overlap between different SRs was then quantified through their correlated weight fluctuations across bootstrap replicas, yielding an alternative estimate of the covariance:

$$\begin{aligned} \mathrm{cov}_{ij} &= \langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle \\ &\equiv \frac{\sum_b Y_{i,b} Y_{j,b}}{N_{\mathrm{boot}}} - \frac{\sum_{b'} Y_{i,b'} \;\cdot\; \sum_{b''} Y_{j,b''}}{N_{\mathrm{boot}}^2} \,. \end{aligned} \tag{7.5}$$

A key distinction of this method is that the averaging is performed over bootstrap replicas of the aggregate yields, rather than over individual event-level acceptance tuples. While not critical in the present implementation, the bootstrap approach offers a practical advantage by eliminating the need for a linearly growing acceptance matrix. Instead, it maintains a fixed-size $N_{\mathrm{SR}} \times N_{\mathrm{boot}}$ yield matrix, which may become computationally significant for large event samples. From the covariance matrix obtained via either method, we then define the *overlap matrix*.

$$\rho_{ij} = \frac{\mathrm{cov}_{ij}}{\sqrt{\mathrm{cov}_{ii}\,\mathrm{cov}_{jj}}} \,, \tag{7.6}$$

following the usual Pearson correlation definition. Lower-triangle plots of this symmetric acceptance matrix for the sets of signal regions common to SModelS and MadAnalysis 5 are shown in the Appendix B Figures B.1 and B.2 for 8 TeV and 13 TeV LHC data-analyses respectively, with patterns of highly and partially co-populated SRs visible. Finally, the BAM B between SR-pairs SR_i and SR_j was derived by applying an "acceptable overlap" threshold T such that the overlap between SRs i and j is $B_{ij} = (|\,\rho_{ij}| \leq T)$. The value chosen for T is somewhat subjective, reflecting that for each use-case, there will be a finite value of ρ_{ij} below which double-counting biases are not statistically resolvable: treating these low correlations as zero-correlations avoids blocking useful SR combinations due to irrelevant and noisy correlation estimates.

The procedure described above is implemented in the public Python program TACO (Testing Analyses COrrelations), available at https://gitlab.com/t-a-c-o/taco_code.

7.3. Optimal signal-region combination

To fully take advantage of the WHDFS algorithm, we would need to define edge weights that maximise the exclusion power given a model point. With this, we can select a set of signal regions that results in the maximum expected significance for exclusion. For each specific BSM model, some signal regions will be more sensitive than others. For example, leptophilic models naturally tend to see the most sensitivity in SRs with multilepton signatures; models with enhanced couplings to the third generation have the most impact on t- and b-quark and τ-lepton signatures; and dark-matter models favour jet + missing transverse-energy signatures. In addition, when not all SRs have the same integrated luminosity, SRs in high-luminosity datasets are naturally more sensitive than those in low-statistics ones.

7.3.1. Selection of weights for exclusion

A typically appropriate choice for the edge weights, and one that is motivated by use in collider physics, is the logarithm of the expected likelihood-ratios (LLR) between the signal-model under test and the background-only model, for pseudodata equal to the

expected yields under the background-only model.

$$\omega_i = -2\ln\left(\frac{\mathcal{L}_i(\mu=1,\hat{\hat{\boldsymbol{\theta}}}_e)}{\mathcal{L}_i(\hat{\mu},\hat{\boldsymbol{\theta}}_e)}\right) = -2\ln\Lambda_i^{\exp}(\mu)\,, \tag{7.7}$$

where the index i links the i-th weight to SR (node) i and $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_z)$ denotes all of the nuisance parameters, with b_z being the total number of background events. The likelihoods are obtained using the SModelS [81] software package, which implements a mixture of pyhf [187] and Simplified Likelihoods, [226]. The precise form of the likelihood changes with the analysis and depends on the data provided by the experiment. However, the use of likelihoods in SModelS can be summarised by a generalised example:

$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \boldsymbol{\Delta}\big) = \prod_i \big[\mu s_i(\boldsymbol{\theta}) + b_i + \Delta_i\big]^{n_i} \frac{1}{n_i!} \exp\big(-(\mu s_i(\boldsymbol{\theta}) + b_i + \Delta_i)\big) P(\boldsymbol{\Delta})\,. \tag{7.8}$$

Here, $\mathcal{L}(n \mid \mu, s(\boldsymbol{\theta}), b, \boldsymbol{\Delta})$ represents the likelihood function, which is the product of Poisson PDF's (see Section 3.2.3) for each bin. It describes the probability of observing n_i events in a given bin, given the expected signal contribution s_i, background contribution b_i, and signal strength μ at the model point $\boldsymbol{\theta}$. The nuisances, contained within the vector $\boldsymbol{\Delta}$, are described by a non-specific distribution $P(\boldsymbol{\Delta})$ and are coupled to the Poisson. This equation represents a simplified explicit likelihood form, where the new nuisance parameters directly modify the nominal background event yield, $b_i \to b_i + \Delta_i$, and the $P(\boldsymbol{\Delta})$ term imposes a penalty on such deviations It is worth mentioning that nuisances can be completed objects where the elements of $\boldsymbol{\Delta}$ can correspond to many factors that affect the rates for signal and background processes. Influences such as energy scale, b-tagging efficiency, and detector effects are often incorporated; these elementary nuisance responses are generally non-linear, resulting in a completed likelihood function that is difficult to report as an analysis outcome. The Simplified Likelihoods scheme replaces this full complexity with linear responses of expected (background) event yields with an effective nuisance parameter Δ_i for each bin, with the interplay between elementary nuisances absorbed into a covariance matrix ($\boldsymbol{\Sigma}$) between the new effective nuisances.

One option for $P(\boldsymbol{\Delta})$ could be an additional Poissonian corresponding to the control sample containing background events, which can be used to construct a histogram of some chosen kinematic variable, thus constraining b_z. Alternatively, one could replace the elementary nuisances with effective ones, corresponding to the terms in the bin-covariance matrix. In this scenario, $P(\boldsymbol{\Delta})$ would most naturally be a multivariate Gaussian of bin-value mean and covariance $\boldsymbol{\Sigma}$. Additional improvements could be made by adding

some asymmetry to that Gaussian, bringing it closer to the Poisson, which is especially useful for reducing negative-rate sampling. For the moment, the exact form of $P(\boldsymbol{\Delta})$ is of minimal concern when considering the test statistic, as ω_i is the profiled likelihood ratio; thus, the likelihoods are profiled over $\boldsymbol{\Delta}$, giving:

$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b\big) = \max_{\boldsymbol{\Delta}}\big\{\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \boldsymbol{\Delta}\big)\big\}\,. \tag{7.9}$$

The use of the Equation (7.7) is motivated by the following logic:

1. As we are combining a set of direct-search analyses in which no individual significant signal was found, we choose to frame our mission primarily as maximising the volume of model exclusion rather than a discovery. Our null hypothesis is, hence, the BSM signal model, and we seek to overturn it with a preference for the SM at every point in its parameter space.

2. Maximum exclusion power at each point hence corresponds to minimising the rate β of Type-2 errors (false-negatives, i.e. identifying the data as BSM when it is SM), hence maximising the conventional statistical power $1 - \beta$.

3. We hence aim to maximise the expected significance of exclusion Z at each point in the BSM parameter space. Under the assumptions of Wald's Theorem [227, 228], the expected significance is given by the square root of the LLR between the models; hence, maximising the LLR maximises the expected model-exclusion.

As any generated path is by definition composed of SRs, which can be treated as non-overlapping, the total log likelihood-ratio (LLR) of an element subset is just the sum of such weights along its corresponding path candidate C:

$$\Omega_E = \sum_{i \in C} \omega_i\,. \tag{7.10}$$

The use of expected-background pseudodata rather than the actually observed data counts is important to avoid cherry-picking statistical fluctuations. To avoid bias, we identified the optimal element-combinations for each point as if the data had not yet been recorded.

The desired distribution of Ω_E can be found starting with a result due to Wald [227], who showed that for the case of a single parameter of interest, and expanding this to the

sum of minimally correlating LLRs such that:

$$\Omega_E = \sum_i -2\ln\left(\frac{\mathcal{L}_i(\mu=1,\hat{\hat{\boldsymbol{\theta}}}_e)}{\mathcal{L}_i(\hat{\mu},\hat{\boldsymbol{\theta}}_e)}\right) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}\big(1/\sqrt{N}\big)\,, \tag{7.11}$$

where we assume $\hat{\mu}$ is normally distributed with mean μ' and standard deviation σ for a sample size N. With these assumptions and neglecting the $\mathcal{O}(1/\sqrt{N})$ term, one can show that Ω_E follows a non-central *chi-square* distribution with one degree of freedom,

$$\begin{aligned} f(\Omega_E \mid \mu) = \frac{1}{\sqrt{2\pi}}\frac{1}{2\sqrt{\Omega_E}}\Bigg[&\exp\Big(-\frac{1}{2}\Big(\sqrt{\Omega_E}+\frac{\mu-\mu'}{\sigma}\Big)^2\Big) \\ &+\exp\Big(-\frac{1}{2}\Big(\sqrt{\Omega_E}-\frac{\mu-\mu'}{\sigma}\Big)^2\Big)\Bigg] \end{aligned} \tag{7.12}$$

where the $(\mu-\mu')/\sigma$ term defines the non centrality of the distribution. However, by definition, under the SM background hypothesis $\mu = \mu' = 0$; thus, the equation reduces to the χ^2 distribution with one degree of freedom [228].

$$f(\Omega_E \mid \mu) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\Omega_E}}\exp\Big(-\frac{1}{2}\Omega_E\Big) \tag{7.13}$$

This is a crucial result for the project as it provides a robust weight and easy access to p-values via the χ^2 CDF. Alternative test statistics for other use cases, particularly anomaly detection, in which the observed data is compared to background expectations in search of the most consistent, discrepant, non-overlapping subset of measurements, a modified metric is required, but with similar motivation. For such use cases, however, the edge weights are, in general, no longer additive, resulting in a more complicated and CPU-intensive task.

7.3.2. Combination

Using the SModelS toolkit, the weights defined by Equation (7.7) were calculated for all SRs for a given model. This was done by producing SUSY Les Houches Accord data files (SLHA files) [229, 230] and the following classes from SModelS:

```
from smodels.base.model import Model
from share.model_spec import BSMList
from smodels.share.models.SMparticles import SMList
```

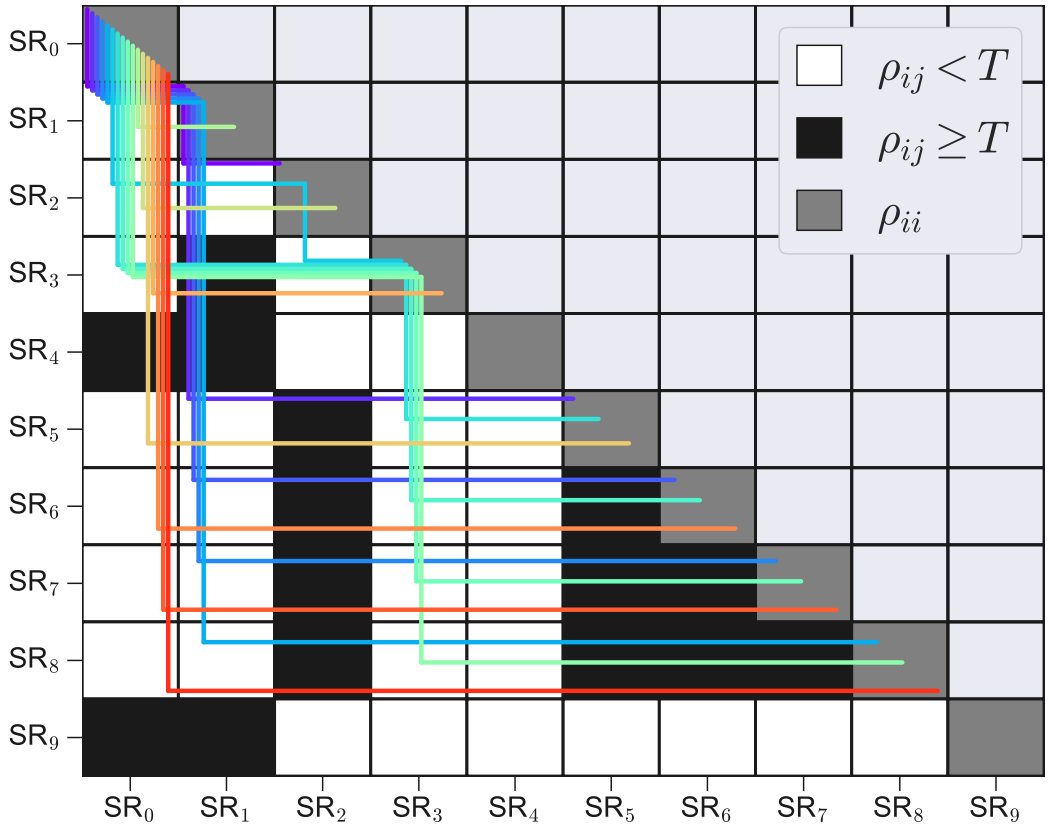


Figure 7.2.: BAM constructed from 10 signal regions ($SR_0 - SR_9$) with values masked according to threshold T. The weights have been calculated using a model point from the "T1" results (Section 7.4. The coloured lines show all allowed paths originating at SR_0.

```
from smodels.decomposition.decomposer import decompose
from smodels.base.physicsUnits import GeV, FB
from smodels.matching.theoryPrediction import TheoryPrediction
```

A Model object could then be called from SModels and subsequently decomposed into SMS topologies:

```
model = Model(BSMparticles=BSMList, SMparticles=SMList)
model.updateParticles(inputFile=SLHA_file)
topology_list = decompose(model, sigmacut=0.005*fb,
minmassgap=minmassgap=5*GeV))
```

From the SModels database, the experimental results were used to generate a theory prediction object using the `TheoryPrediction` class. From this object, the relevant likelihoods were calculated to evaluate ω_i for each SR. Figure 7.2 shows a selection of

SRs contributing to a model point taken from the "T1" SMS topology. The coloured lines show all ranked paths starting from SR_0 with combinations ranked highest to lowest, going from violet to red. With this machinery in place, we could evaluate the optimum combinations over the mass space of simplified models and extend to more complex models.

Before moving to the results of the TACO project, it is worth examining how the combinations were used to estimate exclusion. The experimental results in the SModelS database can be categorised in two main ways [231–233]:

- **Upper Limit** (UL) results: these correspond to the observed limits on the production cross-section (times the branching ratio) for simplified model topologies as a function of the BSM masses obtained by the experimental collaborations, $\sigma^{\text{UL}}_{\text{obs}}$. The results can also include the expected upper limits, denoted as $\sigma^{\text{UL}}_{\text{exp}}$. All limits are given at 95% confidence level (CL).
- **Efficiency Map** (EM) results: these correspond to signal efficiencies–or more precisely, acceptance $\times$ efficiency, $(\mathcal{A} \times \epsilon)$–for simplified topologies as a function of the BSM masses for the signal regions considered by the corresponding experimental analysis. These results include information about the number of observed and expected events with the corresponding uncertainties for each signal region

For the TACO project, only the "efficiency map" results were used as they contained the event information needed to construct the likelihood functions. However, for results that considered a single topology, it was convenient to present the data in terms of the r-value, which is defined as

$$r = \frac{\sigma_{\text{signal}}}{\sigma_{\text{UL}}}\,, \tag{7.14}$$

where $r \geq 1$ means that an experimental result is excluded. The upper limit is defined at the 95% CL, thus, an isocontour at $r = 1$ is equivalent to the 95% CL.

7.4. TACO results

To demonstrate the effectiveness of our approach, we presented physics results for various BSM-reinterpretation scenarios. Starting with simpler cases, we first demonstrated increases in model-exclusion limits in the context of simplified models, as detailed in

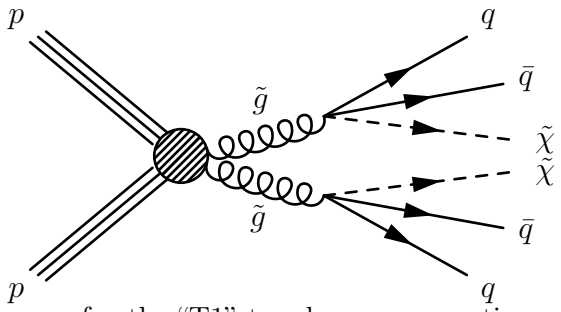


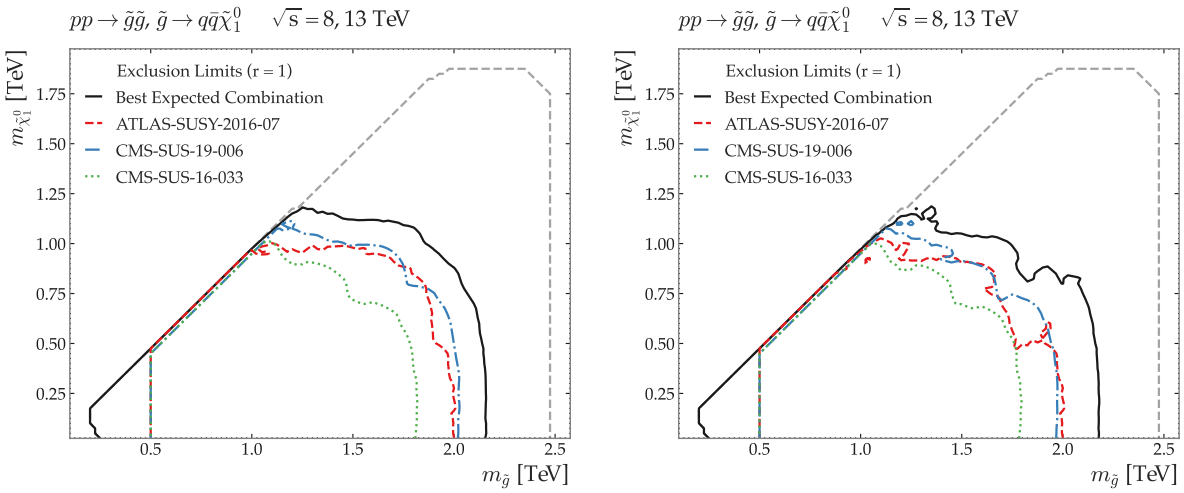
Figure 7.3.: Feynman diagram for the "T1" topology, representing a simplified gluino pair-production scenario followed by the decay into a qq pair and a neutralino $\tilde{\chi}$.

Section 7.4.2. Moving to more complex scenarios, we then showed the impact of our combinations on the pMSSM-19 model, which is discussed in Section 7.4.3. Finally, we concluded by analysing our combination results in the context of a simple t-channel dark matter model, where we fully recast the relevant analyses, as described in Section 7.4.4.

Throughout the project, control regions and overlaps in the background expectations of the signal regions were disregarded. This approach reflected the assumption that the signal regions were sufficiently specific to event topology and kinematic phase space, making distinguishing between signal and background events unnecessary. The removal of reducible backgrounds had already been addressed through pre-selection and SR-cut definitions. As a result, overlap estimation could also have been performed using large background-event samples rather than relying solely on sampling across signal models.

7.4.1. T1 simplified-model combination

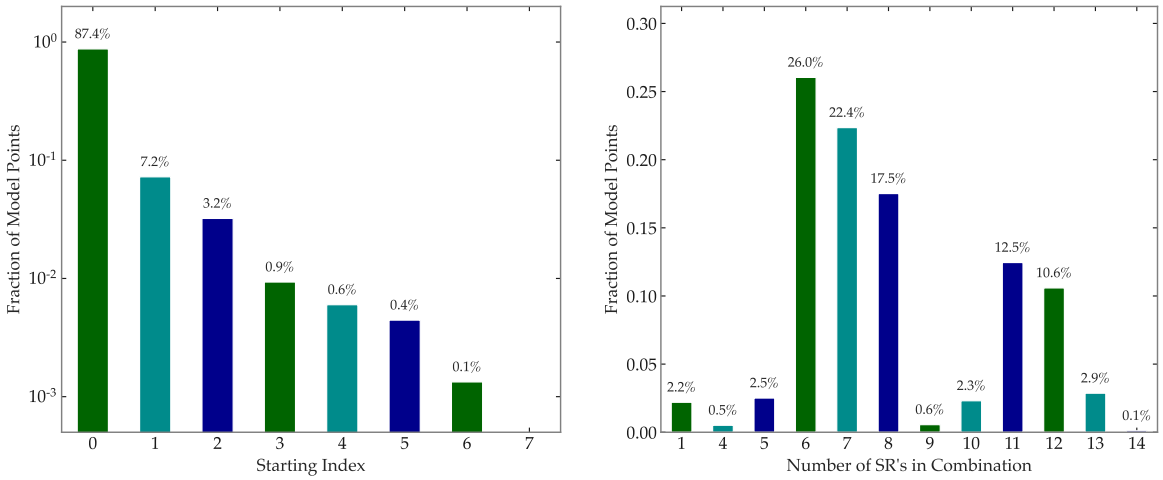
Following the procedure outlined in Section 7.2, an overlap matrix was constructed based on the set of analyses provided in both SModelS and MadAnalysis 5. From the SModelS database, a binary acceptance matrix (BAM) containing 393 SRs was generated using a 1% maximum overlap threshold. The initial test of the TACO formalism involved comparing the combined results to the validation plots for analyses and topologies in the SModelS database. These analyses were selected to ensure consistency within a model space thoroughly mapped and understood by SModelS. The first simplified model examined was the "T1" topology, representing a simplified gluino pair-production scenario where each gluino undergoes a three-body decay. Figure 7.3 shows the corresponding Feynman diagram with $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$, producing a light-flavor quark-antiquark pair along with the lightest stable particle (LSP) $\tilde{\chi}_1^0$.



(a) Expected **(b)** Observed

Figure 7.4.: Validation plots comparing the (a) expected and (b) observed results of the TACO SR-combination against three individual-analysis limits from the SModelS results database for the T1 topology. The lines represent the exclusion limits based on r-values, where $r = 1$ corresponds to the 95% confidence-level exclusion boundary. The CMS-SUS-19-006-ma5 analysis is labelled in this way because its efficiency map was derived for SModelS using MadAnalysis 5 rather than directly from the experimental analysis data. The grey dashed line marks the edge of the efficiency maps.

Figure 7.4 shows the validation plots for the T1 topology, illustrating the (a) expected and (b) observed exclusion limits for the combined signal regions (SRs) compared to three individual analyses. The contour lines represented the exclusion limits in terms of the ratio between the predicted cross-section σ_{pred} and the upper limit on that cross-section σ_{UL}, expressed as $r \equiv \sigma_{\text{pred}}/\sigma_{\text{UL}}$, where $r = 1$ marked the exclusion limit at a 95% confidence level. The left-hand plot shows the theoretical expected improvement in exclusion gained by the optimised combinations at each mass-parameter point. The right-hand plot shows the actual improvement using the experimental (observed) data. A total of 265 SRs were applied to the T1 topology. At each model point, the number of relevant SRs was determined by identifying those with efficiency maps covering the parameter ranges of the given model point. In SModelS (v2.1), this was done by first decomposing a given new physics model into its SMS components, representing distinct production and decay processes. These components were matched against experimental results, which provide efficiency maps that describe how well different signal regions of a detector respond to specific event topologies. For a given model point, SModelS checks which signal regions have been covered by these efficiency maps and evaluates



(a) Initial node in each combination **(b)** Number of SRs in combination

Figure 7.5.: Fractional distributions of (a) starting SR-index and (b) number of SRs in each combination. Data was drawn from the optimal SR combination identified for each model point in the T1 model space. As described in Section 7.3, the nodes were constructed from an ordered set of SRs optimised for each point in the model space. Plot taken from the TACO paper [194]

their impact on the model's parameter space. In normal operation, when available, SModelS utilises covariance matrices to combine multiple signal regions in a statistically robust manner, ensuring that the best-constraining regions are identified for exclusion or validation of the model [81, 234]. However, for our purposes, this combination procedure was suppressed in favour of our combination criterion.

Once the relevant SRs had been identified, they were ranked based on the expected upper limit (UL) on the predicted yield (luminosity $\times$ cross-section $\times$ efficiency) for each SR. This process of selecting and ranking the signal regions was incorporated into the BAM, after which the WHDFS SR-selection algorithm was implemented. As shown in Figure 7.4, the combined results pushed the exclusion line approximately 150 GeV beyond the best individual analysis available in the SModelS database.

Upon further examination of the combined results, Figure 7.5 (a) displays the distribution of starting (i.e., lowest) SR indices over the set of maximum-sensitivity combinations. This supported the statements made in Section 6.2, which suggested that the efficiency of the path-finding process was greatly enhanced by sorting the BAM by individual SR sensitivities. The histogram indicated that, when ordered in this way, the optimal combination typically appeared early in the iteration process, allowing many later path sets to be excluded when it became clear that they could not outperform the current best

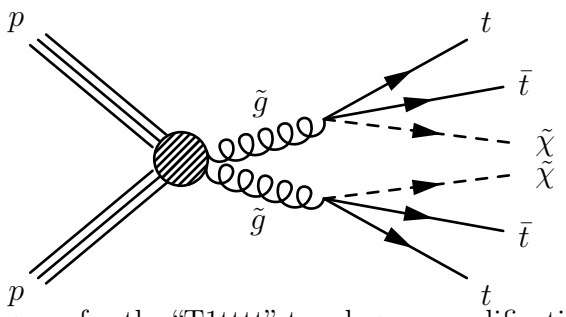


Figure 7.6.: Feynman diagram for the "T1tttt" topology, a modification of the T1 model in gluino decays exclusively into top quark–antiquark pairs and neutralinos $\tilde{\chi}$.

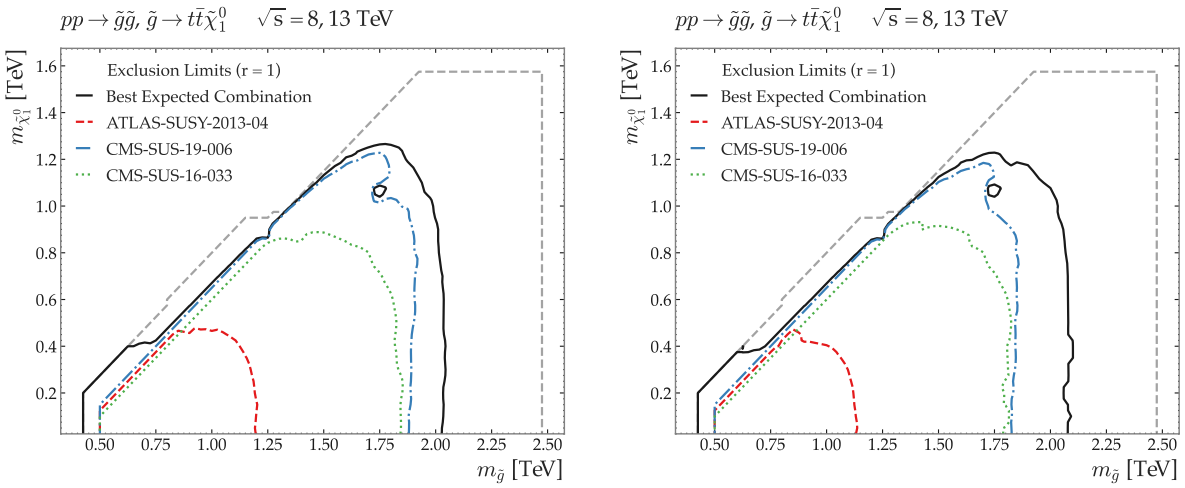
set. The right-hand plot of Figure 7.5 showed the percentage prevalence of the number of SRs in each optimal-sensitivity combination, with typically 6–10 of the available 265 SRs being utilised. This small number conveniently allowed for expensive statistical methods, such as coherent profiling or marginalisation of systematic uncertainties across analyses, which would have been prohibitively costly over the full set of 265 (and growing) SRs.

7.4.2. T1tttt simplified-model combination

The T1 model represents a minimal scenario where gluinos decay into a pair of jets and a neutralino, making it simpler to analyse due to the lower complexity of final states. In our analysis, this serves as a benchmark for SUSY searches, where current experimental results from LHC exclude gluino masses up to around 2 TeV based on missing transverse energy and jet signatures. Building on the T1 analysis, the T1tttt model introduces more complexity where the gluino decays exclusively into pairs of top quarks and antiquarks ($\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$). Figure 7.6 shows the corresponding Feynman diagram.

Including top quarks in the final state significantly increases the complexity of the analysis, necessitating advanced techniques for accurate event identification and reconstruction. This added complexity stems from various characteristics of top quark decays. High-momentum, or "boosted," top quarks produce decay products—jets and leptons—that tend to cluster closely within the detector, requiring the application of complex jet substructure methods and boosted object tagging techniques to distinguish individual components within these densely populated regions.

Additionally, the top quark almost invariably decays via the b-quark, resulting in multiple b-jets in the final state. Precisely identifying these b-jets relies on b-tagging algorithms that detect distinguishing features, such as displaced vertices associated with the relatively long-lived B-mesons. The presence of multiple b-jets further complicates



(a) Expected **(b)** Observed

Figure 7.7.: Validation plots comparing (a) the expected and (b) the observed outcomes of the combination results with three individual analyses available in the SModelS results database for the T1tttt topology. The lines display the exclusion limits in terms of the r-value, where $r = 1$ corresponds to the 95% confidence level. The dashed-grey line marks the boundary of the efficiency maps. Plot taken from the TACO paper [194].

the analysis, as each b-jet must be accurately associated with its parent particle to enable a correct reconstruction of the event.

Furthermore, the complex decay chain of top quarks frequently generates leptons and missing transverse energy due to undetectable neutrinos, which adds additional analytical challenges. Accounting for these leptons and missing energy requires precise calibration and robust background rejection methods to infer the missing components and differentiate top quark signals from background processes. Despite these complexities, experimental searches using SModelS v2.1 have imposed stringent constraints on this model, excluding gluino masses up to approximately 1.8 TeV, depending on the mass of the neutralino.

Figure 7.7 presents the results for T1tttt topology, with (a) the validation plot for the expected and (b) the observed outcomes, once again comparing the exclusion ranges of the combined signal regions (SRs) with those from three individual analyses. The plot construction followed the same approach as in Figure 7.4. Similarly, the primary contribution to the T1tttt combination came from the CMS-SUS-19-006 analysis [213] (depicted by the blue dot-dashed exclusion line). As before, there was a notable expansion of the 95% exclusion contour when compared to the individual SR results, with the

combination smoothing out the particular weakness of the most constraining analysis near $m_{\tilde{\chi}_1^0} \sim 1.1\,\text{TeV}$.

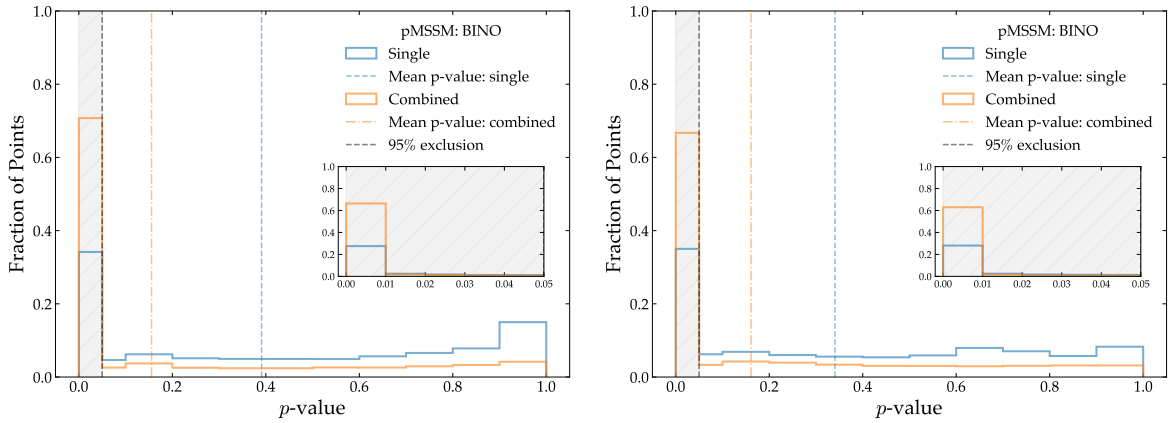
7.4.3. pMSSM-19 reinterpretation

With the machinery previously established to construct the BAM from signal regions (SRs) based on a specified model point, the analysis was expanded to encompass more complex models. The 19-parameter phenomenological Minimal Supersymmetric Standard Model (pMSSM-19) was selected as a case study for reinterpretation due to its significantly larger degrees of freedom than those typically considered in experimental publications. Data points were drawn from the ATLAS 2015 pMSSM-19 scan paper [235], independently for the two scenarios explored in that work: bino-like and wino-like lightest supersymmetric particles (LSPs). ATLAS had classified these points according to their viability against 8 TeV ATLAS data, providing an *a priori* valuable set for subsequent re-evaluation against 13 TeV LHC data.

With a high-dimensional model, the analysis had to move away from exclusion contours expressed in 2-dimensions and consider the data behaviour in a generalised manner. To do this, the p-values from the analyses of 13 TeV LHC Run-2 data were computed for the first (randomly ordered) 27,000 model points in both the bino and wino scenarios separately. These points were processed through the SModelS analysis chain, with those outside the bounds of the SModelS efficiency maps discarded, leaving approximately 20,000 points in each run.

The test statistic ω_i (Equation (7.7)) was calculated using the SModelS best-single-expected signal region selection and best-expected-combination procedures. The resulting p-value distributions from the pMSSM-19 bino-LSP reinterpretation are shown in Figure 7.8. The histograms revealed that in both the (a) expected and (b) observed cases, the combination method shifted a substantial fraction of points from below the 95% exclusion limit into the excluded category. This resulted in an increase in the exclusion fraction from approximately 35% to 70% of all points in the ATLAS set of bino-like models.

This shift in the mean from single to combined could have been interpreted as evidence of the exclusionary capacity of the TACO approach. However, before drawing any conclusions, it was essential to carefully examine how the model points behaved at the bin-to-bin level. While the only definitive statement that can be made from a p-value



(a) Expected **(b)** Observed

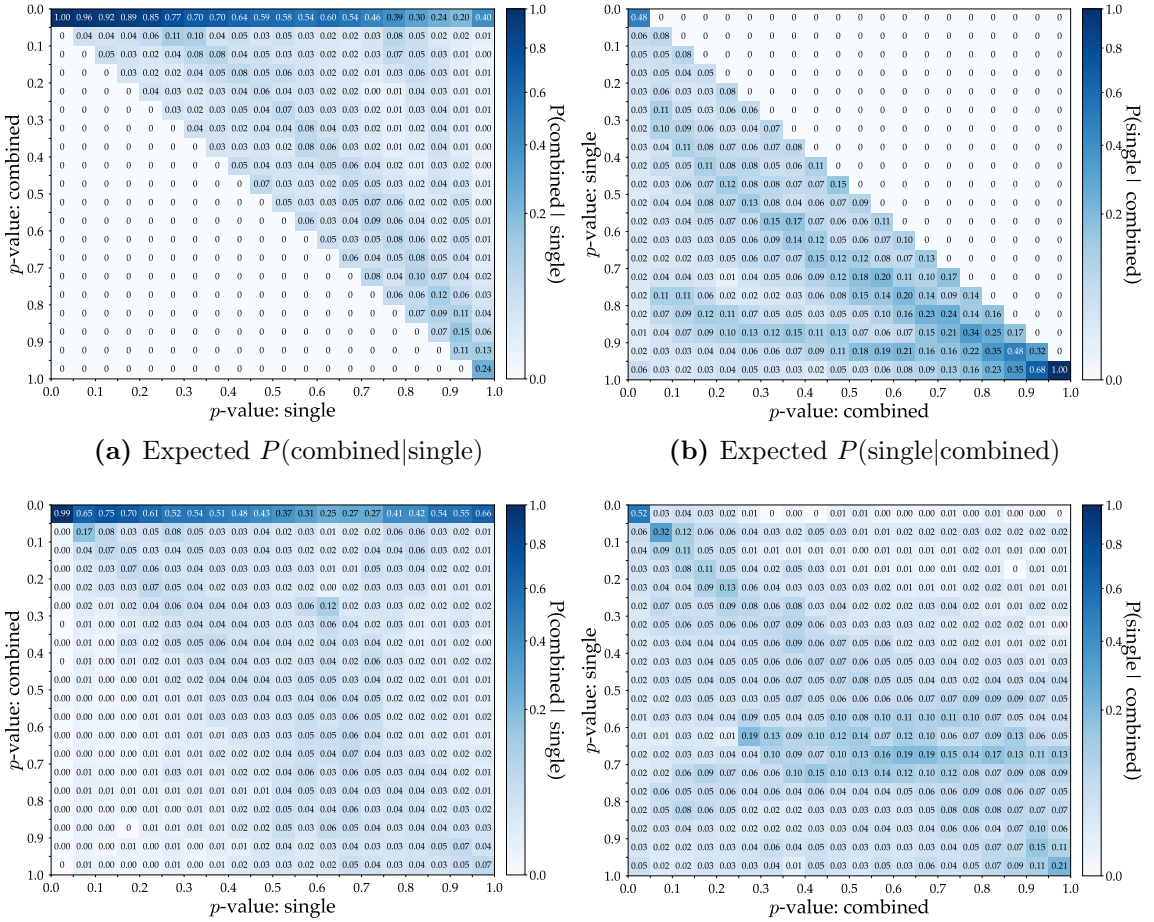
Figure 7.8.: Results from the pMSSM-19 bino reinterpretation using the TACO combination method. p-values were calculated from a selection of 22 000 points taken from the ATLAS pMSSM-19 data set. The blue and orange dashed lines show the mean p-values for the single and combined results. The histograms show that in both the (a) expected and (b) observed cases, a large fraction of points are moved beyond the 95% exclusion limit by WHDFS SR-combination

is whether the model is excluded, additional insights can be gained by examining how combining results affects the population of model points.

To achieve this, we used transition matrices, also known as stochastic matrices. These are routinely used in modelling Markov processes, where the future state of a system depends only on its current state and not on the sequence of events that preceded it [236]. A transition matrix is a square matrix in which each entry represents the probability of transitioning from one state to another. For a discrete-time Markov chain with n states, the transition matrix P is an $n \times n$ matrix, where each element p_{ij} represents the probability of transitioning from state i to state j, such that $\sum_{j=1}^{n} p_{ij} = 1$ for all i, ensuring the matrix is stochastic. The rows of P must sum to 1, reflecting that the system must transition to some state in the next time step.

Mathematically, suppose the system's current state is represented by a probability vector $\mathbf{x}_t$, where each entry corresponds to the probability of the system being in a particular state at time t. In that case, the state of the system at the next time step is given by $\mathbf{x}_{t+1} = P\mathbf{x}_t$. This iterative process describes how the system evolves [236].

To interrogate the pMSSM results on a bin-by-bin level, we repurposed the transition matrices. Figure 7.9 presents the matrices for the pMSSM-19 bino dataset, which depicted the probability of a model point "transitioning" between p-value bins based on whether



(a) Expected $P(\text{combined}|\text{single})$ **(b)** Expected $P(\text{single}|\text{combined})$

(c) Observed $P(\text{combined}|\text{single})$ **(d)** Observed $P(\text{single}|\text{combined})$

Figure 7.9.: Transition matrices showing the pMSSM-19 bino outcomes. These matrices represented the likelihood that a model point shifted from one p-value bin to another, using the SR-combination approach instead of the more conservative single-best-SR method employed by prior recasting tools. The columns of each subfigure separated the transition behaviour based on the movement to combined p-value distributions, given the performance in individual SRs and the origins of each combined-SR p-value range in the individual SR outcomes. The top row of subfigures presented the expected transitions, while the bottom row displayed the observed transitions.

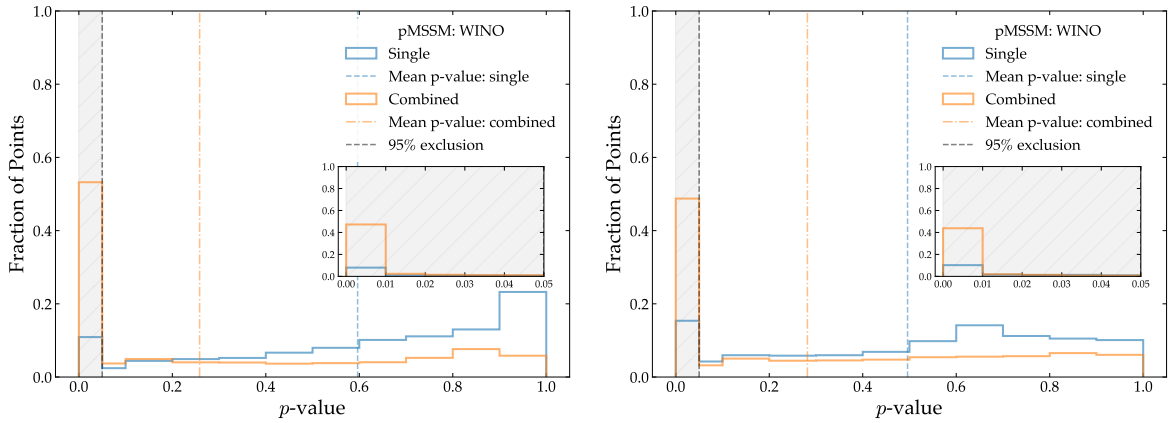
the single-SR or combined-SR LLR construction method had been applied. This could either be viewed as the probability of points in a specific single-SR bin transitioning to various combined-SR bins or as the "origin distribution" of points that ended up in a given combined-SR p-value bin. Both perspectives were informative and displayed in the subfigures' left- and right-hand columns, respectively, with expected and observed outcomes shown in the corresponding rows. Since the values in each matrix represent

probabilities $P(\text{row} \mid \text{column})$, the sum across the column values equals 1 (the transpose of the usual transition matrix).

In this completed project, we began by analysing the expected $P(\text{combined}|\text{single})$ result, which is presented in the top left of Figure 7.9 (a). This result demonstrates how the TACO combination changes the p-values of model points based on their initial p-value obtained from the single best-expected SR. The overall structure of the transition pattern was consistent with the use of SR combinations to *enhance* exclusionary power. Therefore, it was expected that the transitions would predominantly lie in the upper triangle of the $P(\text{combined}|\text{single})$ matrix. Notably, the transitions were largely characterised by shifts into the excluded $p \in [0, 0.05)$ bin, originating not only from neighbouring "nearly excluded" single-SR bins but also spanning a wide range of single-SR p-values. This demonstrated that 40% of the least excluded single-SR points could be excluded when independent SR combinations were employed. Additionally, sub-leading transition patterns were observed in the matrix, indicating below-threshold increases in exclusion that could potentially exceed the threshold with more analyses. The expected $P(\text{single}|\text{combined})$ results, shown in Figure 7.9 (b), were concentrated in the lower triangle of the matrix, as anticipated, and also exhibited similar structures in the transition pattern.

Turning to the observed case in the lower plots of Figure 7.9, the findings became more intricate as the transition distribution became dilute. The prominent transition into the exclusion bin, highlighted in plot (a), was reproduced in the observed case shown in plot (c), as anticipated from the histogram results. The "negative transition" of model points, which moved opposite to what was expected, was attributed to over-fluctuations in the observed yields of the SRs used to compute the combined result. This appeared to be a statistical artefact, likely arising from the combination of multiple SRs, though it only occurred in a small percentage of cases (approximately 5%). Referring back to Figure 7.8 (b), the percentage of points in the exclusion bin jumped from around 35% to over 90% when using the combined SRs. Therefore, the negative transitions seen in plot (d) of Figure 7.9 represented only a small portion of the model points.

Figure 7.10 presents the pMSSM-19 wino-LSP reinterpretation results. When compared to the bino results in Figure 7.8, a similar overall shift toward the exclusion bins was observed, with migration into the 95%-exclusion bin increasing from approximately 15% (single) to over 50% (combined) in both the expected and observed cases. Notably, the wino scenario exhibited a larger fraction of expected single-SR points with p-values greater than 0.3, indicating that a larger population of points at moderate and high-p had the potential to be improved through SR combination.



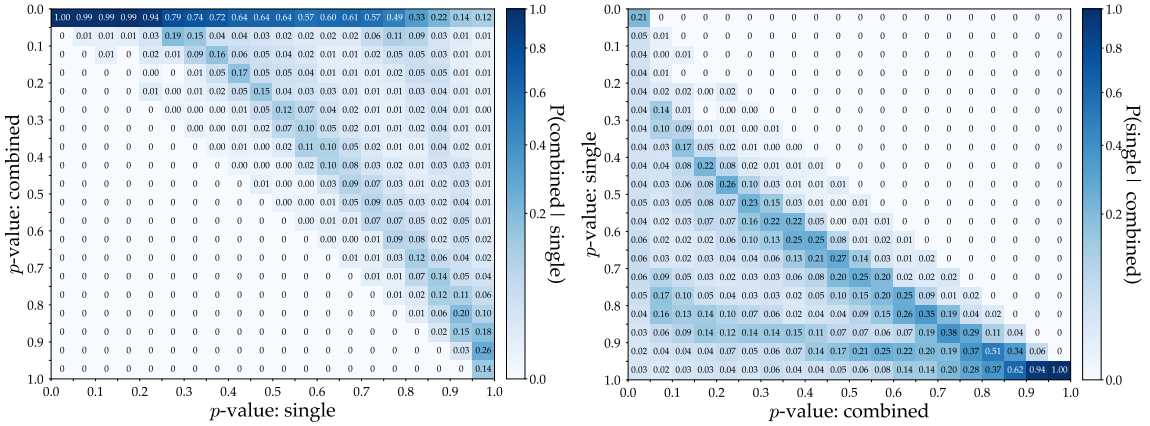
(a) Expected **(b)** Observed

Figure 7.10.: Results from the pMSSM-19 wino reinterpretation were obtained using the TACO combination method. p-values were derived from a sample of 20 000 points selected from the pMSSM-19 dataset. The blue and orange dashed lines represent the average p-values for the individual and combined results, respectively. The histograms illustrate that, in both the (a) expected and (b) observed scenarios, a significant portion of the points was shifted beyond the 95% exclusion threshold through the WHDFS SR-combination method.

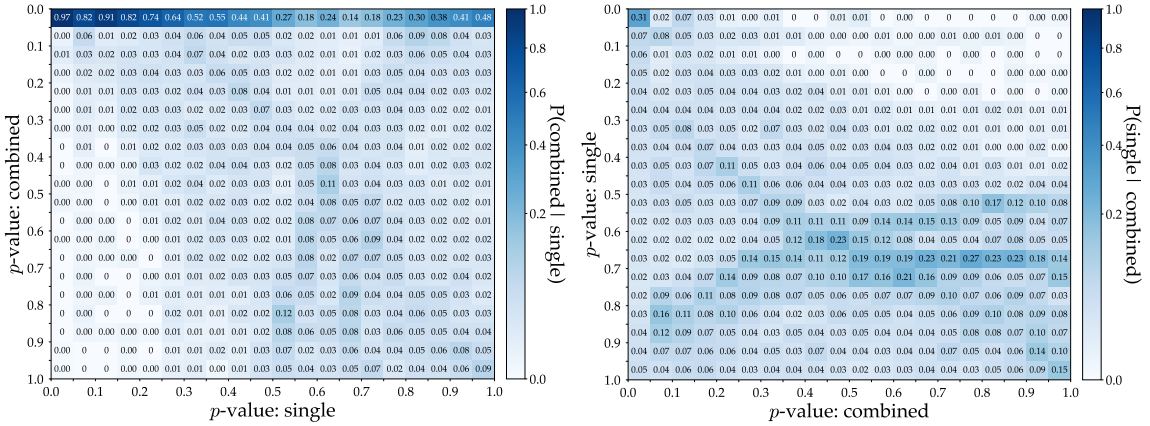
Upon reviewing the transition matrices in Figure 7.11, similar trends to the bino case were identified. The $P(\text{combined}|\text{single})$ plots highlighted the clear shift into the exclusion bin, as shown in the 1D histogram, and the observed plots again showed some degree of negative transitions. However, these were insufficient to increase the overall population of points in the higher-p bins.

The expected transitions into the 95%-excluded p-value bin extended less far along the single-SR p-value spectrum compared to the bino-LSP scenario. Only 12% of the least-excluded single-SR points (those with single-SR $p > 0.95$) were predicted to transition into the combined-SR exclusion bin. In practice, as observed in the yield plots, over-fluctuations in the SR yields resulted in greater exclusion than anticipated for weakly constrained single-SR model points, with nearly 50% of the least-excluded points ultimately being ruled out in combination.

The filament structures observed in both the expected and observed transition plots were more pronounced in the wino scenario, allowing clearer identification of their origins. One of these structures, represented by the shallower lower line in subfigure (b), was associated with the single-SR peak structures at approximately 0.95 for the expected case and 0.65 for the observed case. The other structure followed the main migration trend,



(a) Expected $P(\text{combined}|\text{single})$ **(b)** Expected $P(\text{single}|\text{combined})$

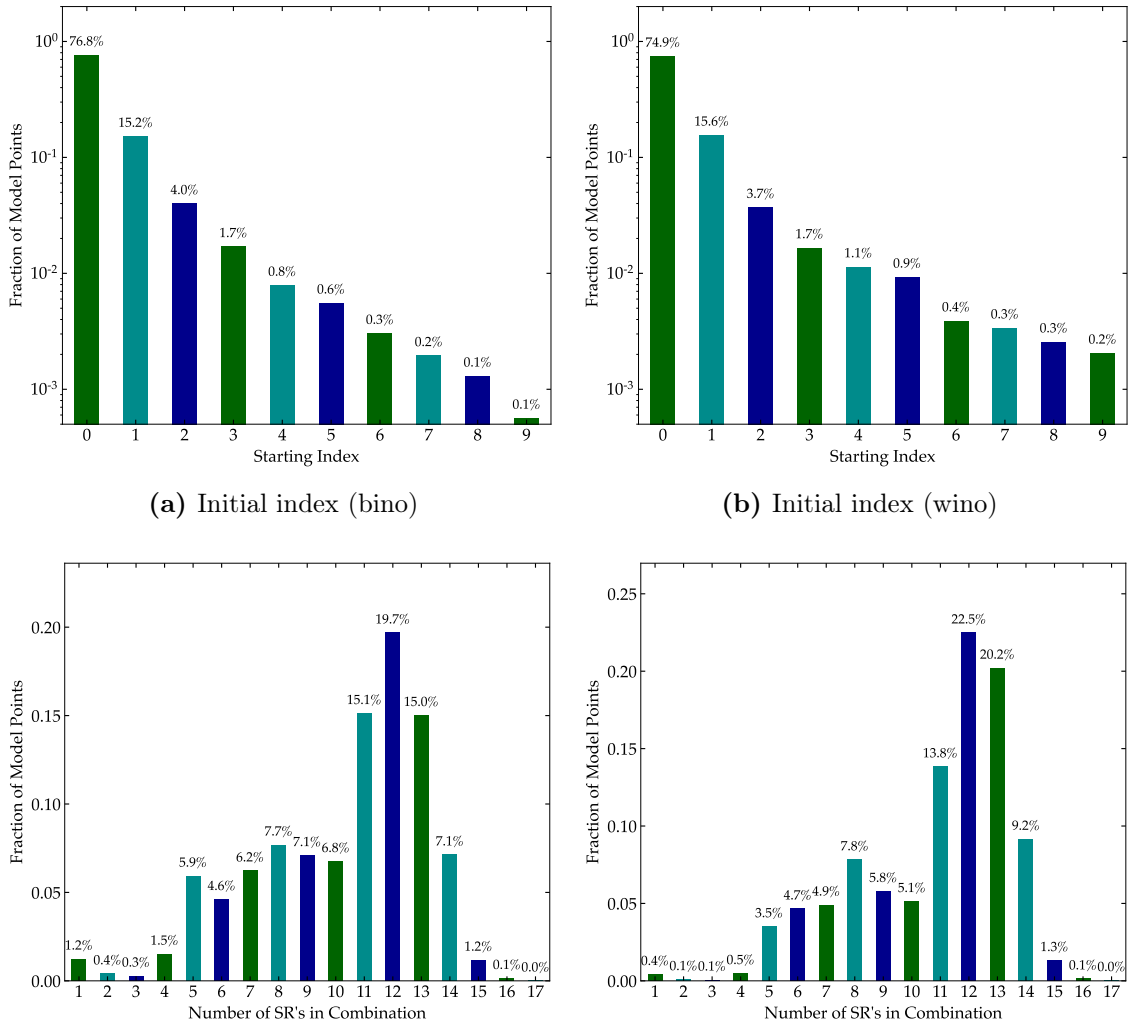


(c) Observed $P(\text{combined}|\text{single})$ **(d)** Observed $P(\text{single}|\text{combined})$

Figure 7.11.: Transition matrices showing the pMSSM-19 wino findings. These matrices represented the likelihood that a model point shifted from one p-value bin to another, utilising the SR-combination approach instead of the more conservative single-best-SR method commonly employed by recasting tools. The columns in the sub-figures separated the transition patterns based on the movement into combined p-value distributions, given the performance in individual SRs, and the origins from single-SR results for each range of combined-SR p-values. The upper and lower rows of the sub-figures presented the predicted and actual transitions, respectively

indicating that the dominant contribution to a given combined p-value bin came from a single-SR p-value bin that was 0.2 units higher. These transition patterns underscored the potential for further improvements in the model-point exclusion fraction, given the availability of additional SRs.

As in the interpretation of the simplified model from Section 7.4.2, the performance of the WHDFS SR-combination algorithm was evaluated for both the bino- and wino-LSP



(a) Initial index (bino)

(b) Initial index (wino)

(c) Number of SRs in combination (bino)

(d) Number of SRs in combination (wino)

Figure 7.12.: Fractional distributions of starting-SR (a, b) and number of SRs (c, d) in each combination. Data is taken from the optimum SR-combination found for each model-point for both the pMSSM-19 bino and wino reinterpretations. As mentioned in Section 7.3, the combinations are constructed from a differently ordered set of SRs for each point in the model space, so the identity of the zeroth SR is free to change from point to point.

pMSSM-19 reinterpretations. The initial SR index distributions were presented in the top row of Figure 7.12, exhibiting a similar bias toward low starting indices as observed in Figure 7.5.

The bottom two plots of Figure 7.12 illustrated the distribution of the number of SRs per combination for (c) the bino and (d) the wino reinterpretations. The rapid decrease in this distribution was attributed to the hereditary condition, which progressively

reduced the number of available SRs with each iteration of the path-building process. Consequently, a critical point was reached, after which the cumulative reduction of available SRs became statistically noticeable in the average number of SRs. In the pMSSM-19 bino and wino analyses, this critical threshold occurred at around 11 SRs.

7.4.4. *t*-channel dark-matter

As a final demonstration of the effectiveness of the TACO approach, we examined one of the t-channel dark matter models investigated in Refs. [237, 238]. In this model, the Standard Model was extended to include a fermionic dark-matter candidate χ and a scalar mediator Y, both of which interacted with the right-handed up-quark. The model's Lagrangian is given by

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{kin}} + \Big[y\big(\chi u_R\big)\ Y^\dagger + \text{H.c.} \Big] , \tag{7.15}$$

where $\mathcal{L}_{\text{SM}}$ represented the Standard Model Lagrangian, $\mathcal{L}_{\text{kin}}$ contained the kinetic and mass terms for the new particles, and y described the strength of the interaction between the mediator, dark matter, and up-quark. In this model, the *full* new-physics signal involved three components:

1. direct production of dark matter with a hard jet from initial-state radiation ($pp \to \chi\chi j$);
2. on-shell production of mediator pairs followed by their decay into dark matter and jets ($pp \to YY^* \to \chi j \chi j$);
3. and the associated production of a mediator decaying into χj and a dark-matter state ($pp \to \chi Y \to \chi(\chi j)$).

This signal could be sought through analyses that targeted multiple jets and missing transverse energy, with each signal component yielding distinct jet multiplicities and properties. We focused on reinterpreting the results from the ATLAS-SUSY-2015-06 [200], ATLAS-SUSY-2016-07 [201], CMS-SUS-16-033 [209], and CMS-SUS-19-006 [213] analyses to determine which mediator and dark-matter mass combinations were consistent with the data, with the new-physics coupling fixed at $\mu = 1$. All the analyses were incorporated into the MADANALYSIS 5 Public Analysis Database [239], and the corresponding recast

codes, along with validation notes, were made available through Refs. [240–244] and the database web page.[1]

To estimate the exclusion limits from each analysis, we employed MadGraph5 v2.6.5 [218] to generate leading-order (LO) hard-scattering events. The signal contributions were grouped into two sets based on jet multiplicity at the parton level. The first matrix element described dark-matter pair production with a hard jet ($pp \to \chi\chi j$), while the second concerned mediator pair production and decay ($pp \to YY^* \to \chi j \chi j$).

The associated production of a dark-matter state with a mediator was included in the first subprocess, as it led to the same final state ($pp \to \chi Y \to \chi(\chi j)$ with the on-shell mediator Y). This subprocess captured scenarios where a mediator produced in association with a dark matter particle decays within the detector acceptance. These events contribute to the dark matter search channels, specifically those characterised by missing transverse momentum. Neglecting this contribution would underestimate the signal yield and, therefore, the potential for excluding model points for cases where the mediator mass is relatively low, as these configurations become kinematically feasible.

These matrix elements were combined with the NNPDF 2.3 LO [219] parton distribution functions, generating 200,000 signal events per model point, reducing the statistical uncertainty in signal yield predictions. In exclusion studies, the precision of signal yield estimates directly impacts the confidence level of exclusion limits, particularly when the signal-to-background ratio is small. Generating 200,000 events ensures a sufficient sample size across mass-parameter space, enabling a robust estimate of the exclusion potential.

Hadronization and parton showering were handled by Pythia 8 v8.240 [245], which simulates the evolution of partonic final states into observable hadrons and the subsequent particle showers, accurately modelling particle interactions as they appear in collider data. The parton showering and hadronisation process contributes significantly to jet formation, which is an important component in searches that rely on identifying missing transverse momentum and jet multiplicities, as it reflects the underlying partonic activity and mediator decay structure.

The CMS and ATLAS detector responses were simulated using Delphes 3 [183], with custom detector parameterisations from each recast code. These simulations account for detector-specific resolutions, efficiencies, and acceptance effects, translating parton-level predictions to observed event signatures. Employing custom parameterisations ensures

[1] See http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase.

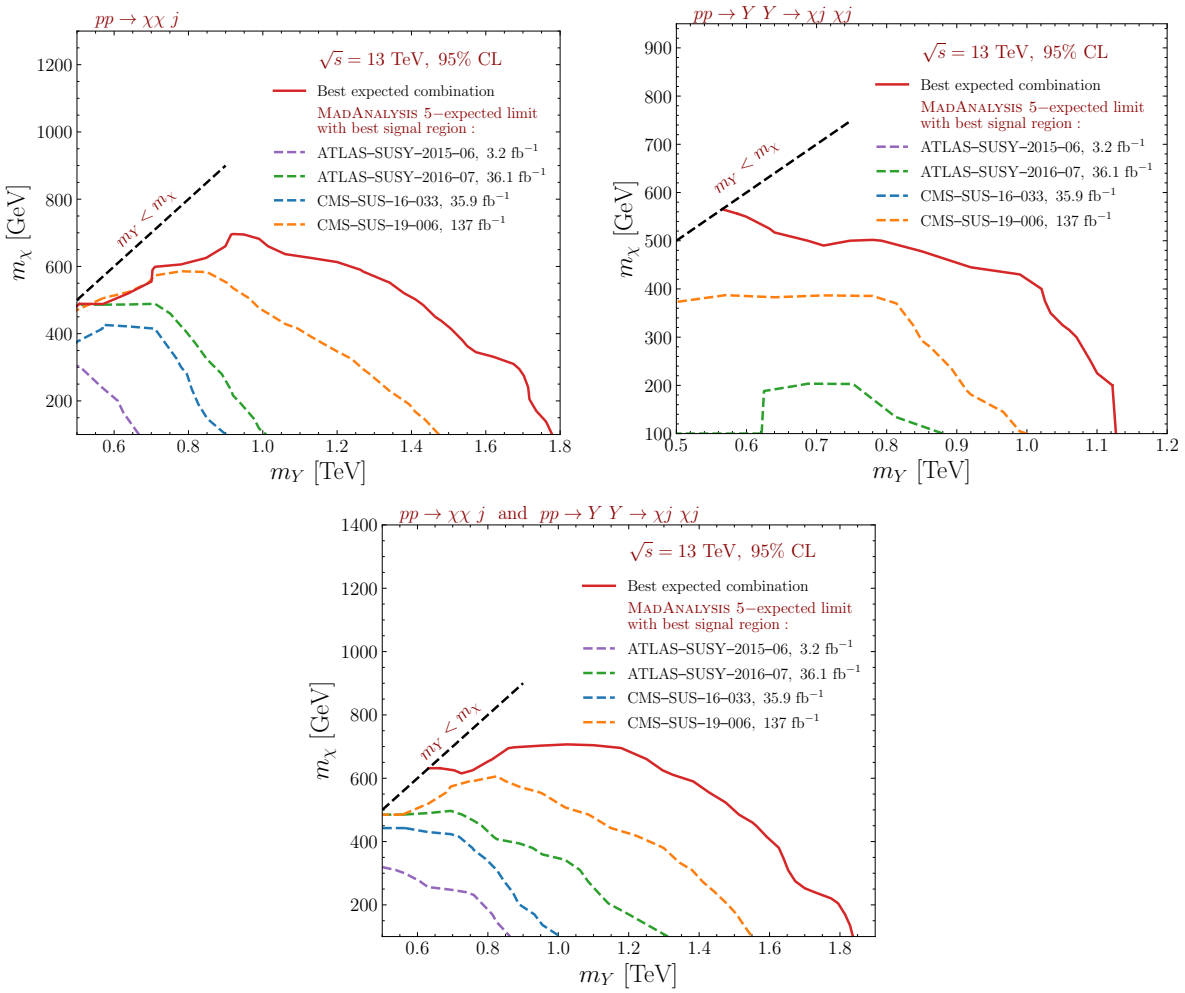


Figure 7.13.: Exclusions at 95% confidence level for the studied t-channel dark matter model. The exclusions are shown separately for the processes $pp \to \chi\chi j$ (upper left) and $pp \to YY^* \to \chi j \chi j$ (upper right), and their sum (lower panel). Dashed lines depict limits from the individual analyses: ATLAS-SUSY-2015-06 (purple), ATLAS-SUSY-2016-07 (green), CMS-SUS-16-033 (blue), and CMS-SUS-19-006 (orange). The solid red line represents the combined exclusion derived by our method.

that the detector response reflects the known effects of each experiment's geometry and technology, thus enhancing the accuracy of the exclusion limits derived from this analysis.

The results, presented in the (m_Y, m_χ) plane in Figure 7.13, displayed exclusions for mediator masses $m_Y \in [0.5, 1.8]$ TeV and dark-matter masses $m_\chi \in [0.1, 1]$ TeV. Exclusions are shown separately for single-jet events ($pp \to \chi\chi j$, upper left), dijet events ($pp \to YY^* \to \chi j \chi j$, upper right), and the combined limits from the full new-physics

signal (lower panel). In each panel, the dashed lines corresponded to individual exclusions derived from the ATLAS-SUSY-2015-06 (purple), ATLAS-SUSY-2016-07 (green), CMS-SUS-16-033 (blue), and CMS-SUS-19-006 (orange) analyses, which respectively probed $3.2\,\text{fb}^{-1}$, $36.1\,\text{fb}^{-1}$, $35.9\,\text{fb}^{-1}$, and $137\,\text{fb}^{-1}$ of data. These limits were obtained by selecting the signal region with the strongest expected exclusion for each benchmark.

Not surprisingly, the CMS-SUS-19-006 analysis, which utilised the largest data set, provided the strongest individual exclusion. For single-jet events (Figure 7.13, upper left), mediator masses up to 1.5 TeV were excluded for small dark-matter masses m_χ. By comparison, the ATLAS-SUSY-2016-07 and CMS-SUS-16-033 analyses, which used about one-third of the Run 2 data, only excluded mediator masses below 900 GeV to 1000 GeV for the same m_χ values. More compressed spectra, which are harder to detect due to softer final states, were also better constrained by the CMS-SUS-19-006 analysis. In contrast, the early Run 2 ATLAS-SUSY-2015-06 analysis only reached exclusions around 500 GeV for new-physics masses.

A similar pattern was observed for the dijet component (Figure 7.13, upper right). The CMS-SUS-19-006 analysis was more sensitive to both larger masses and compressed spectra compared to the earlier analyses, which showed little sensitivity. The reduced sensitivity in this case was due to phase-space suppression from the production of two heavy mediators. As a result, the combined exclusion limits for the full new-physics signal (Figure 7.13, lower panel) were nearly identical to those from the single-jet scenario, with an improvement of about 100 GeV for light dark-matter masses. This demonstrated that considering the full signal yielded better limits than an approximate treatment.

Figure 7.13 also illustrates the effect of combining the four analyses, both for the individual signal components (upper panels) and after combining them (lower panel). By determining the overlaps between the analyses' signal regions (see Sections 7.2 and 7.3), we were able to perform a combination despite the analyses targeting similar topologies (multijet and missing energy). This demonstrated the advantage of a quantified measure of overlap over informal estimates of orthogonality. The combination allowed for improved parameter space coverage, as shown by the solid red lines in Figure 7.13. This combination significantly improved, especially for split-mass spectra (light dark-matter, heavy mediator) and compressed spectra. For $m_\chi \approx 100\,\text{GeV}$, mediator masses up to 1.9 TeV are reachable, whereas scenarios with a dark-matter mass $m_\chi < 600$ GeV get excluded from $Y < 1.2$ TeV

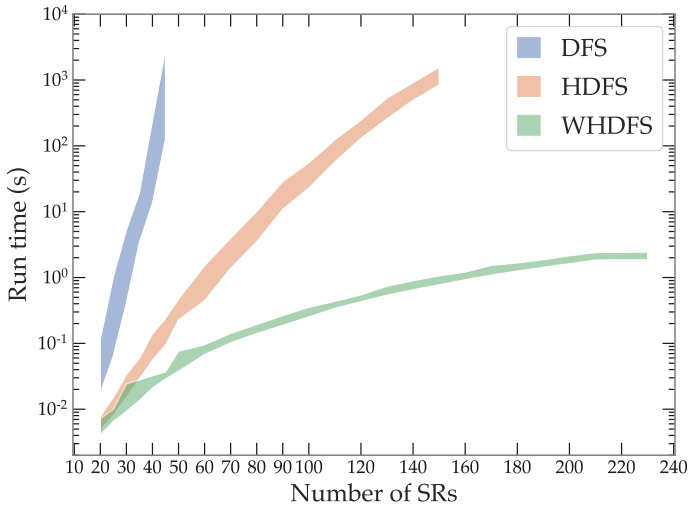


Figure 7.14.: Comparison of CPU-runtime scaling against a number of features between the standard depth-first search (DFS), the hereditary DFS HDFS, and the weighted WHDFS algorithms. The width of the lines was determined by the spread of results taken over 100 trials.

7.4.5. Performance

In BSM scans considering many thousands or possibly millions of model points and hundreds of SRs, the computation of which combination to use for each point must typically be made in seconds, or preferably less. In Section 6.2, we saw that the performance of the HDFS and WHDFS algorithms depended on the dimensionality N and the fraction of allowed combinations f_A. The difference in performance was demonstrated in Figure 6.6, where the number of iterations was used as a proxy for efficiency in finding the optimum path. This relationship could now be tested in a real use-case scenario.

To accurately measure the algorithm performance, 20 mass points from the "T1" simplified gluino-pair with over 240 available SRs were chosen. To generate a range of results, we calculated the optimal combinations for each of a set of reduced element collections $\{SR_0, \ldots, SR_{N-1}\}$ and its corresponding $N \times N$ BAM submatrix. We randomly selected SRs to avoid bias and maintained the original BAM fraction of allowed combinations, $f_A \approx 0.75$. The number of elements $m \leq n$ in the reduced element sets was evaluated from $m = 20$ to 80 in steps of 10 and from $m = 80$ to 140 in steps of 20. The upper limit on m was determined by the requirement to find 20 mass points (for timing-uncertainty estimation) with at least that many supported elements.

Figure 7.14 shows the CPU-performance comparison between the HDFS and WHDFS algorithms. There is an additional baseline comparison using a depth-first search (DFS) algorithm from the PYTHON NetworkX package [193]. The plot clearly shows that the DFS does not scale sufficiently for physics purposes, as it requires hundreds of seconds by 40 SRs, with extremely strong exponential scaling. The HDFS algorithm fares much better, scaling up to 100 SRs with a flatter exponential growth than DFS and with slightly sub-exponential growth thereafter. Although an improvement on DFS, HDFS requires $\mathcal{O}(100)$ seconds for 100 SRs, which is insufficient for most applications that require a best combination only. The further optimisations enabled by the WHDFS formulation show a flatter still scaling exponent, with sub-exponential growth that becomes particularly flat for large element counts. 230 SRs were obtained in around 2 seconds, which is very compatible with adaptive sampling and indicates little issue in scaling further toward thousands of elements. These performance gains indicate the effectiveness of the WHDFS algorithm and its ability to meet the current practical requirements of large-scale analysis combinations.

7.5. TACO conclusions

In this project, we argued that enhancing the BSM search capabilities of the LHC and other colliders, especially in light of the numerous null results from Run 2 direct searches, required combining analyses to improve sensitivity to subtle, dispersed-signal models that had not been fully considered before.

However, such combinations could not be done naively due to overlapping event acceptances across analyses. In the absence of long-term, top-down coordination within experimental collaborations to prevent phase-space overlaps—or public tracking of which collider events contributed to which signal regions across published analyses—a *post hoc* approach was necessary to estimate the degree of overlap. To address this, we developed the TACO method, using the SModelS and MadAnalysis 5 analysis databases to guide the simulated-event population across all recastable signal regions. We introduced a new feature in the MadAnalysis 5 analysis framework to compute overlap coefficients between pairs of signal regions through Poisson bootstrapping.

We also demonstrated how this overlap information could effectively identify the optimal subset of non-overlapping signal regions for excluding a given BSM model or model point. The computationally complex task of evaluating all allowed combinations

of roughly 400 signal regions—expected to increase—was transformed into a directed acyclic graph construction problem. This construction was made efficient by using a binarised overlap matrix, which excluded partially overlapping graphs, and by ordering signal regions based on their expected log-likelihood ratio (LLR). The best combination of signal regions could then be determined through a weighted hereditary depth-first search (WHDFS) algorithm, analogous to solving a weighted longest-path problem. The overlap estimation and the signal region combination code are publicly accessible from the https://gitlab.com/t-a-c-o/taco_code repository.

Using the TACO WHDFS algorithm to compute optimal SR subsets proved a viable alternative to directly applying a correlation matrix for correlated χ^2 or other measures across all signal regions. The latter approach is prone to numerical instabilities in covariance inversion and is computationally costly due to the challenge of profiling likelihoods across the entire set of signal regions. This graph-based method thus offers a promising pathway not only for composite likelihood calculations in reinterpretations but also as a dimension-reduction tool for visualising and understanding how dominant analyses and signal regions evolve across model spaces.

We tested the TACO method for overlap estimation and optimal subset selection using various BSM models of increasing complexity, including a SUSY simplified model, ATLAS' 8 TeV pMSSM-19 scan re-evaluated using 13 TeV data and a t-channel dark matter model. In all cases, we observed that combining SRs algorithmically significantly extended experimental limit-setting reach, typically by $\mathcal{O}(100)$ GeV in mass parameters for both simple and complex BSM models considered with existing reinterpretation analyses. Transition matrices studied for the SR combinations indicated that the improvements were not merely marginal, pushing nearly excluded model points over the threshold but substantial, with combinations of weaker signal regions contributing to a more robust exclusion.

This approach demonstrated that combining BSM direct-search data *post hoc* is feasible and computationally efficient, eliminating the need for conservative strategies that use only one SR per event topology. As a scalable and empirical computational method, TACO can handle hundreds of potentially overlapping analyses, far exceeding the capacity of manual or subjective analysis. Nonetheless, the method has its limitations. The assumption of effective orthogonality for signal regions with overlap coefficients $\rho_{ij} < T$ is critical for computational efficiency. However, this can be addressed by integrating correlated LLR evaluations on the reduced set of SRs. Furthermore, the current method does not account for systematic uncertainties, although uncertainties

accessible through event reweighting can be incorporated into future work. We hope that this method and the corresponding toolkit will serve as a useful reference for how BSM combinations are approached in the future, with analysis routines submitted to major reinterpretation frameworks and the introduction of event-bootstrapping tools beyond MadAnalysis 5.

7.5.1. Retrospective analysis

At the time of writing, it has been over a year since the publication of the *Strength in numbers: Optimal and scalable combination of LHC new-physics searches* (TACO) paper [194]. In that time, there has been ongoing discussion among the authors and the broader LHC reinterpretation community regarding the project's strengths and weaknesses. This section will cover key aspects of these discussions, focusing on the paper's methodological innovations, its applicability to current high-energy physics challenges, and the limitations encountered in its real-world implementation. A notable strength of the TACO approach lies in its ability to provide a scalable framework for combining results from multiple LHC new-physics searches. By leveraging advanced statistical techniques and developing new selection algorithms, the framework optimises the sensitivity to new physics signals while maintaining computational feasibility, even for large-scale datasets. The novelty of this approach has attracted significant attention, particularly in contexts where traditional methods struggle with data scalability.

Control regions

Certain weaknesses have also been identified, particularly concerning the assumptions in calculating the overlap matrix. While necessary for tractability, these assumptions may introduce biases when applied to data sets with varying degrees of correlation. This is particularly true between control regions, a specific subset of the experimental data used to validate or constrain the background predictions in an analysis. They are typically a region of phase space where new physics signals are not expected to appear, making them ideal for understanding and modelling the standard processes that contribute to the background [246].

Control regions are designed to contain mostly well-understood, background-only events, with minimal or no contribution from the hypothetical signal being searched for. They help to fine-tune the background estimation by comparing the predicted

background (usually derived from simulation or theoretical models) with the actual data. This comparison allows physicists to adjust or reweight their models to better reflect reality, thus improving the accuracy of the background prediction in the signal region. The effectiveness of a control region relies on the assumption that the factors influencing the background processes are similar in both the control and signal regions, except for the presence of the signal itself [7].

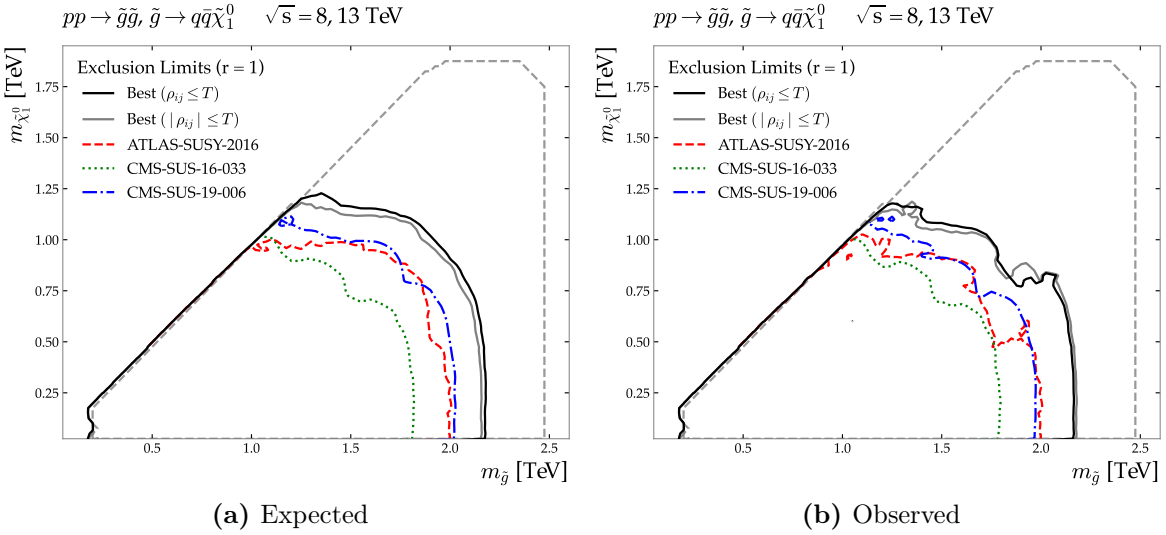
The scenario that caused the most concern was where the control region of one SR physically overlaps an alternative, combinable SR. This situation is theoretically quite realistic as the overlaps are, in essence, calculated via the estimate of shared events, and control regions do not share signal events by design. This would require a model that populated an SR and its corresponding control region; let's call these SR_A and CR_A, with $\mathrm{SR}_B \in \mathrm{CR}_A$. For the case of no signal, the background estimate is unaffected, and the exclusion "power" is unaffected. For the alternative case, the background estimate for SR_A contains signal events resulting in a higher yield expectation, meaning that ω_i is inflated, resulting in an overly conservative estimate.

An alternative approach would be to estimate the overlaps in terms of the SM background. This would improve the overlap estimate, but it would also be extremely computationally expensive. Using only signal events restricted to the SMS topologies, generating the overlap matrix took over two months. As such, the approach balanced traceability and efficiency reasonably well, producing a statistically well-motivated technique for estimating the overlaps between SRs.

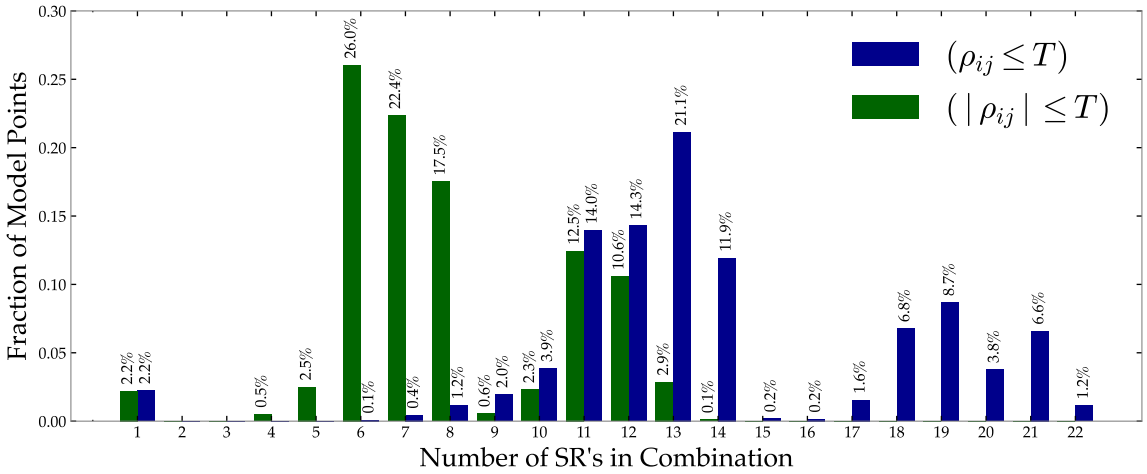
An absolute mistake

When reflecting on the use of the overlap matrix, specifically in terms of creating the BAM via the "arbitrary" threshold T, it was realised that a mistake had been made. When determining the threshold under which to allow combinations, we defined elements of the BAM B as $B_{ij} = (\mid \rho_{ij} \mid \leq T)$. The logic for this definition was that we wanted minimally "overlapping" SRs, and close to zero was minimal. However, upon further consideration of ρ_{ij}, it was realised that negative or "anti-correlated" values would also be suitable for combination. The "anti-correlation" provided information regarding the fluctuations of events that were being discarded by taking the absolute value of ρ_{ij}.

To investigate the effect of redefining the threshold to allow negative values, the analysis of the "T1" topology was redone using the new BAM definition. The first



(a) Expected **(b)** Observed

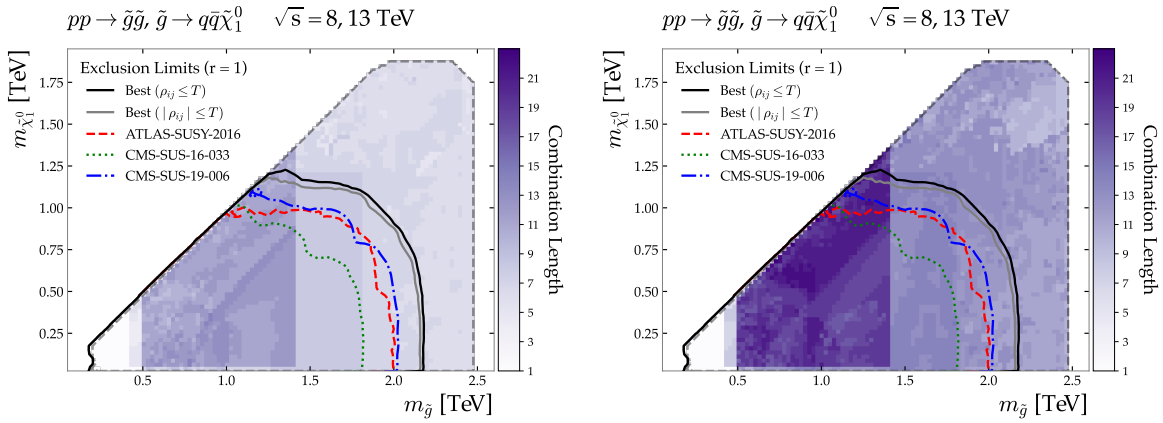


(c) Path Length Comparison

Figure 7.15.: Comparison of results using the updated definition of the BAM. The new results are shown by the black line, with the previous results shown in grey.

result that should be mentioned is that changing the BAM configuration increased the total fraction of allowable combinations f_A (Equation (6.9)) of both ATLAS and CMS analyses at $\sqrt{s} = 8, 13$TeV, from 72% to 85%. Treated separately, ATLAS and CMS at $\sqrt{s} = 13$TeV increased from 36% to 60% and 41% to 69%, respectively. And finally, at $\sqrt{s} = 8$TeV, ATLAS and CMS increased from 25% to 44% and 29% to 58%, respectively. As predicted, this increase in f_A subsequently increased the computation time logarithmically.

Figure 7.15 shows the results in terms of both definitions of the BAM. For the expected case, there is a small but not insignificant increase in exclusion, changing from 58% to

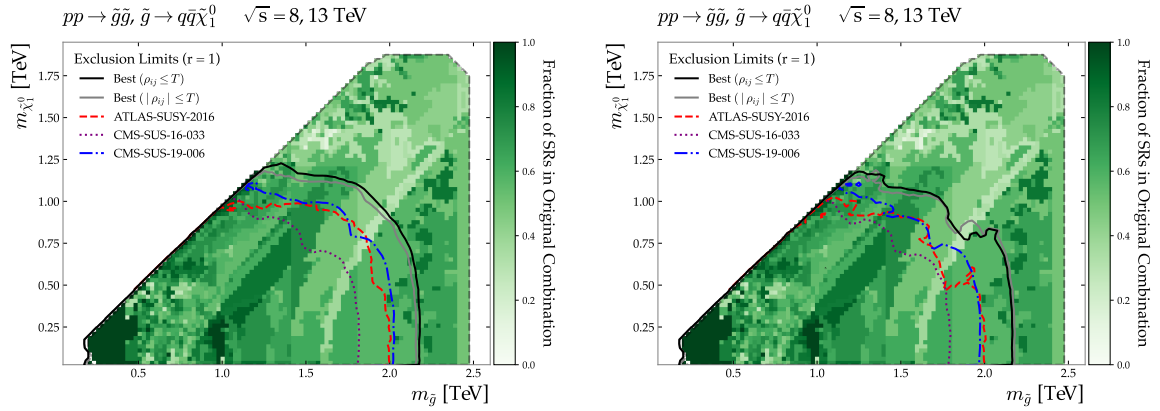


(a) Expected Results $B_{ij} = (|\ \rho_{ij}\ | \leq T)$ **(b)** Expected Results $B_{ij} = (\rho_{ij} \leq T)$

Figure 7.16.: Comparison of results using the updated definition of the BAM, including a heat map of combination lengths for each model point. The left-hand plot shows the original published results, and the right-hand plot shows the length using the new BAM configuration

60% exclusion using the new BAM calculation. For the observed case, adding new SRs made little difference, changing from 56.45% to 56.38%, a slight reduction of 0.07%. The lower plot of Figure 7.15 shows the change in combination length. For the previous iteration, the majority of the combinations were around six or seven SRs long; moving to the new configuration, the number of SRs increases to around 13, with a significant number of points getting around twenty SRs. This result was surprising as one would think that a large increase in the number of SRs per combination would be translated into exclusion "power".

A short comparative analysis was performed to understand why the effect had been so slight. Following from plot *c* of Figure 7.15, Figure 7.16 shows the lengths of each combination as a heat map over the "T1" mass plane. The left-hand plot shows the original published results, and the right-hand plot shows the length using the new BAM configuration. The first and most obvious feature in both plots is the over-representation of data for $m_{\tilde{g}} < 1.45$ TeV. This is due to the free combination of $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV data, which are assumed to be uncorrelated by default. Looking at the upper section of the exclusion line at $m_{\tilde{\chi}_1^0} \approx 1.25$ TeV, the increase in available data provides an extra "bump" in exclusion for both results. The combination length within this region is apparent in the secondary peaks in Figure 7.15, which occur at around twelve SRs for the original data and nineteen for the new. Moving to the upper right



(a) Expected **(b)** Observed

Figure 7.17.: Plots showing the fraction of the original result represented in the new combination. Both plots a include the exclusion ($r = 1$) lines calculated using the updated BAM configuration, with the left-hand plot (a) using expected data and the right-hand plot (b) using observed data.

of each plot, the increase in combination length appears approximately homogeneous, suggesting the additional SRs provided little extra exclusionary "power".

The next point of enquiry was comparing the SRs combination per model point. For a meaningful comparison, the original results were used as a baseline, meaning that for each model point on the "T1" mass plane, a list of SRs contributed to the published result. Each one of the lists was then compared to the new combination in terms of the fraction of the old SR list represented in the new, i.e. what fraction of the original result is still in the new combination. Figure 7.17 shows the results of this analysis, with the left-hand plot (a) showing the expected results and the right-hand plot (b) showing the observed. The combinations were calculated using the expected values; thus, the combinations used in each plot were the same. It was found that the fraction of original SR represented at each model point averaged at 0.60 with a standard deviation of 0.16. For both sets of results, the dominant analyses were ATLAS-SUSY-2016-07 [201] and CMS-SUS-19-006 [213], both of which were searched in final state jets with missing transverse momentum at $\sqrt{s} = 13$TeV. The main difference was the increased representation of individual SR from these analyses.

A visual example of this is the filament structure seen in Figure 7.17 starting at $\approx m_{\tilde{g}} = 1.6$TeV. This band curves up and to the right and shares, on average, 35% of the SRs from the original result. Comparing the SRs within this filament showed that the original result used four to six SRs from CMS-SUS-19-006, whereas the new result

used upwards of twelve. Interestingly, they were almost completely different sets of SRs in each case, only sharing one or two SRs. Looking at plot b, following this filament structure up to approximately $m_{\tilde{g}} = 2$ TeV, we see a loss of exclusion driven by this selection.

The weights (ω_i) introduced by the new SRs revealed that in most cases, the SRs retained in the new combination from the old are the most exclusionary. This makes sense as the WHDFS algorithm would select the SRs with the most "power", and if they were combinable in the original result, they would still be combinable under the new BAM configuration. Thus, we still see the same combinations of analyses, i.e. ATLAS-SUSY-2016-07 and CMS-SUS-19-006, but the new configuration allows for a potentially different mix of SRs from the analyses.

Unfortunately, the additional choice didn't improve the original results to the extent that would warrant a complete recalculation of all results. However, the realisation that the BAM could be configured differently provided an interesting opportunity to investigate the behaviour of the WHDFS algorithm.

Reinterpretation outlook

Looking at the results of this comparative analysis, it is safe to say that the original conclusions of the TACO project remain valid. The main criticism of the project was the method by which the overlaps were calculated and how realistic this approach was. This raises an interesting issue for reinterpretation studies in general: to get the most out of the data available from search analyses like the ones used in this study, there are three options. First, we accept that there will be overlaps in SRs, so for reinterpretation, we only choose the data where we know the correlations. This would limit the data to that for which the analysis provides correlations or that we know is uncorrelated by default (i.e. different energies, time or experiment). The second option is to estimate the correlations via methods like the one presented in the TACO paper. However, as pointed out, this estimation is computationally expensive and biased to be overly conservative. The third option is to design the SRs such that they don't overlap in the first place. This last option is the most idealistic and challenging as it requires the standardisation of SRs across experimental collaborations; however, at the time of writing, consideration is being given to this idea by the reinterpretation community.

Chapter 8.

Anomaly detection: proto-models

Anomaly detection attempts to identify generic unexpected signatures in the data. Traditional BSM searches are generally hypothesis-driven, where signatures of particular BSM scenarios, such as supersymmetry or extra dimensions, are explicitly targeted. However, these approaches can overlook phenomena that do not conform to the predefined signatures. Anomaly detection offers a complementary strategy by leveraging data-driven methods to identify regions in the data that exhibit significant discrepancies compared to SM predictions. For example, by constructing representations of the SM background, often through machine learning or other statistical techniques, and flagging regions with excesses not explained by statistical fluctuations or known physical processes.

This chapter combines the PathFinder package with an anomaly search, reinterpreting search analyses from the ATLAS and CMS experiments. In general, anomaly detection looks for subtle deviations from the SM expectations, which may indicate the presence of new physics without relying on specific BSM models. In addition to this goal, this chapter will demonstrate how the whdfs algorithm provides an ideal selection tool for choosing analyses.

This project builds on a previous study conducted by the SModelS collaboration in 2021. While I was not involved in the original research, I provide an overview of it in Section 8.1. My contribution to this work includes implementing the whdfs algorithm and conducting the statistical analysis required to establish a robust test statistic for anomaly detection. As of the time of writing, the project is still ongoing, and the results presented here are preliminary.

8.1. Proto-models version 1

This project follows from previous research presented in the paper "*Artificial Proto-Modelling: Building Precursors of a Next Standard Model from Simplified Model Results*" [233]. The original paper explored an innovative approach to identifying signs of new BSM physics. The authors proposed a methodology using a Markov Chain Monte Carlo (MCMC)-like "random walk" algorithm to generate proto-models — theoretical frameworks involving new particles. These models are tested against experimental results from the Large Hadron Collider (LHC), specifically using data from the ATLAS and CMS experiments integrated with the SModelS framework. This framework analyses simplified-model results, focusing on regions where deviations from the SM might be significant. By assessing various analyses and signal regions that potentially conflict with the SM while still adhering to existing LHC constraints, the method looked to identify patterns that could suggest new physics.

This section will summarise the original research and discuss pertinent elements carried over to the current project in greater detail. We will start with the definition of a proto-model, move through the free parameters of the MCMC "walk," and finish with the results and conclusions of the original research.

8.1.1. What is a proto-model?

Proto-models are simplified theoretical frameworks that suggest the existence of new particles or interactions beyond the Standard Model (SM). They serve as preliminary versions of more complex theories and are tested against experimental data to evaluate their validity. In the context of the work presented in Chapter 7, which focused on exclusion, proto-models function as alternative hypotheses (H_1) in anomaly detection, contrasting with the Standard Model null hypothesis (H_0). Proto-models are constructed by introducing sets of hypothetical particles with specific parameters–masses decay channels and production cross sections–and then comparing their predictions with results from experiments like those conducted at the LHC. They serve as "prototypes" that can either be discarded if incompatible with experimental constraints or further developed into more comprehensive models of new physics. The approach simplifies the search for new phenomena by focusing on basic components rather than fully developed, intricate theories. The "Proto-Modelling" paper introduces proto-models as collections or "stacks"

of simplified models, where the number of BSM particles, as well as their masses, production cross sections and decay branching ratios, are taken as free parameters [233].

Approximation likelihoods were constructed using simplified-model constraints via the SModelS [76] package, allowing for a MCMC type walk through the parameter space. The MCMC walks refer to randomised "steps" in the parameter space, translating to random changes in the BSM model via the number of particles, their masses, production cross sections, and decay branching ratios. The purpose was to identify proto-models that evaded all available constraints while explaining potential dispersed signals in the data. This ambitious objective did, however, come with some large caveats; the authors emphasised that the proto-models were not intended to be UV complete, nor were they a consistent, effective field theory. With these limitations, the authors admit that the purpose of the analysis was to guide future experimental and theoretical efforts toward the possible construction of the Next Standard Model (NSM)

As the proto-models were not intended to be fully consistent theoretical models, they were, by definition, unbound by higher-level theoretical assumptions, such as representations of the SM gauge groups or higher symmetries. As such, the following constraints were imposed:

- Only consider particles with masses within the reach of the LHC for a specific proto-model.
- All BSM particles are odd under a $\mathcal{Z}_2$-type symmetry, so they are always pair-produced and cascade decay to the lightest state.
- The Lightest BSM Particle (LBP) is stable and electrically and colour neutral, and hence is a dark matter candidate
- Except for the LBP, all particles are assumed to decay promptly.
- Each particle's production cross-section was treated as a free parameter, effectively reducing the degrees of freedom by absorbing the effects of spin and multiplicity.

With these constraints applied, the original proto-models MCMC "walk" was conducted with the following pool of 20 (SUSY-inspired) particles under consideration:

- Light quark partners $X_q(q = u, d, c, s)$: A single partner was allowed for each light-flavour quark. Unlike SUSY models, two independent particles (left and right-handed squarks) for each flavour are not considered.

- Heavy quark partners $X_b^i, X_t^i (i = 1, 2)$: unlike the light quark partners, two independent particles for each flavour were considered. The large amount of available data from LHC searches meant that two states were included to allow enough (degrees of) freedom to accommodate the data.
- Gluon partner X_g: One new colour-octet particle, analogous to a gluino in SUSY.
- Electroweak partners $X_W^i, X_Z^j (i = 1, 2; j = 1, 2, 3)$: Two electrically charged and three neutral states were allowed. These could correspond to charginos and neutralinos in the MSSM (with the neutral higgsinos being exactly mass-degenerate) or to the scalars of an extended Higgs sector with a new conserved parity. The lightest neutral state (X_Z^1) was assumed to be the LBP and, hence, a dark matter candidate.
- Charged lepton partners $X_\ell (\ell = e, \mu, \tau)$: a single partner for each lepton flavour was considered for light-flavour quarks.
- Neutrino partners X_ν: again, one partner for each neutrino flavour (ν_e, ν_μ, ν_τ) was considered.

A full list of particles considered in the proto-modelling paper is given in Table 8.1, summarising the BSM particles and decay channels.

8.1.2. Free parameters

Masses

As previously defined, the lightest BSM particle was required to be the X_Z^1. Thus, all the masses considered in the "walk" had to satisfy the relation $m(X) \geq m(X_Z^1)$. States that exist in two- or three-fold multiplicities were mass-ordered. Only masses below 2.4 TeV were considered, as they were assumed to be within the LHC reach at the time of publication. Some additional mass requirements were necessary to tailor the machinery to the available database; more details can be found in the original publication Ref. [233]

Decay branching ratios

The branching ratios (BRs) of the allowed channels must add up to unity with the decay modes of each new particle consistent with its quantum numbers. The BSM particles and decay channels were restricted to those shown in Table 8.1. Not all configurations were considered in the paper; in particular, X_W and X_Z decay into $X_\ell + \ell/\nu$, $X_\nu + \nu/\ell$, and $X_{W,Z} + \gamma$ were not taken into account. Specific decay modes were turned off if one of the daughter particles was not present in the proto-model or if the decay was kinematically forbidden.

Production cross sections

The initial values of the cross sections were calculated assuming the BSM particles to be MSSM-like (see section 2.2.5). This value could be rescaled freely by signal strength multipliers κ. For instance, the pair production of X_g is given by:

$$\sigma(pp \to X_g X_g) = \kappa_{X_g X_g} \sigma(pp \to \tilde{g}\tilde{g}) \,, \tag{8.1}$$

where the mass of the gluino $\tilde{g}$ is set to the X_g mass. The rescaling factors $\kappa_{X_g X_g}$ were then used as free parameters of the proto-model. SLHA templates were used to define the masses and BRs, with the reference SUSY cross sections computed using PYTHIA 8.2 [172] and NLLFAST 3.1 [247–254].

Particle	Decay Channels	Particle	Decay Channels
X_q	qX_Z^j, $q'X_W^i$, qX_g	X_W^1	WX_Z^i
X_t^1	tX_Z^j, bX_W^i, WX_t^1, tX_g	X_W^2	WX_Z^j, ZX_W^1, hX_W^1
X_b^1	bX_Z^j, tX_W^i, WX_t^1, tX_g	$X_Z^{j\neq 1}$	WX_W^i, $ZX_Z^{k<j}$, $hX_Z^{k<j}$
X_t^2	tX_Z^j, bX_W^i, ZX_t^1, WX_b^1, tX_g	X_l	ℓX_Z^j, νX_W^i
X_b^2	bX_Z^j, tX_W^i, ZX_b^1, WX_t^1, bX_g	$X_{\nu l}$	$\nu_l X_Z^j$, ℓX_W^i
X_g	$q\bar{q}X_Z^i$, $q\bar{q}'X_W^i$, $b\bar{b}X_Z^i$, $t\bar{t}X_Z^j$, btX_W^i, qX_q, bX_b^i, tX_t^i		

Table 8.1.: BSM states and decay channels used for proto-model construction. The contents of this table are in reference to and were taken from Ref. [233]

8.1.3. Likelihoods and constraints

The extent to which the likelihoods could be calculated depended heavily on the information the experimental collaborations provided. As mentioned in Section 7.3.2, the results available in the SModelS database are mainly split into upper-limit (UL) type and efficiency map (EM) type. The likelihood options were dependent on the data available, and starting with the least ideal case, the options were:

1. If only observed ULs were available, then the likelihood became a constraint in the form of a step function at 95% CL.

2. If expected and observed ULs were available, the likelihood was approximated using a truncated Gaussian:

$$\mathcal{L}(\mu \mid \mathrm{D}) = \frac{c}{\sqrt{2\pi}} \frac{\sigma_{\mathrm{ref}}}{\boldsymbol{\sigma}_{\mathrm{obs}}} \exp\left(-\frac{(\mu\sigma_{\mathrm{ref}} - \sigma_{\mathrm{max}})^2}{2\boldsymbol{\sigma}_{\mathrm{obs}}^2} \right), \text{ for } \mu \geq 0\,, \tag{8.2}$$

   where the likelihood is a function of the signal strength parameter μ and the reference cross-section σ_{ref} predicted for a given proto-model for data D and normalising constant c. The value of $\boldsymbol{\sigma}_{\mathrm{obs}}$ was approximated using $\sigma_{\mathrm{exp}}^{\mathrm{UL}}$, and σ_{max} was chosen such that the truncated Gaussian correctly reproduced the 95% CL. Equation (8.2) was, admittedly, a crude approximation. However, it did contain useful information.

3. For EM-type results, a *simplified likelihood* [255] could be used where nuisances (ξ) are introduced and $p(\xi)$ is assumed to follow a Gaussian distribution located at zero with a variance of δ^2, such that:

$$\mathcal{L}(\mu \mid \mathrm{D}) = \frac{(\mu s + b + \xi)^{n_{\mathrm{obs}}} e^{-(\mu s + b + \xi)}}{n_{\mathrm{obs}}!} \exp\left(-\frac{\xi^2}{2\delta^2} \right), \tag{8.3}$$

   where n_{obs} is the number of observed events in the signal region under consideration, b and s are the number of expected background and signal events, and $\delta^2 = \delta_{\mathrm{b}}^2 + \delta_{\mathrm{s}}^2$ are the signal plus background uncertainty. Where n_{obs}, b and δ_b were taken from experimental results, the uncertainty on the signal (δ_s) was assumed to be 20%.

4. The ideal case was an EM-type result with a statistical model. This was only available for a handful of analyses at the time of publication, and as such, none were used in this iteration. Nevertheless, the publication of covariance matrices or, in some cases, full likelihoods allows for combining multiple SRs. The benefit of providing such information is that it increases the amount of information available

from a signal analysis. Without, only a single "best" SR from each analysis can contribute to the global likelihood estimate.

Constructing a global likelihood using multiple results relied upon the assumption that results from distinct experiments and taken from different LHC runs were approximately uncorrelated. This is a reasonable assumption to make and one that was used in the TACO project (Chapter 7). The proto-modelling paper also conducted a "by-eye" assessment, which involved identifying results with different final states in their signal regions (e.g., fully hadronic final states vs. final states with leptons) and defining them as uncorrelated. The final result was similar to the BAM from Section 7.2.2 in that it is a symmetric binary matrix where each element defined the combinability of the SRs corresponding to the column and row. The likelihoods from uncorrelated analyses were then combined by taking the product:

$$\mathcal{L}_{\text{BSM}}(\mu) = \prod_{i=1}^{n} \mathcal{L}_i(\mu)\,, \tag{8.4}$$

where μ is the global signal strength parameter, which was bound by an upper limit obtained from the most sensitive analysis,

$$\mu \times \sigma = 1.3 \times \sigma_{\text{obs}}^{\text{UL}} \Rightarrow \mu < \mu_{\text{max}} = 1.3\frac{\sigma_{\text{obs}}^{\text{UL}}}{\sigma}\,. \tag{8.5}$$

where σ is the signal cross-section and the 1.3 factor allowed for a 30% violation of the 95% CL observed limit. Minor violations were allowed to account for the fact that a few are statistically allowed to be violated when simultaneously checking limits from many analyses.

8.1.4. MCMC-type walker

Before exploring the *walker* algorithm, it is worth looking at the test statistic K. During the MCMC-type walk, K was maximised and, as such, was designed to increase for proto-models that satisfy the given constraints. Thus, for a proto-model, for each combination of results $c \in C$, the auxiliary quantity K^c was defined as:

$$K^c = 2\ln\frac{\mathcal{L}^c_{\text{BSM}}(\hat{\mu}) \cdot \pi(\text{BSM})}{\mathcal{L}^c_{\text{SM}}(\mu = 0) \cdot \pi(\text{SM})} \tag{8.6}$$

where $\mathcal{L}^c_{\text{BSM}}$ is the likelihood for a combination of experimental results *given* the proto-model. Being the BSM likelihood, this was evaluated at the signal strength value $\hat{\mu}$, which maximised the likelihood and satisfies $0 \leq \hat{\mu} < \mu_{\text{max}}$. $\mathcal{L}^c_{\text{SM}}$ is the corresponding SM likelihood, given by the usual $\mathcal{L}^c_{\text{SM}}(\mu = 0)$. The $\pi(\text{BSM})$ and $\pi(\text{SM})$ are the priors for the proto-model and SM, respectively. The choice of priors was crucial to formulating the test statistic as it allowed for the introduction of a penalty for model complexity. For the SM, a flat prior was assumed ($\pi(SM) = 1$) as the SM was a common factor for all combinations and, as such, did not affect the comparison between different proto-models. For the BSM prior, the following distribution was chosen:

$$\pi(\text{BSM}) = \exp\left[-\left(\frac{n_{\text{particles}}}{a_1} + \frac{n_{\text{BRs}}}{a_2} + \frac{n_{\text{SSMs}}}{a_3}\right)\right] \tag{8.7}$$

Where $n_{\text{particles}}$, n_{BRs} and n_{SSMs} refer to the number of new particles, branching ratios and signal strength multipliers, respectively. The parameters a_1, a_2, and a_3 were chosen to be 2, 4, and 8, respectively. This was chosen such that one particle with one decay and two production modes was equivalent to one free parameter in the Akaike Information Criterion (AIC) [256]:

$$\text{AIC} = -2\ln\left[\mathcal{L}_{\text{max}}(\hat{\boldsymbol{\theta}})\right] + 2k\,, \tag{8.8}$$

where $\mathcal{L}_{\text{max}}(\hat{\boldsymbol{\theta}})$ is the maximised profiled likelihood and k is the degrees of freedom. The test statistic K is defined as the maximum value of K^c

$$K = \max_{\forall c \in C} K^c\,. \tag{8.9}$$

The test statistic was assumed to roughly correspond to a $\Delta\chi^2$–distributed variable. However, the authors made it clear that the choice of prior in Equation (8.7) along with the introduction of the critic and the selection of the maximum K^c, made the exact distribution of K unknown.

The *walker* algorithm comprised several stages or building blocks. A procedural flow chart is shown in Figure 8.1, detailing the stages of a single *walker* iteration, which follows the subsequent steps:

- Starting with the Standard Model, the builder created proto-models by randomly adding or removing particles and changing any proto-model parameters as follows:

- **Add a new particle**: one of the BSM particles not yet present in the model could be randomly added. Once added, the new particle's mass was drawn from a uniform distribution between the LBP mass and 2.4 TeV, conditions on the generations already present, i.e. $m_{X_b^2} > m_{X_b^1}$. The new particle was initialized with random branching ratios and signal strength multipliers set to one. Adding a particle is programmed to occur more often for models with low test statistics and/or with a small number of particles.

- **Remove an existing particle**: one particle in the model was randomly selected and removed. All the production cross sections and decays involving the removed particle were deleted, and the remaining branching ratios were normalized, so they summed to unity. The removal of a particle was set to occur more often for models with low test statistics and/or with a large number of particles.

- **Change the mass of an existing particle**: The mass of a randomly chosen particle was changed by an amount δ_m according to a uniform distribution whose exact interval depends on the test statistic and the number of unfrozen particles in the model, with better-performing models making smaller changes. This change was always performed if no other changes had been made in the proto-model in a given step.

- **Change the branching ratios**: the branching ratios of a randomly chosen particle was changed. This change can occur in three ways: First, a random decay channel can set its BR to 1, and all other channels are closed. Second, a random decay channel can be closed and third, each decay channel can have its BR modified by a distinct random amount δ_{BR} drawn from a uniform distribution between $-a$ and a, where $a = 0.1/(\text{number of open channels})$. After any of these changes, the branching ratios were normalized to make sure they sum to unity.

- The proto-model was then passed on to the critic, which checked it against the database of SMS results to determine an upper bound on signal strength (μ_{max}).

- The *combiner* identified all possible results combinations– assuming that results from different LHC runs and distinct experiments (ATLAS or CMS) were combinable–and constructed a combined likelihood for each subset.

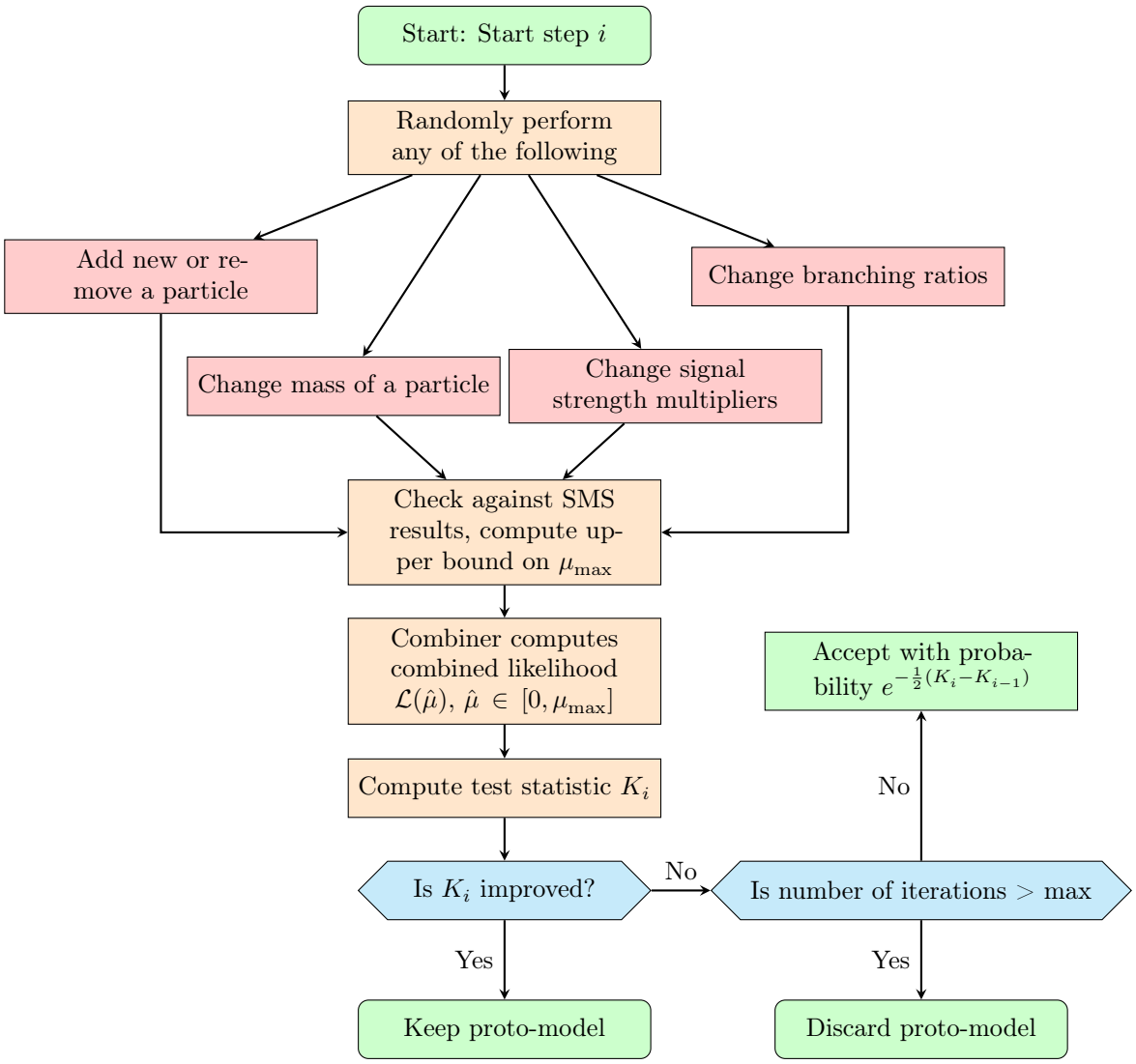


Figure 8.1.: Procedural flowchart of a single interaction of the "walker" algorithm presented in the proto-modeling paper [233]. See the text for further details.

- Using the combinations provided by the *combiner*, the *walker* computed the test statistic K for the proto-model. The new proto-model was kept if the K value was higher than the one obtained in the last step. If it was lower than the previous K, the step was accepted with a probability of $\exp[-\frac{1}{2}(K_i - K_{i-1})]$, where i was the index of the current step.

Over many iterations of the above procedure, the *walker* could identify proto-models that evade all simplified-model limits and, simultaneously, could explain dispersed signals in the data.

8.1.5. Results and conclusions from v1

The *walker* algorithm was applied to the SModelS database, performing ten runs, each employing 50 walkers and 1,000 steps/*walker*. The results are summarized in Figure 8.2, showing each run's proto-models with the highest K value. Besides the X_Z^1 LBP, all models included one top partner, X_t^1, and one light-flavor quark partner, $X_{d,c}$, and their test statistics were at $K = 6.76 \pm 0.08$, which, argue the authors, showed the stability of the algorithm. The X_μ particle introduced in run 5 was due to small ($\approx 1\sigma$) excesses in the CMS search for sleptons, CMS-SUS-17-009 [257] and the ATLAS search for electroweakinos, ATLAS-SUSY-2016-24 [258]; the prior's penalty for introducing the additional particle, $\Delta K = -1.25$, was overruled by the increase in the likelihood ratio, $\Delta K = 1.94$.

The top performing proto-model from Figure 8.2 obtained a test statistic value of $K = 6.90$ and was generated in step 582 of the 29th *walker* in run 9. It had X_t^1, X_d and X_Z^1 masses of 1166, 735 and 163 GeV, respectively, and produced signals in the $t\bar{t} + E_{\text{miss}}^{\text{T}}$ and jets $+ E_{\text{miss}}^{\text{T}}$ final states with SUSY-like cross sections. The effective signal strength multipliers were found to be $\hat{\mu} \times \kappa_{X_t^1 \tilde{X}_t^1} \approx 1.2$ and $\hat{\mu} \times \kappa_{X_d \tilde{X}_d} \approx 0.5$, corresponding to $\sigma(pp \to X_t^1 \tilde{X}_t^1) \approx 2.6$fb and to $\sigma(pp \to X_d \tilde{X}_d) \approx 24$fb at $\sqrt{s} = 13$ TeV and both the X_1^t and X_d directly decay to the lightest state (X_Z^1) with 100% BR. The full results can be found in the proto-modeling paper Ref.[233], A global p-value was estimated using a kernel density estimate (KDE) for the test statistic K. A p-value of 0.19 was determined which did not reach the threshold typically used to claim a significant discovery (set at 5σ or 0.00003%).

The paper demonstrated that the proto-model approach could navigate the complexity of LHC data to identify signals of new particles, such as top partners and light-flavour quark partners. The best-performing proto-models feature particles with masses around 1.2 TeV, 700 GeV, and 160 GeV, and the authors calculated a global p-value of approximately 0.19 for the SM hypothesis. Extensive thought and effort went into avoiding the "look-elsewhere effect" by constructing upper limits on μ and introducing priors with penalties for free parameters. Although still approximate in some instances, these procedures attempted to ensure that the signals identified were not statistical flukes but potential avenues for discoveries. The paper emphasizes the importance of building upon existing simplified models to guide future theoretical and experimental research in high-energy physics, laying the groundwork for a potential "next Standard Model".

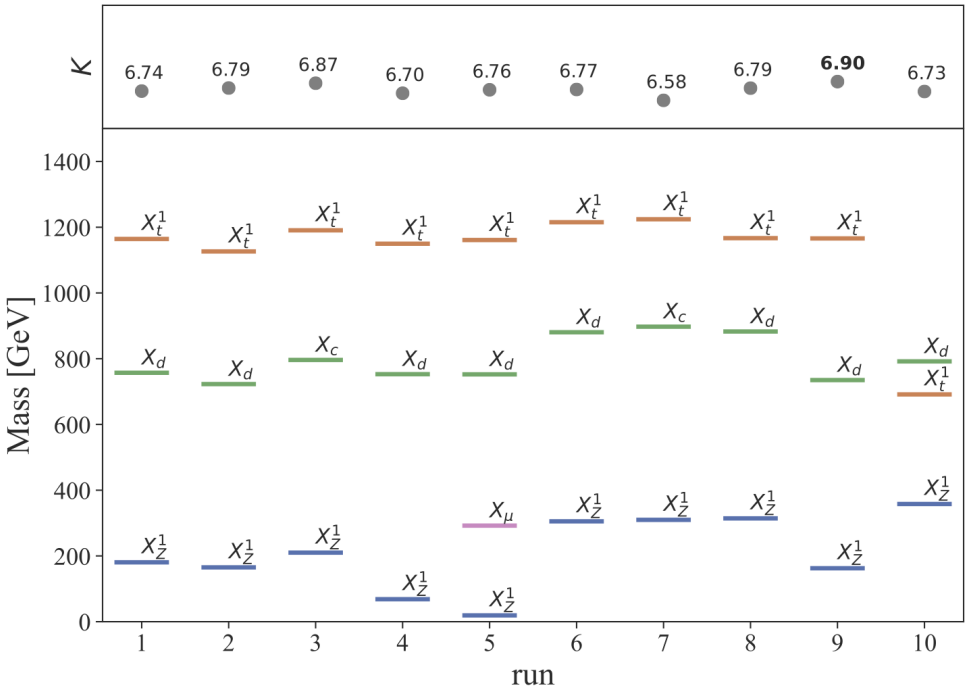


Figure 8.2.: Particle content and masses for the proto-models with highest test-statistic K obtained in each of the ten runs performed over the SModelS database. The corresponding K values are shown at the top. This plot was taken from the results of the "proto-modelling" paper Ref. [233]

8.2. Proto-models version 2

At the time of the original "proto-modelling" paper [233], the authors stressed the results were intended as a proof-of-principle. The variety and type of simplified-model results available from SModelS v1.2 [234] limited the realisation of the proto-model builder. However, since publication, the SModelS framework and database have seen a large expansion in available data and capability (see section 5.1.3). For version 1, 40 ATLAS and 46 CMS analyses were included in the SModelS database v1.2.4. For version 2, SModelS v3 [184] with database v3, contains 62 ATLAS and 63 CMS analyses, at $\sqrt{s} = 8$ and 13 TeV. Crucially for the proto-model project, the database contains 6,349 efficiency maps from 1,430 distinct signal regions and 117 different SMS topologies [217].

The frustration of such a large results database was that the original *walker* couldn't compute all possible combinations, as the increasing number of combinations quickly created a computational bottleneck. Fortunately, the PathFinder algorithms developed

during the TACO protect (see Chapter 7) provided a mechanism to efficiently select subsets of minimally correlated results from large numbers of analyses/SRs.

Before simply plugging the PATHFINDER algorithms into the proto-models analysis chain, consideration had to be given to what effect this would have on any resulting test statistic and whether alternative test statistics would be required. To this extent, the following sections will consider various test statistics for the combinations.

The second proto-models analysis is ongoing at the time of writing, with only preliminary results. Thus, much of this chapter will concentrate on the work done to integrate the PATHFINDER algorithm into the analysis chain and the considerations made when changing context from exclusion to anomaly detection. We will finish with a project update, looking at the changes made to the analysis chain and presenting preliminary results.

8.2.1. PATHFINDER for anomaly detection

In Chapter 7, the WHDFS algorithm was used to find the optimum set of exclusionary SRs. Fundamental to this procedure was the definition of the test statistic ω, in Equation (7.7), that served as the edge weight for the WHDFS algorithm. We defined ω to be the logarithm of the expected negative log-likelihood ratio (NLLR) between the signal model ($\mu = 1$), under test, and the background-only model at a signal strength $\hat{\mu}$ that maximised the likelihood:

$$\begin{aligned} \Lambda_i^{\text{exp}}(\mu) &= \frac{\mathcal{L}_i^{\text{exp}}(\mu = 1, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}_i^{\text{exp}}(\hat{\mu}, \hat{\boldsymbol{\theta}})}\,, \\ \omega_i &= -2 \ln \Lambda_i^{\text{exp}}\,, \end{aligned} \tag{8.10}$$

where i runs over the combination C, with $\hat{\hat{\boldsymbol{\theta}}}$ and $\hat{\boldsymbol{\theta}}$ refering to the conditional ML estimators of the expected $\boldsymbol{\theta}$ given a signal-strength parameter μ. This test statistic was chosen because, under the assumption of Wald's theorem [227], the expected exclusion is given by the square root of the NLLR between models. Hence, the NLLR maximises the expected model exclusion. Another motivation for this choice was that ω is additive, with $\Omega^{\text{exp}} = \sum \omega_i$, where Ω^{exp} has an asymptotic distribution approaching χ^2 with one degree of freedom according to Wilks' theorem [150, 228].

For anomaly detection, the choice of test statistic becomes more involved as the data moves away from expectations and into observed data. This section will examine the available test statistics options and how one was finally selected for use in the proto-models project.

8.2.2. The test statistic α_i'

The first step in any statistical analysis is to interrogate the question. For anomaly detection, the question is, for a given BSM theory or model point, "do significant discrepancies exist in the data compared to SM predictions?" Let's, for the moment, ignore the complexity of choosing a given BSM model point, assume we have a model point in mind and concentrate on quantifying the discrepancy compared to the SM. The first and somewhat obvious divergence from ω is the shift to using observed data; using expected background pseudodata rather than observed data was important to avoid cherry-picking statistical fluctuations. However, for anomaly detection, observed data has to be compared to SM expectations, requiring a modification of the Equation (8.10). The naïve choice for such a modification would be

$$\begin{aligned} \Lambda'_{\mathrm{A},i} &= \frac{\mathcal{L}_i^{\mathrm{obs}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}_i^{\mathrm{obs}}(\hat{\mu}, \hat{\boldsymbol{\theta}})}\,, \\ \alpha'_i &= -2\ln\Lambda'_{\mathrm{A},i}\,, \end{aligned} \tag{8.11}$$

i.e., α'_i is twice the observed negative log-likelihood ratio (NLLR) between the SM ($\mu = 0$), under test, and the signal model at the signal strength that maximises the likelihood.

The problem with α'_i is that switching to observed likelihoods meant that the sum of the LLRs was no longer equal to the log of the combined likelihood ratio. This is because the product of minimally correlating likelihoods optimised at $\hat{\mu}$ does not necessarily equal the product of optimised likelihoods,

$$\mathcal{L}_{\mathrm{comb}}^{\mathrm{obs}}(\hat{\mu}) \leq \prod_{i\in C} \mathcal{L}_i^{\mathrm{obs}}(\hat{\mu}_i)\,, \tag{8.12}$$

This inequality made the process of evaluating the edge weight in the WHDFS algorithm a more complicated and CPU-intensive task. However, the inequality could be leveraged

to provide an upper limit by combining Equations (8.11) and (8.12)

$$\begin{aligned} \bar{\Omega}' &= \sum_{i \in C} \alpha'_i \geq \Omega' \\ \Omega' &= -2 \ln \left(\frac{\mathcal{L}^{\text{obs}}_{\text{comb}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}^{\text{obs}}_{\text{comb}}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \right), \end{aligned} \tag{8.13}$$

where Ω' defines the combined NLLR. We can define an upper limit to the weight available in a current combination (C) by assuming all results that are potentially combinable with the current set (A_c) to be combinable and taking the sum of the remaining weight

$$\bar{\Omega}'_{\text{UL}} = \sum_{i \in C} \alpha'_i + \sum_{j \in A_c} \alpha'_j \,. \tag{8.14}$$

The WHDFS algorithm was altered such that the "current weight" (Ω') and the "remaining weight" ($\bar{\Omega}'_{\text{UL}}$) were calculated using functions defined (optionally) by the user. This meant that the Ω' could be calculated using the combined likelihood ratio, and the $\bar{\Omega}'_{\text{UL}}$ could provide an upper limit to the weight available to any given set. The solution vastly reduced the computational efficiency of the WHDFS algorithm, resulting in the algorithm taking minutes to run per model point rather than fractions of a second for simple additive weights. This could be a serious issue for a test statistic required to run over many thousands of proto-model points, however, this was still manageable given large computational resources.

With a partially workable solution to the combined weight calculation, the focus turned to understanding the asymptotic distribution of Ω'. When discussing the distribution of the test statistic Ω_E for the expected NLLR under the SM hypothesis (see Section 7.3.1), we defined the general case to be a non-central χ^2 (Equation (7.12)). The distribution was simplified further by Wilks' theorem; however, as previously mentioned, the definition of $\bar{\Omega}'$ no longer satisfies the conditions of the theorem, and the non-centrality parameters of Equation (7.12) no longer vanish. However, the non-centrality terms in the distribution of Ω' will cancel under the SM hypothesis. Retaining the original assumption that μ is Gaussian distributed, the general distribution for a test statistic t_μ takes the form:

$$\begin{aligned} f(t_\mu \mid \mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{t_\mu}} \Bigg[& \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} + \sqrt{\lambda_\mu} \right)^2 \right) \\ & + \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} - \sqrt{\lambda_\mu} \right)^2 \right) \Bigg], \end{aligned} \tag{8.15}$$

where λ_μ is unknown for $t_\mu = \bar{\Omega}'$, and $\lambda_\mu = 0$ for $t_\mu = \Omega'$. To compute the significance of a test statistic t_μ and, ultimately, obtain a p-value, we needed to understand $f(t_\mu \mid \mu)$. Fortunately, a suite of tools was available as a codebase from the previous proto-model anomaly search by the SModelS group [233]. This included an Experimental Results Modifier (ERM), which allowed for the production of "MC fake" versions of the SModelS database where the observed data is replaced by values sampled from the SM background. Testing the distribution of t_μ required event yields for a given BSM point under the SM hypothesis; thus, the toy database was restricted to data with available efficiency maps (EM). Generating data for EM-type results first requires sampling from the background distributions; for results with covariance matrices or full statistical models, this is done by sampling events from all SRs at once. For the cases without, a normal distribution is sampled with a mean corresponding to the background estimate (b) and a variance of the squared background error (σ_b). The sampled value is then entered as the expectation parameter of a Poissonian distribution. The value drawn from the Poissonian is then used as the "fake" observation, i.e. the fake event yield.

For the proto-models paper, the pseudo-databases were used to produce fake databases, which were used by the *Walker* algorithm. For each pseudo-database, the proto-models with the highest test statistic K were combined to estimate the density of the test statistic K under the SM hypothesis via a kernel density estimator. A similar strategy was employed to test the assumption that Ω' was $\chi^2 \sim$ distributed by adjusting the ERM to procedurally generate hundreds of pseudo-databases, which were used to create the BAM and weights for the PathFinder module. Unfortunately, it soon became apparent that maximising the combined likelihood ($\mathcal{L}_{\text{comb}}$) within the whdfs algorithm was infeasible due to the computational complexity increasing the time taken per model point beyond the point where it could be used in an anomaly search.

8.2.3. The test statistic α_i

Using the infrastructure in place to evaluate the previous test statistic, an alternative that removed the need to maximise $\mathcal{L}^{\text{obs}}$ for each set was considered.

$$\begin{aligned} \Lambda_{\text{A},i}(\mu) &= \frac{\mathcal{L}_i^{\text{obs}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}_i^{\text{obs}}(\mu = 1, \hat{\hat{\boldsymbol{\theta}}})}\,, \\ \Omega = \sum_i \alpha_i &= \sum_i -2 \ln \Lambda_{\text{A},i}. \end{aligned} \tag{8.16}$$

Here, once again, $\mathcal{L}(\mu = 1, \hat{\hat{\boldsymbol{\theta}}})$ is the likelihood of the signal plus background model and $\mathcal{L}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})$ is that of the background-only hypothesis. The test statistic, α_i, can be understood as the difference between $\alpha'(\mu = 0)$ and $\alpha'(\mu = 1)$, which, following log rules, equals the ratio of ratios, thus cancelling out the $\hat{\mu}$ terms

$$\alpha_i = -2\ln\left(\frac{\mathcal{L}^{\text{obs}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}^{\text{obs}}(\hat{\mu}, \hat{\hat{\boldsymbol{\theta}}})}\right) + 2\ln\left(\frac{\mathcal{L}^{\text{obs}}(\mu = 1, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}^{\text{obs}}(\hat{\mu}, \hat{\hat{\boldsymbol{\theta}}})}\right) = -2\ln\left(\Lambda_A(\mu)\right) . \tag{8.17}$$

Removing the maximisation step means that the sum of α_i ($\sum \alpha_i = \Omega$) can be used directly as a test statistic. A similar statistic is discussed in "*Asymptotic formulae for likelihood-based tests of new physics*" [228], where the ratio is presented as $\mathcal{L}_{s+b}$ / $\mathcal{L}_b$ as opposed to $\mathcal{L}_b$ / $\mathcal{L}_{s+b}$, following the same logic one can express α_i as

$$\alpha_i = -2\ln(\Lambda_{A,i}(\mu)) = -2\ln\left[\mathcal{L}_i^{\text{obs}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})\right] + 2\ln\left[\mathcal{L}_i^{\text{obs}}(\mu = 1, \hat{\hat{\boldsymbol{\theta}}})\right] . \tag{8.18}$$

We once again assume the validity of the Wald approximation [227], which states that in the limit of large statistics (N), for the case of a single parameter of interest, the LLR is:

$$\mathcal{L}^{\text{obs}}(\mu, \hat{\hat{\boldsymbol{\theta}}}) = \frac{(\hat{\mu} - \mu)^2}{\sigma^2} + \mathcal{O}\left(1/\sqrt{N}\right) . \tag{8.19}$$

Applying Equation (8.19) to Equation (8.18), one can approximate α_i as

$$\alpha_i \approx \frac{\hat{\mu}^2}{\sigma^2} - \frac{(\hat{\mu} - 1)^2}{\sigma^2} = \frac{2\hat{\mu} - 1}{\sigma^2} , \tag{8.20}$$

where σ^2 is the variance of $\hat{\mu}$. Assuming $\hat{\mu}$ follows a Gaussian distribution, the distribution of α_i will also be Gaussian with an expectation value and variance of:

$$\mathrm{E}[\alpha_i] = \frac{2\mu - 1}{\sigma^2}, \quad \mathrm{Var}[\alpha_i] = \frac{4}{\sigma^2}, \quad \mathrm{Std}[\alpha_i] = \frac{2}{\sigma} , \tag{8.21}$$

where μ is the true unknown value of the signal strength parameter. The left-hand plots of Figure 8.3 show a normally distributed random variable X. The upper plot shows four histograms representing example distributions of $\hat{\mu}$ over four different mean values $\mathrm{E}[\mu]$ with unit variance. The lower plot transforms X via Equation (8.20), giving the expected distribution of α_i (with colours corresponding to the relevant $\hat{\mu}$). The right-hand plots show the variance (upper) and the expectation value (lower) of the transformed samples as a function of σ (standard deviation of $\hat{\mu}$) with μ fixed at 0 and 1. Looking a Figure 8.3

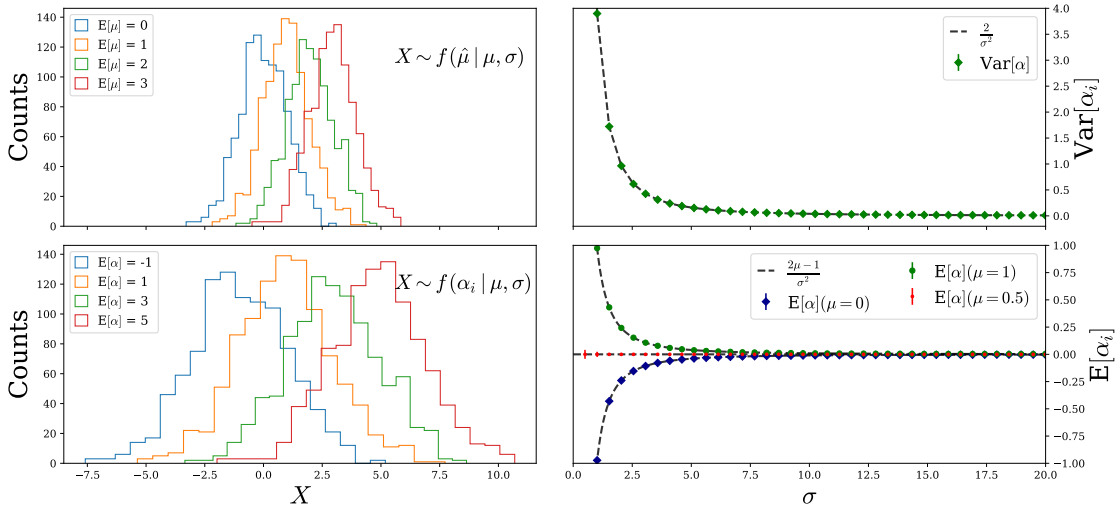


Figure 8.3.: Relation between distributions of μ' and α_i following the the transform from Equation (8.20). The left-hand plots show a normally distributed random variable X. The upper plot shows four histograms representing example distributions of μ' over four different mean values $E[\mu]$ with unit variance. The lower plot transforms X via Equation (8.20), giving the expected distribution of α_i. The right-hand plots show the variance (upper) and the expectation value (lower) of the transformed samples as a function of σ (standard deviation of μ') with μ' fixed at 0, 0.5 and 1.

and Equation (8.21) it is clear that if the expectation value of $\hat{\mu}$ is less than 0.5 the resulting expectation of α_k will be $\mathrm{E}[\alpha_i] < 0$ approaching zero at large $\mathrm{Var}[\hat{\mu}]$. This makes intuitive sense when considering the NLLR in the case where the data heavily favours the SM hypothesis ($\mu = 0$); thus, $\mathcal{L}(\mu = 0) > \mathcal{L}(\mu = 1)$ meaning that the ratio would be greater than one and α_i would be less than zero.

As no significant evidence for SUSY had been reported from the ATLAS or CMS experiments–where significant implies a 5σ signal from a single search analysis–we can comfortably assume that if any BSM signal does exist in the data, it is close to the SM expectations. Thus, to understand the significance of such a signal, we need to understand the behaviour of α_i for SM-only data. In addition to the pseudo-database generator, the distributions of α_i, $\hat{\alpha}_i$, Ω, and $\hat{\Omega}$ were investigated in a simplified environment, independent of the SModelS framework. This was done by breaking the analysis chain down into components:

1. A well-defined background and signal model with cuts applied to isolate signal regions.

2. Background and signal events sampled b and $s + b$ models.

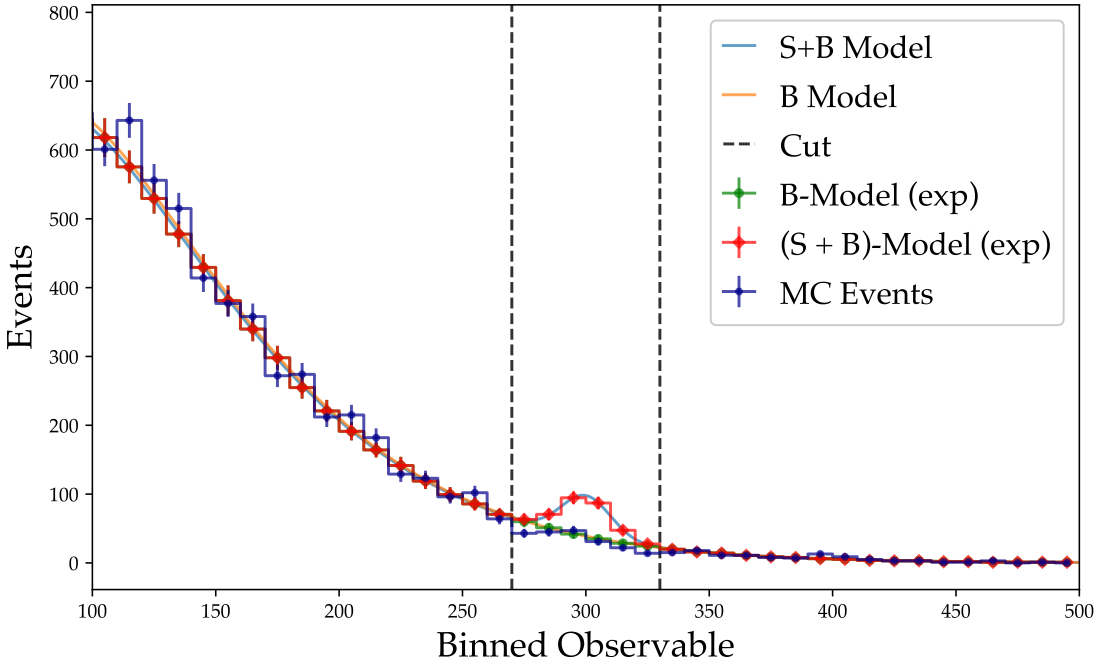


Figure 8.4.: Example of the toy model described by Equation (8.22), comprising a gamma-distributed background with a Gaussian signal. The signal fraction has been set to $f_s = 0.015$, producing a well-pronounced bump in the background tail.

3. A representative likelihood equation with effective nuisances (as described in Section 7.3.1).

4. A binary acceptance matrix describing the overlap between signal regions

A simple signal-background model was chosen, a linear combination of a gamma-distributed background (P_b) and a Gaussian signal (P_s) with a mean located on the right-hand tail of the background. The general form of the model was

$$P(x|\theta_b, \theta_s, f_s) = (1 - f_s)P_b(x \mid \theta_b) + f_s P_s(x \mid \theta_s)\,, \tag{8.22}$$

where θ_b and θ_s contain the parameters of the distributions and f_s is the signal fraction, maintaining overall normalisation via the opposing $1 - f_s$ term. The f_s term should not be confused with the signal strength parameter μ, as f_s controls the contribution of the signal PDF to the overall model and does not necessarily correspond to the number of signal events. Equation (8.22) was used to generate a range by defining values of θ_b, θ_s and f_s and applying a "kinematic" cut. Events could then be generated by applying Monte-Carlo (MC) sampling, which involves drawing random samples from an expected distribution.

In this case, the predicted distribution represents a probability density, but the likelihood of specific particle interactions or decay channels is often used. The MC method used here is the inverse-CDF transform MC and begins by identifying the probability density function (PDF) of the process under investigation. A cumulative distribution function (CDF) is then derived from the PDF, and random numbers are generated uniformly between 0 and 1. These random numbers are mapped to the corresponding values of the CDF to sample from the original distribution. This procedure ensures that the generated events follow the same statistical properties as the theoretical model, such as the shape of the distribution. Figure 8.4 shows an example implementation of the sampling method. MC sampling was used to generate multiple sets of background and signal events, using the average value and standard error per bin as the expected B (green) and S+B (red) counts. Example events (blue) were then generated using single MC runs with Poisson errors on the bin counts. It is clear from Figure 8.4 that the "MC events" were generated containing no signal; this will become an important detail when discussing how to interpret the significance of the statistic.

With the tools to generate events from a given background and signal model, the next task was to select a likelihood function. Two different functions were chosen for thoroughness, representing two very different approaches to the same task. The first function contains a single nuisance that uniformly scales both signal and background [7]:

$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \xi\big) = \prod_i \big[\xi(\mu s_i(\boldsymbol{\theta}) + b_i)\big]^{n_i} \frac{1}{n_i!} \exp\big(-\xi(\mu s(\boldsymbol{\theta})_i + b_i)\big) P(\xi)\,, \tag{8.23}$$

where ξ acts as a global scale factor on the expected rate, which approximately accounts for the systematic uncertainties. The probability distribution for ξ is given as:

$$P(\xi|\sigma_\xi) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left[-\frac{1}{2}\left(\frac{1-\xi}{\sigma_\xi}\right)^2\right]. \tag{8.24}$$

This distribution peaks at $\xi = 1$ and has a width characterised by

$$\sigma_\xi = \frac{\sigma_b^2 + \mu^2\sigma_s^2}{(b+\mu s)^2}\,. \tag{8.25}$$

The profiled likelihood is defined as:

$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b\big) = \max_\xi \big\{\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \xi\big)\big\}\,, \tag{8.26}$$

where the maximisation occurs over the scalar quantity ξ, the second likelihood equation follows the general form defined in Equation (7.8):

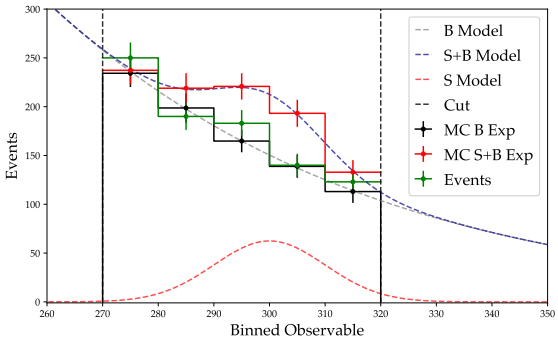
$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \boldsymbol{\Delta}\big) = \prod_i \mathcal{P}\big(n_i \mid \mu, s_i(\boldsymbol{\theta}), b_i, \Delta_i\big) \frac{1}{(2\pi)^{n/2} \det \boldsymbol{\Sigma}} \exp\big(\frac{1}{2}\boldsymbol{\Delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}\big)\,, \quad (8.27)$$

where the $P(\boldsymbol{\Delta})$ from Equation (7.8) has been replaced with a multivariate normal distribution which follows the SIMPLIFIED LIKELIHOODS scheme [255] using linear responses of expected (background) event yields with an effective nuisance parameter Δ_i for each bin. The interplay between elementary nuisances is absorbed into a covariance matrix ($\boldsymbol{\Sigma}$), a diagonal matrix containing the background variances. The profiled likelihood is defined as

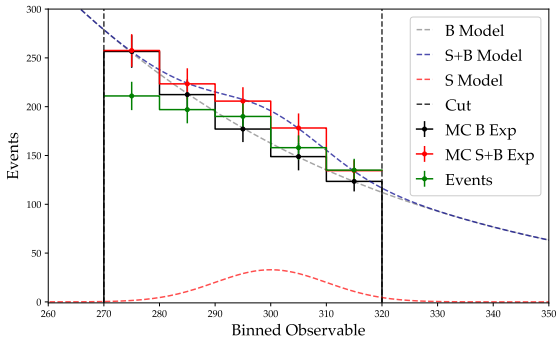
$$\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b\big) = \max_{\boldsymbol{\Delta}} \big\{\mathcal{L}\big(n \mid \mu, s(\boldsymbol{\theta}), b, \boldsymbol{\Delta}\big)\big\}\,, \quad (8.28)$$

where the maximisation occurs over the vector quantity $\boldsymbol{\Delta}$. It's worth noting that if we assume zero covariance (or zero off-diagonal terms) for $\boldsymbol{\Sigma}$, Equation (8.27) reduces independent instances of Equation (8.24) for each bin and thus can be profiled over each bin independently. For convenience, when referencing the profiled likelihoods from Equations (8.26) and (8.28) the corresponding sub-script will be used to differentiate between the two, i.e. $\mathcal{L}_\Delta$ and $\mathcal{L}_\xi$.

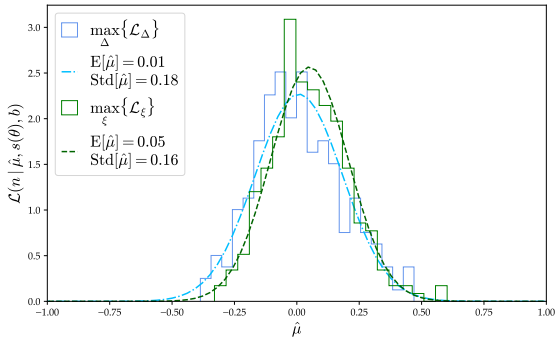
A series of test observables were defined and used to calculate the likelihoods to test the assumptions' validity in defining the profiled likelihood, α_i, and its corresponding distribution. Figure 8.5 shows the results from two toy models; both models were initiated with identical background and signal distributions but with different signal fractions f_s. For model one, f_s is set such that expected signal expectations are predominantly outside of the one-sigma band of the background expectations, meaning that the signal is easily differentiated from the background hypothesis. For model two, f_s is reduced to the point where the error bars overlap. Thus, the signal is not easily differentiated from noise. The upper plots show the binned observable expectations with an example set of "Events". The central plots (c and d) show the maximised $\hat{\mu}$ distribution for 500 events. The results are presented for both versions of the profiled likelihood $\mathcal{L}_\xi$ and $\mathcal{L}_\Delta$ The lower plots (e and f) show the distribution of the NLLR α_i, which according to Equations (8.20) and (8.21) should be normally distributed with an expectation value and standard deviation defined by the distribution of $\hat{\mu}$.



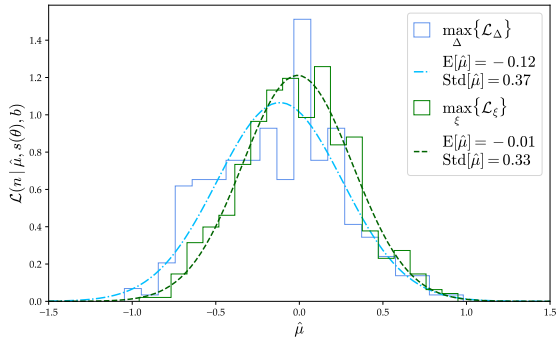
(a) Toy Model One $f_s = 0.004$



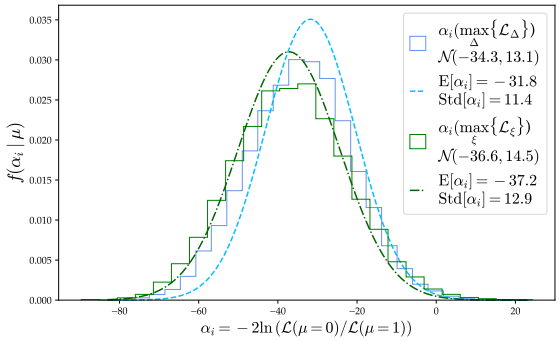
(b) Toy Model Two $f_s = 0.002$



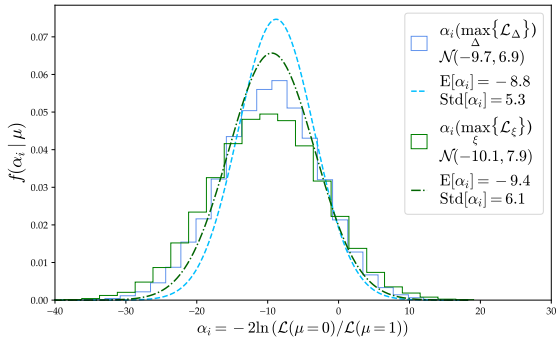
(c) Distribution of $\hat{\mu}$, Model One



(d) Distribution of $\hat{\mu}$, Model Two



(e) Distribution of α_i, Model One



(f) Distribution of α_i, Model Two

Figure 8.5.: Comparison of the NLL and NLLR behaviour for two toy observables (upper plots) following the relations derived in Equations (8.17) through (8.21). The plots confirm the relationship between α_i and $\hat{\mu}$ using the likelihood functions shown in Equations (8.23) and (8.27)

The results from the toy model confirmed the expected relations between $\hat{\mu}$ and α_i produced a consistent result for both versions of the profiled likelihood, which is not surprising considering the simplicity of the toy model. While investigating these distributions, it became clear that the α_i's dependence on $\hat{\mu}$ would become problematic

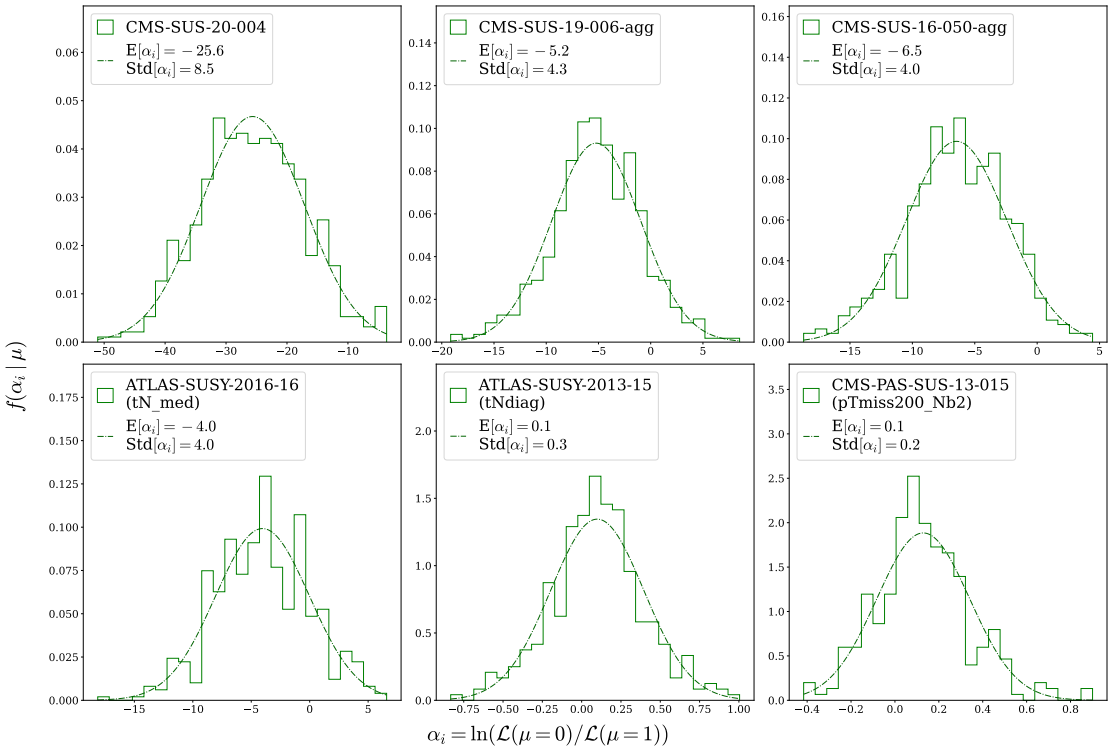


Figure 8.6.: distribution of α_i for pseudo data produced under the SM hypothesis using the *Experimental Results Modifier* for the SModelS database. The results were generated for three search analyses (upper) and three SRs (lower)

when combining the likelihoods. If each SR_i has a unique $\text{E}[\alpha_i]$ and $\text{Var}[\alpha_i]$, the sum of the unknown expectation and variances of the set would define the distribution of the sum. Therefore, defining the significance would require $\text{E}[\alpha_i]$ and $\text{Var}[\alpha_i]$ for each SR at each model point. Thus, for each point in BSM space, one would have to sample events to estimate the distribution of $\hat{\mu}$ under the SM hypothesis. Once again, this calculation would increase the computational complexity beyond the point of being practical.

Despite the possible impracticality of α_i, the next step was to evaluate the distribution of α_i using the proto-models pseudo-data generator simulating the SModelS database. There was still information to be gained, not only in the practical sense of troubleshooting the code base but in whether the distribution α_i would still confirm the "normalness" of $\hat{\mu}$. Figure 8.6 shows the distribution of α_i for pseudo data produced under the SM hypothesis for a proto-model point using three search analyses (upper) and three SRs (lower). The results confirm that α_i follows a normal distribution, indicating that $\hat{\mu}$ is also normally distributed. The plots demonstrate the variability of $\text{E}[\alpha_i]$ and $\text{Var}[\alpha_i]$ over the different SRs; when combining optimum sets using the WHDFS algorithm, this variation will affect the final distribution.

8.2.4. The combined test statistic $\hat{\Omega}$

Two key insights came from the derivations of Ω' and Ω. The optimisation of μ for combined likelihoods in Ω' produces test statistics described by a non-central χ^2. This distribution reduced to χ^2 when working with expected yields; the insight here was that this reduction occurred when the value of μ equalled the expectation of μ (μ'). The second insight came from the factorisation quality of Ω_i, defined by the sum of the individual evaluations of α_i. Combining these two qualities required a holistic reconsideration of the anomaly-detection procedure.

In the original plan for the proto-model analysis chain, the result selection would sit between the model point selection and combined likelihood evaluation. In this mode, the WHDFS algorithm plays a passive role, selecting the best set of SRs / Analyses for a given proto-model. To exploit the attributes of the previous test statistics, the combinations would have to be evaluated such that the product of $\mathcal{L}(\mu = 1, \hat{\boldsymbol{\theta}})$ is equal to the combined likelihood maximised at $\hat{\mu}_{\mathrm{c}}$:

$$\mathcal{L}(\hat{\mu}_{\mathrm{c}} \mid \hat{\boldsymbol{\theta}}') = \prod_{i}^{l} \mathcal{L}_i(\mu = 1 \mid \hat{\boldsymbol{\theta}}_i)\,. \tag{8.29}$$

This condition would require the recursive optimisation of $\hat{\mu}$ where the evaluation of $\hat{\mu}$ for a combined set of results is used to modify the expected signal yields via the cross-sections of the model until $\hat{\mu}$ converges to unity. This procedure modifies the proto-model point by multiplying the cross-sections by the signal strength parameter. It is important to note that μ is a global signal strength. Therefore, the signal cross-sections are given by $\sigma = \mu \times \kappa \times \sigma_{\mathrm{SUSY}}$, where κ are the signal strength multipliers defined in Section 8.1. Technically, once the value of μ which maximises the likelihood ($\hat{\mu}$) is determined, all signal strength multipliers are rescaled by $\kappa \to \hat{\mu} \times \kappa$ and μ is taken as 1. The modification of yields directly affects the WHDFS weights; thus, for each recursion, the best combination has to be re-evaluated, affecting $\hat{\mu}$ for the combined set. This could easily result in an infinite loop where $\hat{\mu}_A$ modifies the model such that set B is chosen to calculate $\hat{\mu}_B$, which modifies the model such that set A is chosen to calculate $\hat{\mu}_A$. With this situation in mind, a limit on the number of loops is required, along with a minimum tolerance on the optimised $\hat{\mu}$ to prevent small non-converging fluctuations around unity.

If successful, the recursion loop will return a new model point and a set of results that, when evaluated for Ω, will give a value approximately equal to Ω',

$$\hat{\Omega} \equiv -2\sum_{i}^{l} \ln\left(\frac{\mathcal{L}_i^{\text{obs}}(\mu=0,\hat{\hat{\boldsymbol{\theta}}}_i)}{\mathcal{L}_i^{\text{obs}}(\mu=1,\hat{\hat{\boldsymbol{\theta}}}_i)}\right) \approx -2\ln\left(\frac{\mathcal{L}_{\text{comb}}^{\text{obs}}(\mu=0,\hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}_{\text{comb}}^{\text{obs}}(\hat{\mu}_{\text{c}},\hat{\boldsymbol{\theta}})}\right). \tag{8.30}$$

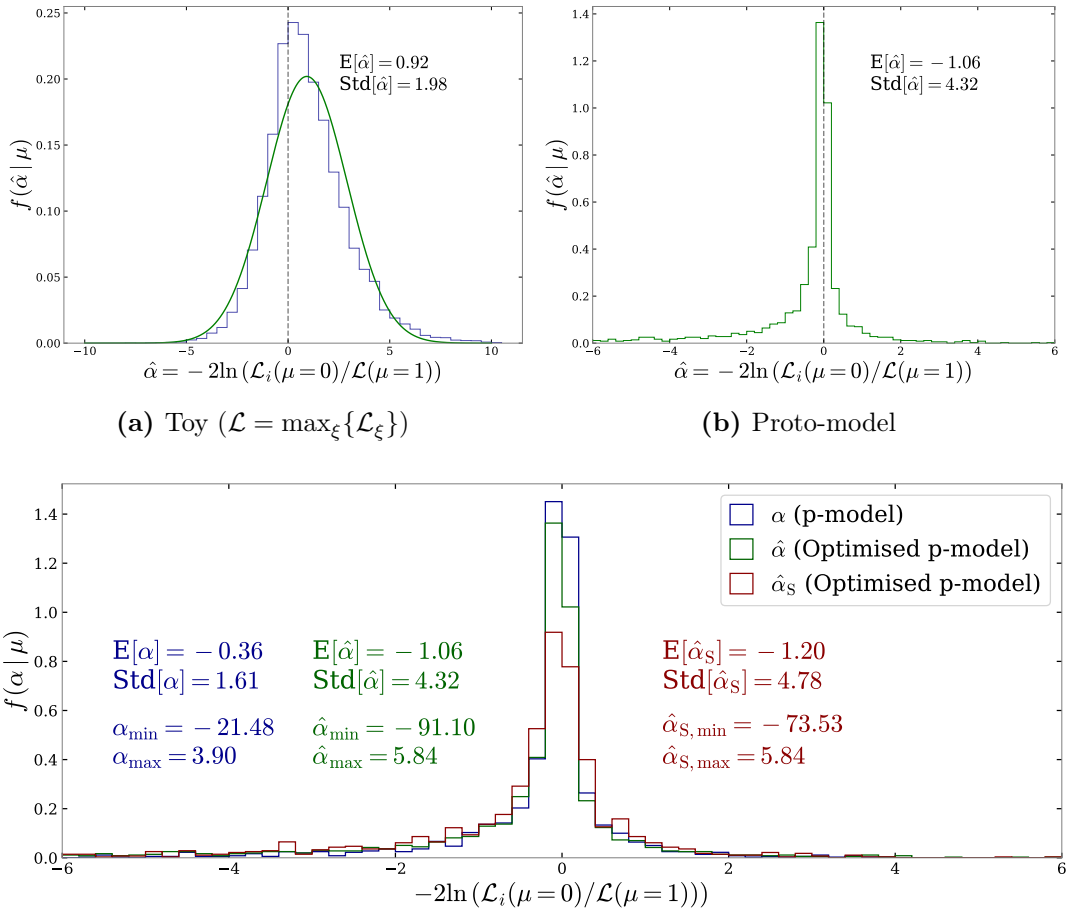
This procedure essentially adds a parameter optimisation stage to the proto-models analysis chain. We should be careful in identifying the difference between this equation and that of test statistics Ω' in Equation (8.11) and Ω in Equation (8.16). Here, $\hat{\Omega}$ is defined as when Ω' and Ω are approximately equal. This simultaneously retains the additive weights for the WHDFS algorithm to optimise and constrains the asymptotic distribution of the sum. We can also define Ω' as the maximum of all possible combinations under the optimised model point Ω'_{c}:

$$\hat{\Omega} = \max\left(\sum_{i}\hat{\alpha}_i\right) = \max\left(\Omega'_{\text{c}}\right). \tag{8.31}$$

Figure 8.7 shows the distribution of α for both the toy data and the proto-models pseudo-database results created using ERM. This shows the test statistic over all SRs (α) instead of a single SR α_i. The left-hand plot (a) shows the results for 2,000 bootstraps of 50 "toy" SRs like those shown in Figure 8.4. The right-hand plot (b) shows the distribution from 50 bootstrapped "fake universes" pseudo-databases containing 62 analyses [217] generated under the SM hypothesis. It is clear from the comparison that there is now a deviation in results between the two examples. The distribution is normal for the toy (a), as expected from the definitions of α_i and $\hat{\alpha}_i$ when compared to the distributions of α_i in Figure 8.4 it is clear the optimisation procedure shifts the distribution to the right.

The upper right-hand plot (b) was generated using 50 bootstraps of the pseudo-database produced under the SM (no signal) assumption using the latest proto-model result (to date). The deviation from the normal distribution is unsurprising, considering the results shown in Figure 8.6, where the different analyses and SRs had widely dispersed distributions. When considered together and under the optimisation condition, the results in Figure 8.7 have an expectation of 1.06 with a range of values going from $\hat{\alpha}_{\text{min}} = -91$ to $\hat{\alpha}_{\text{max}} = 5.8$, the plot has been truncated to provide greater detail for the main body of the histogram.

As previously discussed, the summation term of Equation (8.30) is equivalent to Equation (8.16) and was shown to be normally distributed. However, the maximisation



(a) Toy ($\mathcal{L} = \max_\xi\{\mathcal{L}_\xi\}$) **(b)** Proto-model

(c) α, $\hat{\alpha}$ comparison

Figure 8.7.: The distribution of α. The left-hand plot (a) shows the results for 2,000 bootstraps of 50 "toy" SRs like those shown in Figure 8.4. The right-hand plot (b) shows the distribution from 50 bootstrapped pseudo-databases created using ERM containing 62 analyses [217] generated under the SM hypothesis. Plot (c) compares the distribution of the values before and after the proto-models optimisation step $\alpha \to \hat{\alpha}$. $\hat{\alpha}_S$ is $\hat{\alpha}$ filtered for significance such that results are either full statistical models or the best performing SRs from a single analysis

condition imposed by the right-hand side means that $\hat{\Omega}$ is positive definite and, making the usual assumptions regarding $\hat{\mu}$, one would assume it to be χ^2 distributed with one degree of freedom. Plot (c) of Figure 8.7 shows the difference in the distribution moving from α to $\hat{\alpha}$, where α is the test statistic before the $\hat{\mu}$ optimisation step. Because the optimisation is performed on the highest scoring set of SRs, the optimisation moves the proto-model point into a region supported by that subset. This broadens the distribution, as evidenced by the increase in the standard deviation and the gap between the extreme minimum and maximum values. The plot also shows the distribution of $\hat{\alpha}_S$; this is the

distribution of the $\hat{\alpha}$ filtered for significance such that results are either full statistical models or the best-performing SRs from a single analysis. The implications of $\hat{\alpha}_S$ will be discussed further in Section 8.2.5.

Finally, looking at Equation (8.30), there is an extra consideration to make, as the summation over α_i introduces a new degree of freedom (DOF) in the form of l, the length of the combination. This can be considered a DOF because the best combination of results is recalculated at each iteration of the $\hat{\mu}$ optimisation, meaning that l is not a fixed quantity during this process.

8.2.5. Subsets and the look-elsewhere effect

As discussed in Section 6.3.1, the WHDFS algorithm can be run in two modes: allowing subsets and not allowing subsets. For the current use case, we want to select the best sample of results from a large selection of analyses, but we also want that sample to represent reality. So, should it be dropped from the set when the WHDFS algorithm chooses a selection of l results if dropping the lth result improves the overall test statistic? The answer to that question is no, as to do so is the definition of "cherry picking". Thus, no subsets are allowed to be considered, meaning that if every analysis or SR is deemed combinable, all results will enter into the test statistic.

By not allowing subsets, we force the inclusion of results that support the null hypothesis if such results can be included. This is not an insignificant statement as it implies that we avoid including what would be a penalty term by forcing the inclusion of all available data. The proto-models model builder deals with model complexity by estimating an AIC-type penalty on the free parameters (see Section 8.1). If we were to select only the best result for a given proto-model point, then we would have to apply something similar. It is worth modelling such behaviour statistically and comparing it to the data to understand how negative weights affect a combination.

Suppose, for the moment, we ignore the hereditary condition in the WHDFS algorithm and consider the fraction of allowed combinations f_A to equal the probability of being allowed to add a result to the set at any one step. In that case, we can calculate the probability of finding a negative weight in any set of results.

- f_A: The probability of being allowed to add a result to the set, ignoring the cumulative hereditary condition for the moment.

- f_s: The probability of not being allowed to add the result to the set. $1 - f_A$
- $P(w^-)$: The probability of a negative weight.
- $P(w^+)$: The probability of a positive weight, $P(w^+) = 1 - P(w^-)$.
- l the length of a given set.
- n number of available results.

If we assume that the f_A and $P(w^+)$ are independent quantities, then the probability of finding a negative weight in a set of length l is:

$$\begin{aligned} P(\text{All positive}) &= P(w^+)^l \\ P(\text{At least one negative}) &= 1 - P(w^+)^l \,. \end{aligned} \tag{8.32}$$

The more complicated relation is the distribution and expectation of l. Let's consider the hereditary condition where the probability of being allowed to add a result decreases with the number of results already in the set. In terms of building sets of results, this condition introduces a dependency between the probability of adding an element to the set (f_A) and the current size. Thus, we need to define $f_A(l)$, the probability of being allowed to add the l-th result, which decreases as l increases. The probability of being allowed to add each successive result decreases exponentially as:

$$f_A(l) = f_A^l \,. \tag{8.33}$$

The length of a set is dependent on two factors: the probability of any one step being allowed $f_A(l)$ and the upper limit on the number of steps available n. Modelling this behaviour requires the definition of a Bernoulli-type variable:

$$\begin{aligned} X_i = 1 \quad &\text{(success) with probability} \quad f_A^{L_{i-1}} \,, \\ X_i = 0 \quad &\text{(failure) with probability} \quad 1 - f_A^{L_{i-1}} \,, \end{aligned} \tag{8.34}$$

where i runs from 1 to n and L_n is defined as

$$L_n = \sum_{i=1}^{n} X_i \quad X_i \in \{0, 1\} \,. \tag{8.35}$$

The distribution of L_n approximates the distribution of all set-lengths identified by the HDFS algorithm with no subsets allowed. Figure 8.8 shows the distribution of lengths as identified by the HDFS algorithm using a randomly generated BAM S with $f_A = 0.75$. The

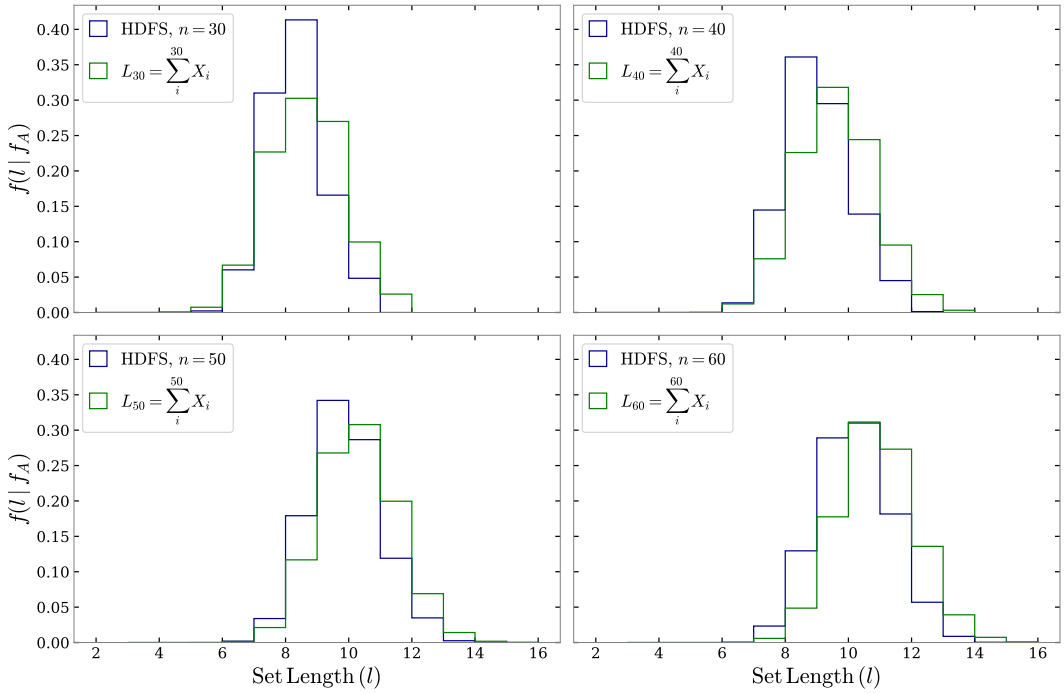
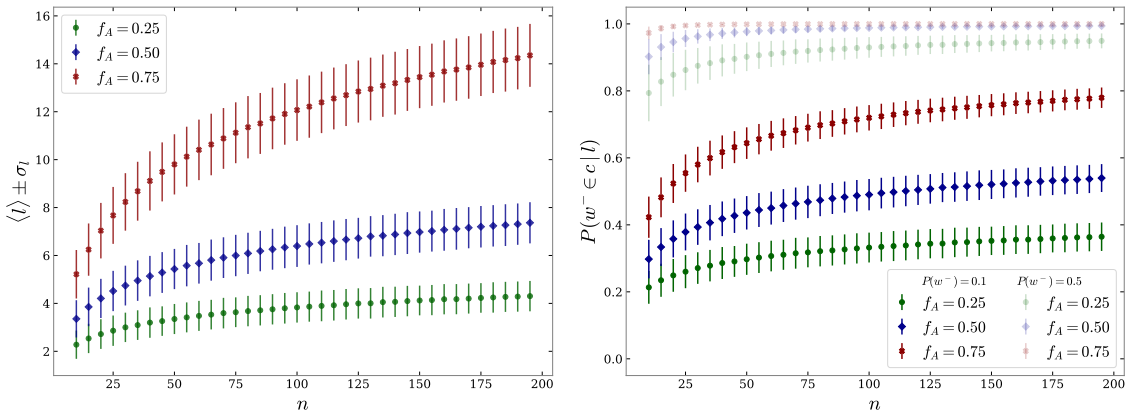


Figure 8.8.: The distribution of set lengths l calculated using HDFS algorithm and the predicted distribution with $l \sim L_n$ defined in Equation (8.35)

distribution of L_n is shown in green and was calculated for 10,000 trials. The plot shows the results running over four different BAM sizes with increasing elements (n). Figure 8.8 shows that n can be well approximated by the construction of Bernoulli-like trials L_n. Simulating the distribution for allowed subsets would require the evaluation of L_n from 0 to n as this follows the construction of the HDFS algorithm, which runs n graphs per $n \times n$ BAM, with each jth iteration having source $k_0 = j$ and sink $k_m = n$. The resulting distribution would be shifted to lower values and broadened to accommodate the extra values.

Figure 8.9 consists of two plots; the left-hand plot (a) shows the relationship between the BAM size and the mean length of all sets constructed with no subsets. The right-hand plot (b) converts the result from (a) via Equation (8.32), giving the expectation of negative weights per set chosen from n results. The results have been calculated at $P(w^-) = 0.1$ and $P(w^-) = 0.5$, knowing that for a typical proto-model point f_A is approximately 0.75 and $P(w^-) \approx 0.5$ and n is over 150-200 analyses and SR; thus we would expect to see all results containing at least one negative weight.

Up to now, we have been working under the assumption that f_A and $P(w^-)$ are independent, which is a fair approximation to first order; however, there is a relation



(a) **(b)**

Figure 8.9.: Relationship between the BAM size n and the average length l of the combinations c (a) and the probability of at least one weight in combination being negative $P(w^- \in c \mid l)$

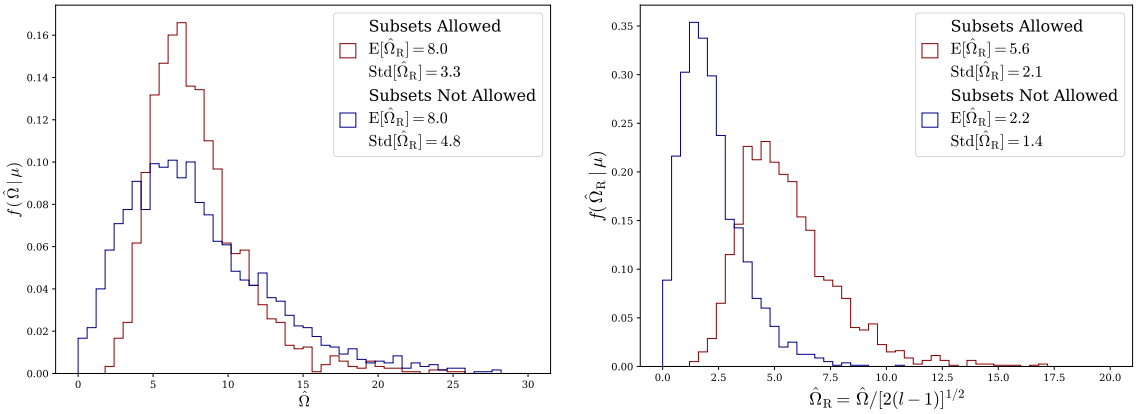
that requires acknowledgement. For this project, the BAM is constructed from analyses and SRs– instead of just SRs used in the TACO project. As discussed in Section 8.1, we assume that results from different experiments or energies are uncorrelated and thus combinable.[1] For analyses from the same experiment and energy, the combinability is assessed "by hand". The complication enters when choosing SRs from an analysis that contains results in favour of the proto-model and others that are either uninformative or in favour of the null hypothesis. In this case, the result favouring the proto-model will be chosen, excluding the alternatives from the set. The effect on the distribution of $\hat{\alpha}$ is shown in Figure 8.7, where $\hat{\alpha}_S$ is $\hat{\alpha}$ filtered such that results are either full statistical models or the best-performing SRs from a single analysis. If there are enough separate combinable analyses to counter this, then the effect would be minimised; however, this does introduce an intra-analyses bias for analyses without a statistical model.

This issue was still being addressed at the time of writing, and several possible solutions were under consideration. One option is to penalise the weight calculated for a single best SR in terms of its significance within the context of $\hat{\alpha}$ such that:

$$\text{Single SR Penalty} \propto \int_{,-\infty}^{\infty} \hat{\alpha} n P(\hat{\alpha} \mid \mu) \cdot \Phi(\hat{\alpha} \mid \mu)^{n-1} \, \mathrm{d}\hat{\alpha} - \mathrm{E}[\hat{\alpha}] \,, \tag{8.36}$$

where $P(\hat{\alpha} \mid \mu)$ and $\Phi(\hat{\alpha} \mid \mu)$ are the PDF and CDF, estimated from $f(\hat{\alpha} \mid \mu)$ (Figure 8.7), and n is the number of SRs in the original analysis. The integrand consists of $\hat{\alpha}$ multiplied

[1] A complete representation of the proto-models BAM is presented in Appendix B, Figure B.3



(a) $\hat{\Omega}$ **(b)** $\hat{\Omega}_\mathrm{R}$ Reduced

Figure 8.10.: Distribution of $\hat{\Omega}$ (a) and $\hat{\Omega}_\mathrm{R}$ (b) for toy data using a the likelihood $\mathcal{L} = \max_\xi\{\mathcal{L}_\xi\}$. The data was generated using MC events under the SM hypothesis (no signal). The best combinations were identified with subsets allowed (red) and not allowed (blue). The results show that allowing subsets biases the results towards shorter combinations.

by the distribution that describes picking the highest value from n random variates. Thus, the integral is the expectation value of the distribution. The expectation of $\hat{\alpha}$– which is approximately zero under the SM hypothesis– is then subtracted to give a penalty weighted by the expectation values of the single SR and the SR in terms of n. Interestingly, the bias for which this penalty is designed would have also affected proto-models version 1; the use of combinations in version 2 compounds this bias and, as such must be addressed before final publication.

One benefit to knowing this selection tendency in advance is that it allows for reducing the BAM size. For analyses evaluated on an SR level with no statistical model, only the most significant SR will be combined. Thus, additional SRs from the same analysis can be dropped from the BAM, as they would always be ignored in favour of the significant result. With a sufficient penalty for such cases, the selection should not affect the original distribution $f(\hat{\alpha} \mid \mu)$; in fact, the distribution of $\hat{\alpha}$ would be the appropriate test for such a penalty. This reduction was tested using the proto-model's ERM to generate multiple versions of the database over a selection of test proto-model points and running the WHDFS algorithm for both reduced and unreduced BAM. The results were matched for each run with the BAM size reducing from 150-200 results to around 60-80.

Assuming that the analyses are representative, running the WHDFS algorithm with no subsets seems to provide a self-regulating test statistic for finding the best-performing set

of results for a given proto-model point. When running the WHDFS algorithm without subsets, the PATHFINDER module will not allow the inclusion of negative weights, so an offset must be applied. The final weight is easily retrieved, providing the offset applied to the `remap_path` method of the WHDFS class. Negative weights can be used when allowing subsets, and as such, we can compare the results with and without the inclusion of subsets. The left-hand plot of Figure 8.10 shows the distribution of $\hat{\Omega}$ calculated using the toy model data with $\mathcal{L} = \max_\xi\{\mathcal{L}_\xi\}$ from Equation (8.26). The distribution of the test statistic $\hat{\alpha}$ (shown in Figure 8.7) had an expectation value of $\mathrm{E}[\hat{\alpha}] = 0.45$ with a standard deviation of $\mathrm{Std}[\hat{\alpha}] = 1.63$. The distribution of $\hat{\alpha}$ was generated using 1,000 bootstraps of 50 toy observables; for a single set of 50 observables, the fraction of negative weights (or the probability of choosing negative weight) was $P(w^-) = 0.40$. For the combined statistic $\hat{\Omega}$ in Figure 8.10, the average path length is 8.5, thus using Equation (8.32) the probability of at least one negative weight in a path is $1 - P(w^+)^l = 0.988$, or 98.8%. The proportion of combinations with at least one negative weight for the "subsets not allowed" histogram in Figure 8.10 is 99%. This is compared to 0.4% for the histogram with subsets allowed. On average, each combination contained 40% negative; this had the effect of pulling the distribution of $\hat{\alpha}$ towards zero. We can define this pull as:

$$\text{Pull} = \frac{\sum_i \Theta(-\hat{\alpha}_i) \cdot (-\hat{\alpha}_i)}{\sum_i \Theta(\hat{\alpha}_i) \cdot \hat{\alpha}_i}, \tag{8.37}$$

where $\Theta(\hat{\alpha}_i)$ is the heavy-side function applied to $\hat{\alpha}_i$, such that the equation gives the ratio of negative to positive values. Because $\hat{\mu}$ is optimized for the best combination $\hat{\Omega}$, its value is always greater than zero. Thus, the pull value must be between zero and one ($[0, 1)$), with zero corresponding to no negative weights and one corresponding to the complete cancellation of positive values. For Figure 8.10, the average pull was found to be $\mathrm{E}[\text{Pull}] = 0.33$ with a standard deviation of $\mathrm{Std}[\text{Pull}] = 0.16$. This confirms that applying the no-subset condition results in a shift to lower values as shown in Figure 8.10.

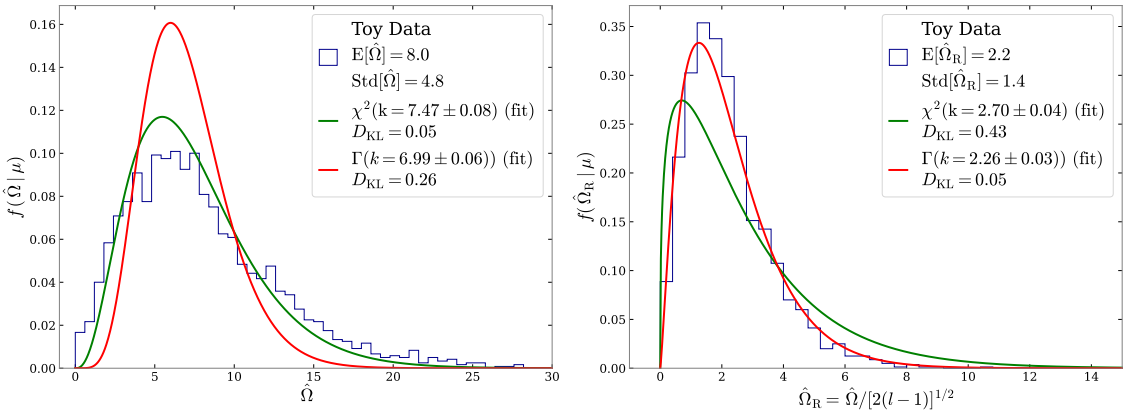
Beyond the single proto-model point, when comparing points in proto-model space, we require a measure of significance. Considering the discussion thus far, we have seen that the number of results entering a given set depends on n and f_A. Yet, as the *Walker* moves in the model space, selecting relevant SMS topologies will alter the number of analyses for each point due to their respective coverage. This brings us back to the extra DOF introduced by l in Equation (8.30). Considering the moments of $\hat{\alpha}$, we can use l to reduce $\hat{\Omega}$ and produce a consistent distribution independent of the number of results. Figure 8.7 showed that the expectation of $\hat{\alpha}$ is close to zero, and assuming the self-mediating effect of not allowing subsets in the final combination, we can assume

that the correction to the first moment vanishes under the SM hypothesis. The second moment, on the other hand, does not vanish and is proportional to the number of results that enter into a combination. If we consider the distribution of $\hat{\Omega}$ to be $\sim\chi^2$, then we can approximately, if not crudely, account for the second moment by dividing through by the square root of the χ^2 variance $(2k)$ where we replace the k with our new degrees of freedom $l-1$.

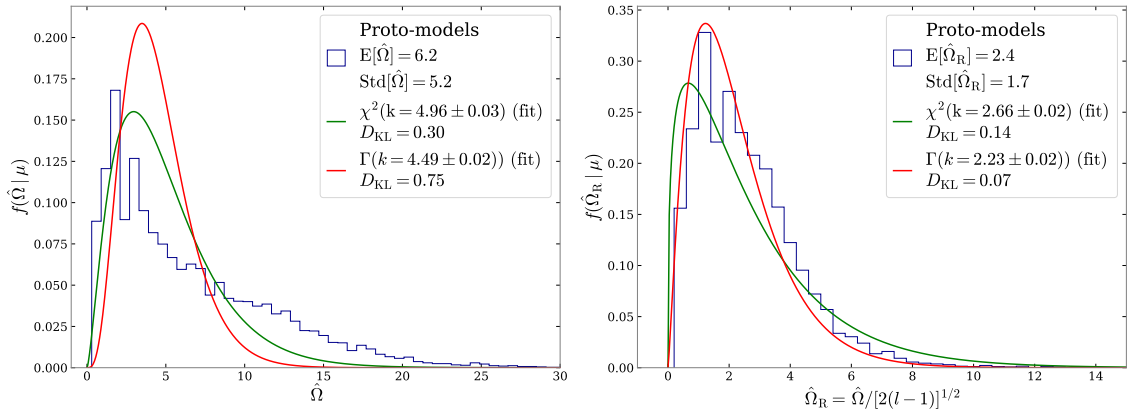
$$\hat{\Omega}_{\mathrm{R}} = \frac{\hat{\Omega}}{\sqrt{2k}} = \frac{\hat{\Omega}}{\sqrt{2(l-1)}}\,. \tag{8.38}$$

Here, the DOF is expressed in the standard form of k, and if a combination is defined as containing more than one result, the minimum rescaling factor becomes $\sqrt{2}$. It is reasonable to assume the rescaled $\hat{\Omega}$ would make a better test statistic for the combination selection, as it includes a penalty on the length. This would be incorrect as such a penalty would bias the WHDFS algorithm towards shorter combinations, less likely to contain results that balance the selection. Because the WHDFS algorithm is only provided with positive weights, it can only increase the final sum by adding more results. When we consider that the BAM and weight are also ordered by decreasing significance, the weights corresponding to a negative test statistic will likely appear in the initial iterations and, thus, in the longest sets (this being a primary feature of the depth-first search algorithm). The left-hand plot of Figure 8.10 shows the subset analysis with the new rescaled test statistic. The difference between the inclusion and exclusion of subsets becomes more apparent as the inclusion favours shorter sets, dropping negative weights. Figure 8.11 shows the distributions of $\hat{\Omega}$ and $\hat{\Omega}_{\mathrm{R}}$ for both toy and proto-model simulations, using data generated under the SM-only hypothesis. The toy data was generated using 2,000 bootstraps of 50 toy observables (like those shown in Figure 8.5). Plot (b) shows the same analyses but performed on proto-model data, which was created by generating 200 SM-only pseudo-databases for 15 proto-model test points (available at https://smodels.github.io/protomodels/). The data is shown in the histograms along with a fit to χ^2 and Γ distributions with associated errors estimated via the Hessian matrix (see Section 4.4). For the proto-models results, 41% of all sets contained at least one damaging result, translating to a pull fraction of 0.2 ± 0.19 from positive. This is a lower value than what was predicted from the toy simulations but not unexpected, considering the intra-analysis bias previously identified.

Although we would expect the distribution to behave as χ^2, the actual distribution is better described by the more general $\Gamma(\theta = 1, k)$ distribution. In Figure 8.11 the



(a) Toy ($\mathcal{L} = \max_\xi\{\mathcal{L}_\xi\}$)



(b) Proto-models

Figure 8.11.: Distributions of $\hat{\Omega}$ and $\hat{\Omega}_\mathrm{R}$ for the toy (a) and proto-model (b) simulations, using data generated under the SM-only hypothesis The distributions are shown pre- and post-reduction ($\hat{\Omega} \to \hat{\Omega}_\mathrm{R}$). The plots show the best-fit results from a single parameter χ^2 and Γ distribution fit where the θ parameter for the Γ distribution is fixed to one. Each fit has a corresponding Kullback–Leibler divergence D_KL calculated using the normalised histogram bin counts

Kullback–Leibler divergence D_KL is used to quantify the "quality" of fit and shows that the Γ distribution (with the scale parameter θ fixed to one) performs better for the reduced data. This is not entirely surprising, given the assumptions we have made with the optimisation step. However, the agreement between the distribution fitted for the toy and the proto-model simulation can't be ignored. The agreement does suggest asymptotic behaviour for the rescaled statistic. Figure 8.12 takes the data from plot (b) of 8.11 and breaks the $\hat{\Omega}_\mathrm{R}$ distribution down by length l. The results show that the $\Gamma(\theta, k)$ fit does vary over different lengths, but the approximation still holds for the purpose of a significance test.

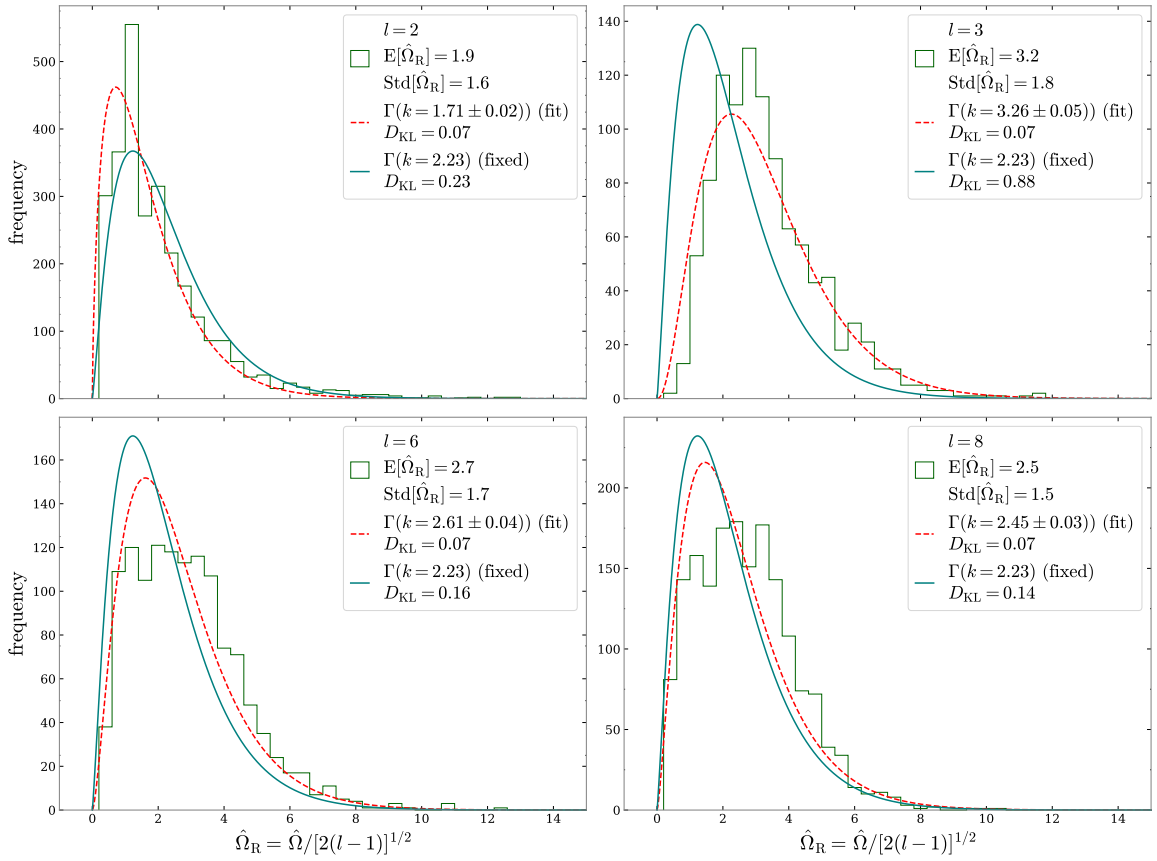


Figure 8.12.: Distributions of $\hat{\Omega}_\mathrm{R}$ for proto-model pseudo-database simulations, using data generated under the SM-only hypothesis. The distributions have been separated by the length of the result set identified by the WHDFS algorithm. The plots show the best-fit and fixed results from a single parameter Γ distribution. The θ parameter for the Γ distributions is fixed to one. Each fit has a corresponding Kullback–Leibler divergence D_KL calculated using the normalised histogram bin counts

8.2.6. Version 2 – analysis chain

With a test statistic defined such that a significance could be estimated for a given proto-model point, we can now look at how this fits into the analysis chain. Some important changes have been made since version 1 to accommodate the changes in the SModelS framework and database. In this section, we will examine these changes and how they affect the analysis chain.

Particle	3-body decay channels	Relative ratios (if fixed)
$X_Z^{j\neq 1}$	$q\bar{q}\,X_Z^k$, $\ell^+\ell^-\,X_Z^k$, $\nu\bar{\nu}\,X_Z^k$	0.7, 0.1, 0.2
	$q\bar{q}'\,X_W^i$, $\ell\nu_\ell\,X_W^i$	0.68, 0.11
X_W^i	$q\bar{q}'\,X_Z^k$, $\ell\nu_\ell\,X_Z^k$	0.68, 0.11
X_W^2	$q\bar{q}\,X_W^1$, $\ell^+\ell^-\,X_W^1$, $\nu\bar{\nu}\,X_W^1$	0.7, 0.1, 0.2

Table 8.2.: Three-body decay channels for the X_Z and X_W particles and their relative ratios (for given k, i) when assuming that they proceed through off-shell Z- or W-boson decays. In practice, at least in the first step, only the channels in blue need to be implemented (for $k = 1$).

Particle content

A specific aim for the second version of the proto-modelling project is to see whether the algorithm singles out the small excesses observed in electroweakino (Wino) analyses, and what are the preferred mass relations. One development required for this is the inclusion of three-body decays for the X_Z and X_W particles. These may or may not assume off-shell W-and Z-boson decays: in the former case, the hadronic/leptonic branching ratios will be fixed (and the mass differences limited); in the latter case, leptonic and hadronic three-body decays will be separate degrees of freedom. In practice, the choice should make little to no difference: at present, the experimental results in the off-shell region only constrain leptonic decays.

In addition to the 20 BSM states and their decay channels considered for constructing proto-models in Table 8.1, we include three-body decays of X_Z and X_W, which are given in Table 8.2. However, not all are needed to characterise the soft-lepton excesses seen in Wino analyses. First, as mentioned, no hadronic results are available in the low-mass compressed regions where the leptonic excesses reside. So we may disregard $q\bar{q}\,X_Z$ and $q\bar{q}\,X_W$ final states in the first step. Second, the decay into $\nu\bar{\nu}\,X_Z$ or $\nu\bar{\nu}\,X_W$ is invisible; the only effect of this channel would be to introduce an additional mass scale. This might be of interest if, e.g., different experimental results pointed to different LSP masses, but given the current uncertainties, this is not relevant. Therefore, the three-body channels to implement in the first step are:

$$X_Z^{2,3} \;\to\; \ell^+\ell^-\,X_Z^1\,,\;\; \ell\nu_\ell\,X_W^1\,; \tag{8.39}$$

$$X_W^{1,2} \;\to\; \ell\nu_\ell\,X_Z^1\,. \tag{8.40}$$

The interest of considering also X_Z^3 and X_W^2 instead of only X_Z^2 and X_W^1 is that this will allow us to see to which extent different excesses are consistent (or need the introduction of different particles to explain them). If the hadronic ($q\bar{q}$) and the neutrino ($\nu\bar{\nu}$) modes are also added, and the relative ratios fixed to those of Z- or W-boson decays as indicated in Table 8.2, this can be used for predicting complementary signals. However, the mass difference between mother and daughter $X_{W,Z}$ should be restricted to the off-shell region. Finally, note that as in [233], X_W and X_Z decays into $X_\ell + \ell/\nu$, $X_\nu + \nu/\ell$, and $X_{W,Z} + \gamma$ are not yet taken into account.

Analyses and likelihoods

As mentioned in Section 8.1.3, the extent to which the likelihoods can be calculated depends heavily on the information the experimental collaborations provided. Since version 1, the SModelS database has expanded to include a total of 157 analyses, with 7,736 individual efficiency maps from 1,529 distinct signal regions covering 130 different SMS topologies [217]. This increase has shifted the ratio of UL- and EM-type results in favour of EM-type, meaning that the likelihood approximation using a truncated Gaussian can be removed from version 2. This is not to say that UL-type results will no longer be used, as they still provide a sanity check for prior exclusion.

There has also been a move towards experiments providing correlation information or full statistical models with the search analyses. This has shifted the likelihood evaluations to *simplified likelihoods v1 and v2* [255] and pyhf [187].

Much work has been done under the premise that constructing a global likelihood from multiple results is reasonable so long as the results are from distinct experiments and taken from different LHC runs. We can't say that the results are entirely uncorrelated; one such source of overlap may be the PDFs used to model background estimates. If such methods are shared between experiments, then this would correlate with the results. However, this would be a 2nd or 3rd-order effect, and as such, it can be neglected. For version 2, the combinability of analyses and SRs is still conducted on a "by-eye" basis. This maintains the original approach, identifying results with different final states in their signal regions (e.g., fully hadronic final states vs final states with leptons).

Test statistic K^c

Before exploring the new *Walker* algorithm, it is worth updating the definition of the global test statistic K. During the MCMC-type walk, K is maximised and, as such, is designed to increase for proto-models that satisfy the given constraints. Thus, for a given proto-model with an optimised combination of results $c \in C$, the auxiliary quantity K^c is now defined as:

$$K^c = \sqrt{\frac{2}{l-1}} \ln\left(\frac{\mathcal{L}_c^{\text{obs}}(\hat{\mu}_s, \hat{\boldsymbol{\theta}})}{\mathcal{L}_c^{\text{obs}}(\mu = 0, \hat{\hat{\boldsymbol{\theta}}})}\right) + 2 \ln\left(\frac{\pi(\text{BSM})}{\pi(\text{SM})}\right) = \hat{\Omega}_R + 2 \ln\left(\frac{\pi(\text{BSM})}{\pi(\text{SM})}\right) \quad (8.41)$$

where we have inserted the definition of the test statistic $\hat{\Omega}$ into Equation (8.6) and rearranged the terms accordingly. The $\pi(\text{BSM})$ and $\pi(\text{SM})$ are the priors for the proto-model and SM, respectively. Defining the priors is still crucial to formulating the test statistic as it introduces a penalty for model complexity. For the SM, a flat prior was assumed ($\pi(SM) = 1$) as the SM is a common factor for all combinations and, as such, does not affect the comparison between different proto-models. Equation (8.7) is still used for the BSM prior, where $n_{\text{particles}}$, n_{BRs} and n_{SSMs} are the number of new particles, branching ratios and signal strength multipliers, respectively. The parameters a_1, a_2, and a_3 are still being optimised to approximate the AIC penalty condition, but the initial values are 2, 4, and 8, respectively.

Architecture

With everything covered thus far, the most significant change moving from version 1 to two is inevitably the *Walker* algorithm. Figure 8.13 shows the full procedural flow chart for the new algorithm, demonstrating the entire iterative cycle. For efficient visualisation, the randomised parameter selection nodes from Figure 8.1 are contained within the "Go to new point" node of Figure 8.13. The test statistic optimisation loop ($\hat{\Omega}$ calculation) is contained within the blue-shaded area in the upper part of the diagram. This loop has a limit of five iterations before a new proto-model is chosen. If the $\hat{\mu}$ is successful, then the global test statistic K^c is computed and compared to the previous iteration.

If the value of K^c improves on the previous iteration, the “fast critic” is employed. This method uses the available UL-type data to check that the proto-model point isn’t excluded by the available data. If K^c is not an improvement on the previous point, then

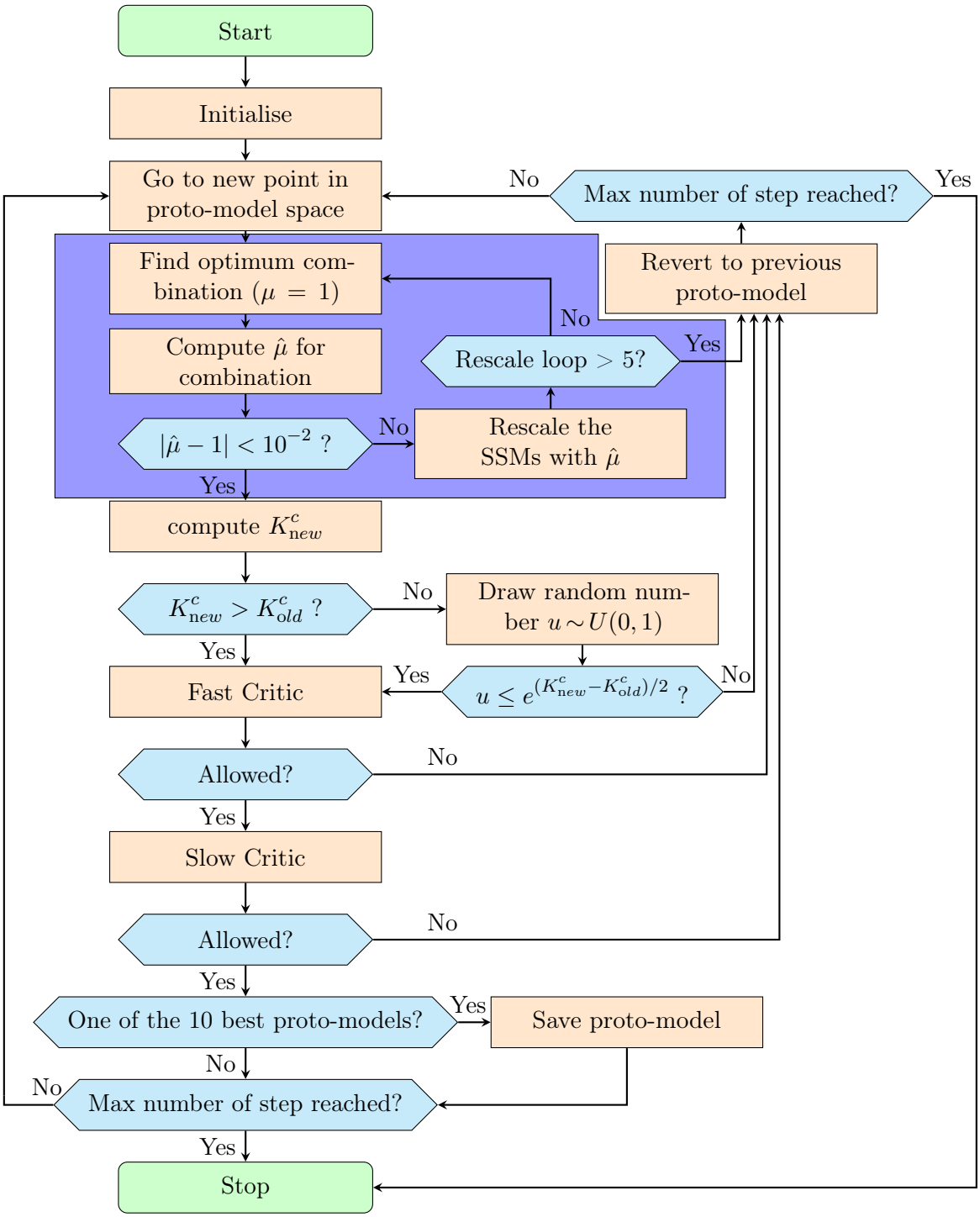


Figure 8.13.: Procedural flowchart of the code architecture. See the text for details. The blue-shaded area contains the test statistic optimisation loop for $\hat{\Omega}$ (Section 8.2.4)

there is still a chance of moving on to the next step if the following condition is met

$$u \leq e^{(K^c_{\text{new}} - K^c_{\text{old}})/2} \quad \text{where} \quad u \sim U(0,1)\,. \tag{8.42}$$

If all checks are passed, we introduce an adversarial counter process that employs a TACO-like procedure to calculate a combined 95% ($r = 1$) exclusion limit for the proto-model point. The method uses the same data available for the calculation of $\hat{\Omega}$ but the weights are changed to $\omega_i = -2\ln(\mathcal{L}_{\text{SM}}/\mathcal{L}_{\text{BSM}})$, as defined in Equation (7.7). Following the procedure in Chapter 7, the combinations are identified using weights calculated from theory expectations, but the final sum is calculated using the observed data. In other words, the combinations are selected using theory, but the exclusion is determined from observations. This adversarial step uses the techniques developed in the TACO project to get the most exclusionary "power" out of the available data. If the proto-model is not excluded, it can be included in the top 10 performing points.

The definitive significance test is not represented in Figure 8.1. Before any assertions regarding significance, it is necessary to assess the test statistic K according to the prescribed methodology for simulating the distribution of the estimator $\hat{\Omega}$. This process will entail generating multiple pseudo-databases and calculating the distribution of K at the proto-model point under the standard model (SM) hypothesis. The value of K demonstrating the highest performance will ultimately constitute the best result.

8.2.7. Preliminary results

At the time of writing, the new *Walker* algorithm is in the early stages of testing. The analysis chain is fully constructed and in operation, but the focus is currently on testing the design and understanding the initial results.

Nevertheless, there have been some preliminary results which we can discuss. The current best-performing proto-model has obtained a test statistic value of $K = 8.47$ and was generated in step 3 of the 95th trial run. The "winning" proto-model consisted of a X_t, X_g, X_Z^2 and X_Z^1 with masses of 1,267, 741 and 578 and 317 GeV, respectively. This produced signals in the $\ell\ell + \text{jets} + E_{\text{miss}}$, $b\bar{b} + E^{\text{T}}_{\text{miss}}$, $t\bar{t} + E^{\text{T}}_{\text{miss}}$ and $\text{jets} + E^{\text{T}}_{\text{miss}}$ final states with SUSY-like cross sections. The mass results are shown in Figure 8.14, along with the contributing analyses.

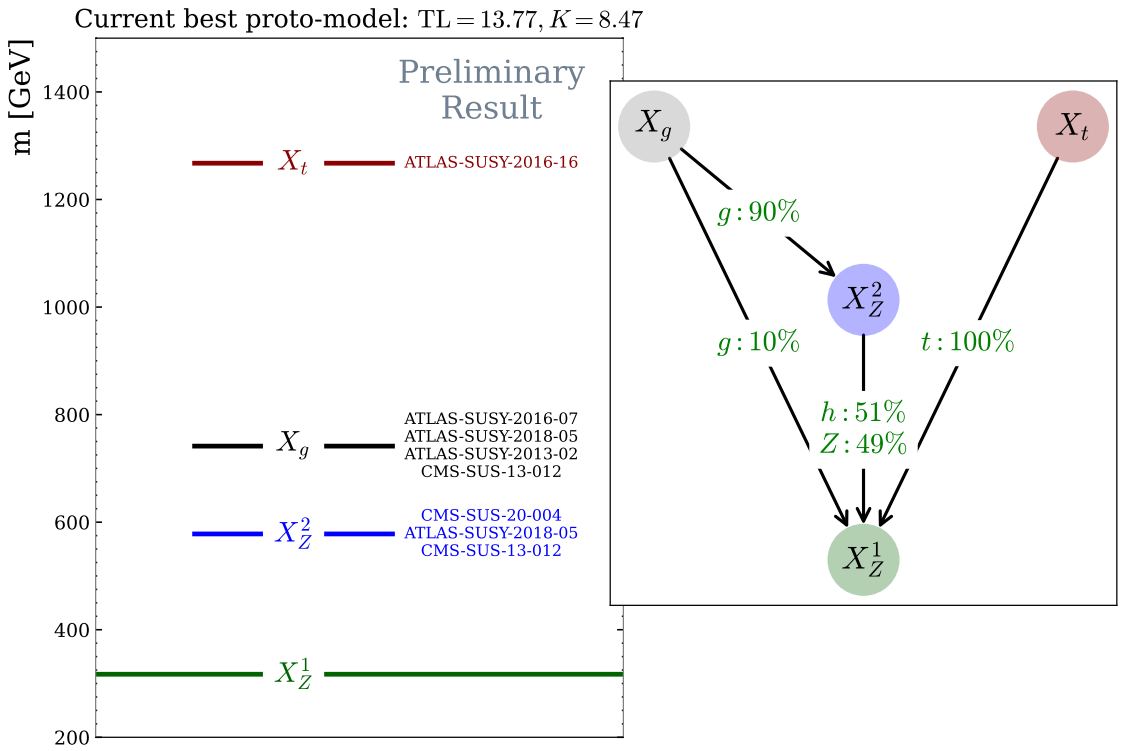


Figure 8.14.: Preliminary result from an early run of the proto-models *Walker* algorithm. The left-hand plot shows the particle type, mass values and contributing analyses, while the right-hand plot shows the decay channels and branching ratios.

The effective signal strength multipliers were found to be $\hat{\mu} \times \kappa_{X_t \tilde{X}_t} \approx 1.1$ and $\hat{\mu} \times \kappa_{X_g X_g} \approx 1.3$, and $\hat{\mu} \times \kappa_{X_Z^2 X_Z^2} \approx 4.6$, corresponding to $\sigma(pp \to \tilde{X}_t X_t) \approx 1.2\,\text{fb}$, $\sigma(pp \to X_g X_g) \approx 3981\,\text{fb}$ and $\sigma(pp \to X_Z^2 X_Z^2) \approx 52.2\,\text{fb}$ at $\sqrt{s} = 13T\text{eV}$. the X_t directly decay to the lightest state (X_Z^1) with 100% BR. However, the X_g decays directly to the X_Z^1 with only a 10% BR; the preferred route is via X_Z^2, which then decays to the (X_Z^1) via the ZX_Z^k or hX_Z^k channels. Table 8.3 gives a full breakdown of the analysis contributions.

Although this is a preliminary result, we can make some interesting observations. The first and most obvious is the cascade decay of the X_g through an intermediate state X_Z^2. Such proto-models were not considered in version 1 as there were not enough results in the database covering this type of process. If nothing else, this is a promising start that indicates a shift to more complex models.

A further observation is that the proto-model serves as an empirical representation of the observed phenomena. When aligned with fundamental theories – candidates for the Next Standard Models – many potential ambiguities must be considered. For example,

Analysis Name	Type	Topology	Obs	Exp	$\approx \sigma$	Particles
ATLAS-SUSY-2016-07 [259]	EM	T2	611	526 ± 31	$2.2\,\sigma$	X_g, X_Z^1
CMS-SUS-20-004 [260]	Comb	TChiHH	0.3 fb	0.1 fb	$2.1\,\sigma$	X_Z^1, X_Z^2
ATLAS-SUSY-2016-16 [261]	EM	T2tt	8	3.80 ± 1.00	$1.7\,\sigma$	X_Z^1, X_t
ATLAS-SUSY-2018-05 [262]	EM	T6ZZ	8	3.38 ± 1.64	$1.6\,\sigma$	X_g, X_Z^1, X_Z^2
CMS-SUS-13-012 [263]	EM	T2, TChiZZ	32	22.80 ± 5.20	$1.3\,\sigma$	X_g, X_Z^1, X_Z^2
ATLAS-SUSY-2013-02 [264]	EM	T2	133	125 ± 10	$0.6\,\sigma$	X_g, X_Z^1

Table 8.3.: Breakdown of results contributing to the latest preliminary results from the proto-model *Walker*

the X_g particle, initially posited as the partner particle of the gluon, could be substituted by a quark due to their experimentally similar signatures. In a similar manner, the X_Z^2 could, with additional data, potentially be attributed to two or more mutually resembling fundamental particles.

From a BSM standpoint, this result is not inconsistent with SUSY with all signal strength multipliers of order unity. If it were the case that the signal strength multipliers with $\hat{\mu} \times \kappa \gg 1$, the resulting small cross sections could imply same-spin scenarios like those predicted by extra dimensions [265–267]. As previously mentioned, having the production cross section as a free parameter while leaving spin undefined gives the *Walker* immense freedom to explore the potential proto-model space. Seeing early indications of SUSY-like results in the signal-strength multipliers is a promising start. A full breakdown of the result discussed in this section can be found at https://smodels.github.io/protomodels/jamie300/.

Chapter 9.

Conclusions

The primary aim of this study was to identify selection techniques for the optimal meta-analysis of BSM physics. This was done by addressing three specific research questions, including (i) *how to select an optimum set of minimally overlapping results*, (ii) *How to implement such a selection technique for model exclusion*, and (iii) *How to implement such a selection technique for anomaly detection*? Through a systematic exploration of these questions, the findings show that through the application and modification of graph-based selection methods, a problem of order $\mathcal{O}(2^n)$ complexity can be reduced to approximately $\mathcal{O}(n \ln n)$. Such a solution reduces the combinatorial challenge of selecting the optimum set of results, which will only get more relevant as we move into the era of high-luminosity LHC.

9.1. Summary of findings

Starting with the HDFS algorithm, in Chapter 6, the HDFS is presented as a potential solution to the problem of choosing an optimum subset of results from 2^n possible combinations. The algorithm is a modification of the well-known depth-first search. The modification excludes nodes prohibited from combination recursively, so the number of available combinations reduces with depth. The BAM matrix, an n by n symmetric matrix that provides a binary description of the allowable combinations, was crucial to evaluating the available combinations. The HDFS algorithm was shown to evaluate all available combinations efficiently. However, it was soon realised that a more efficient method could be devised by sacrificing the requirement to identify all combinations.

The WHDFS algorithm was developed to use weights to optimise the HDFS method. This was achieved by taking advantage of the DFS ordering and tracking the best combination weight. (The "best" here is assumed to be a simple sum of weight, but alternatives could be defined in place). The crucial realisation was that the algorithm could be "short-circuited" by comparing the current best combination weight with an upper limit to the weight available to any future combination in the iteration. This upper limit vastly reduced the number of iterations per evaluation and, thus, reduced the time complexity.

Before fully deploying the HDFS and WHDFS algorithms, it was noticed that specific behaviour was required for different tasks. This led to the development of subset conditions where we defined a priori whether or not subsets were to be included when evaluating a result. This condition mainly affected some edge cases but was an important consideration when dealing with negative weights.

The first application of the WHDFS algorithm in the TACO project where search analyses SRs were combined to improve pre-existing exclusion limits. The project included–among others–authors from the SModelS [184] and MadAnalysis 5 [173] collaborations and made extensive use of both frameworks. Much of the project was dedicated to estimating the overlaps between different SRs from analyses considered to share events. To achieve this, we defined the overlap matrix, estimated by generating BSM events populating a well-defined region in a mass-parameter space constructed from the union volumes covered by the SMS topologies and SRs. By tracking which events populated bins shared by different SRs, we could estimate a correlation-type matrix defined as the overlap matrix.

By specifying a minimum threshold for the allowed overlap between different SRs, we could define the BAM required for the WHDFS algorithm. The NLLR of signal expectation with $\mu = 1$ over the maximised SM expectation with $\mu = \hat{\mu} = 0$ was used for the weights as the sum over a combination provided a well-defined test statistic. The combination analysis was compared to results for two SMS topologies (T1 and T1tttt), model points from the ATLAS 2015 pMSSM-19 scan [235] and 3 t-channel dark matter models. In all cases, there was an evident increase in exclusionary power reinforced by a statistically rigorous treatment of the overlap estimation.

The final project considered in this work is the ongoing research project called proto-models, which examines combinatorial methods to enhance anomaly detection. In many ways, this work is an extension of the TACO project into anomaly detection. The project

continues on previous work done by the SModelS team, presented in the paper "Artificial Proto-Modelling: Building Precursors of a Next Standard Model from Simplified Model Results" [233].

The work presented in this study makes two significant contributions to the proto-models project via two distinct applications of the WHDFS algorithm. In moving the primary objective from model exclusion to anomaly detection, much work had to be done to identify a test statistic capable of being used as an additive weight and measure of significance. This involved understanding the asymptotic distribution of the chosen statistic and how the WHDFS algorithm influenced the distribution of the combined weight. By leveraging model selection techniques and treating the number of results in a combination as a degree of freedom, we could demonstrate the asymptotic behaviour of the test statistic $\hat{\Omega}$. This provided a metric from which an approximate significance could be calculated.

The proto-models MCMC *walker* algorithm also uses a TACO-like adversarial counter metric to critique chosen proto-models. This process follows the exclusion procedure in the TACO project intending to prevent the selection of a proto-model driven by fluctuations present in a database that could be used–via a different metric–to exclude the same point.

At the time of writing, the proto-models project is in the testing and validation stages. However, some preliminary results are presented, containing particles with strength multipliers not inconsistent with SUSY-like expectations. There is also an increase in model complexity due to the increased pool of results in the SModelS database.

9.2. Implications

The work presented in the study is divided into three main parts: algorithmic, statistical, and applications in BSM physics. Admittedly, there is a significant overlap in each part due to the concurrent development and application. However, the implications of this work can be separated into these categories.

The HDFS and WHDFS algorithms presented as part of the PathFinder module provide a specific utility separate from the physics applications. As presented in Chapter 6, the algorithms aligned well with feature selection in machine learning. The construction of the BAM is not limited to correlations but could be extended to joint mutual or Fisher

information. This flexibility means that the algorithm has the potential to be used in many contexts, extending beyond particle physics applications.

Moving onto the implications of the TACO project [194] and leaving aside the SR selection, the development of the overlap matrix provided a solid foundation for future development. Evaluating the overlap matrix was by far the most computationally expensive strategy in the project. However, the project proved the traceability of such computation and demonstrated the result to be a statistically reasonable measure of the events shared between SRs.

One unexpected result shown in this study was the potential to force an unbiased data selection with the HDFS and WHDFS algorithms. For the case of model comparison, where the weights measure the agreement with–or deviation from–a null hypothesis, one can expect the weights to include negative values. In such a case, including as much available data as possible is essential to counter the influence of inevitable statistical fluctuations. Providing the WHDFS algorithm with weights, shifted to be positive, and forcing the non-consideration of any extendable set, i.e. not allowing subsets, the WHDFS will select the highest scoring complete set of results. This was a powerful realisation as this not only suggested a self-regulating metric, but the regulation improved with scale.

From the standpoint of applications in BSM physics, the findings of this study underscore the critical role of reinterpretation within the realm of particle physics. This significance is particularly pronounced when acknowledging the dependence on experimental data to facilitate the recasting of results to alternative models. The combinatorial methods outlined articulate how numerous SRs and analyses can be systematically assessed to identify the optimal subset. However, this endeavour would not have been achievable without utilising reinterpretation frameworks such as SModelS [184], MadAnalysis 5 [173] and Rivet [268].

9.3. Limitations

A consistent theme of this study was how to make the most of the data available from the LHC. To this end, some assumptions and approximations have been made to accommodate the complexity of the task. A good example is the potential effect of control regions in estimating the overlap. This issue was given a great deal of consideration, but there was no straightforward solution. It was eventually determined that the effect would

lessen the exclusionary power of the combined results; thus, any affected result would be conservative in estimation. This was an acceptable trade-off between traceability and accuracy.

One of the most challenging aspects of the study was the development of the proto-models test statistics $\hat{\Omega}$ and $\hat{\Omega}_{\mathrm{R}}$. The nested dependencies on the BAM size n, the allowed fraction f_A and path length l, along with the distribution of weights that entered into the combination, meant that there was no closed form for the distribution of $\hat{\Omega}$. The solution described in Section 8.2.4 was found using independent simulations of the proto-models analysis chain, generating results under the SM hypothesis and replicating the relevant profiled likelihoods. The final distribution was best approximated by a $\Gamma(\theta, k)$ with a scale parameter $\theta = 1$ and shape parameter $k = 2.23$. This relation was confirmed in both toy and pseudo-database simulations; however, the reason for the exact form is still unknown. As covered in Chapter 3, the Gamma distribution $\Gamma(\theta, k)$ is a generalisation of the exponential family of continuous probability distributions, which includes the χ^2 distribution, which is a particular case with $\theta = 1/2$. Considering $\hat{\Omega}$ was expected to have a χ^2 distribution [228], the emergence of a more generalised form is not unreasonable.

The final and most general limitation of the work presented here is that any significant deviation from SM expectations should be used to inform future search analyses and be understood in the context in which they are presented. Any assumptions or approximations made were made with significant consideration, and a conservative strategy is always assumed to be the most appropriate.

9.4. Recommendations

Moving into the era of HL-LHC, it is more important than ever to develop reinterpretation tools capable of handling the increasing number of analyses. The work presented in this study was conducted using search analyses, but increasingly reinterpretation software such as CONTUR, "Constraints On New Theories Using Rivet" [175], probes BSM theories using measurements. This approach takes advantage of the model independence associated with particle-level differential measurements conducted within fiducial regions of phase space. Consequently, these measurements can be compared to BSM physics scenarios simulated in Monte Carlo generators in a very general manner, thereby enabling the exploration of a broader range of final states than is usually encountered. Incorporating a TACO-like combination strategy into this framework would likely necessitate an SM

background estimate of the overlap matrix. This requirement would entail implementing efficient sampling strategies for the MC generators; however, it remains well within the capabilities of contemporary tools.

In more general terms, this study demonstrates that the HDFS and WHDFS algorithms should be used whenever appropriate. Whether it be for the purpose of data selection or optimisation, this work has demonstrated a variety of use cases. The success of these algorithms has been due to the combination of efficiency and scalability.

Given the recent engagement of experiments within the reinterpretation community, it is plausible to anticipate that the collection of analyses will only increase in the forthcoming years. This expectation is further supported by the recent enhancements in providing full statistical models. Therefore, it is reasonable to conclude that the pool of available data is set to grow, necessitating adaptations in the WHDFS algorithm. Certain algorithmic complexities have prevented the development of a multi-threaded/processed implementation; however, there are plans to implement such functionality in the near future.

The proto-models project is currently underway, and the results presented in this study are to be regarded as preliminary. Notable issues, including the single SR selection bias, have been identified during the preparation of this work, alongside a potential resolution. As the project progresses into its operational phase, there are still opportunities to extend the range of particles under consideration. The motivation for this would come from the fact the analyses with the most significant excess come from searches for mono-jet with large missing E_T^{miss}, specifically concerning the TRV1 and TRS1 topologies associated with vector and scalar mediators (X_V, X_S) [269, 270]. There is also the possibility of looking for log-lived–metastable–charginos and gluinos.

9.5. Final remarks

This thesis has advanced the methodological framework for optimising meta-analyses in the context of beyond the Standard Model (BSM) physics by addressing complex selection challenges in collider data reinterpretation. Through the development and application of novel algorithms, namely the HDFS and WHDFS algorithms, this work demonstrated that graph-based methods can effectively reduce the combinatorial complexity of selecting minimally overlapping results from $\mathcal{O}(2^n)$ to approximately $\mathcal{O}(n \ln n)$.

These algorithms have thus proven instrumental in systematically identifying optimal combinations for model exclusion and anomaly detection, providing a scalable and computationally efficient approach that addresses both theoretical and practical concerns inherent in high-luminosity Large Hadron Collider (HL-LHC) data.

The findings underscore the significance of combining exclusion limits in collider-based searches and point to potential applications beyond particle physics. By facilitating a rigorous and statistically robust framework, this work has laid the foundation for future implementations that extend to machine learning and feature selection contexts, highlighting the algorithms' adaptability across domains. Notably, the results emphasize the need for continued development of reinterpretation tools as the volume of available data and analyses increases in the HL–LHC era, a trend that underscores the broader relevance and potential impact of this research within both the particle physics community and data-intensive scientific fields.

In conclusion, this thesis contributes meaningfully to the ongoing efforts within particle physics to refine model selection methods and optimize data utilisation. It reaffirms the importance of reinterpretation frameworks and innovative algorithms in navigating the complexities of modern collider data, paving the way for more refined and statistically robust approaches to BSM discovery and anomaly detection in future research.

Appendix A.

Pseudocode

A. HDFS algorithm

The pseudocode shown in Algorithms 1 and 2 are written in a Pythonic syntax as the code makes use of the generator – denoted by the term *Gen()* – functionality, which allows for efficient iteration ordering. Aspects of the code are heavily influenced by the "all simple paths" method from the Python NetworkX package [193].

Algorithm 1 Hereditary Depth-First Search (HDFS)

Require: source = i
  Target = n
  Cutoff = $n - 1$
  Visited = $[i]$
  Stack = $[Gen(A_i)]$
  $\mathcal{S} = [A_i]$
  while Stack is not empty **do**
    Children = last element of the stack
    c = Next element in Children **or** None **if** Empty
    if c is None **then**
      Drop last element of Stack
      Drop last element of $\mathcal{S}$
      Drop last element of Visited
    else if length(Visited) < cutoff **then**
      if c = Target **then**
        Yield: Visited + [c]
      end if
      Visited += [c]
      if target not in Visited **then**
        $\mathcal{S}_c = A_c \cap S_{c-1}$
        Stack += [Gen(j: **for** index in A_c **if** index $\in \mathcal{S}_c$)]
      else if **then**
        Drop last element of Visited
      end if
    else if length(Visited) = cutoff **then**
      Drop last element of Stack
      Drop last element of $\mathcal{S}$
      Drop last element of Visited
      Yield: Visited + [Target]
    end if
  end while

B. WHDFS algorithm

The pseudocode shown in Algorithm 2 is a modification of Algorithm 1. WHDFS uses the edge weights to calculate an upper limit of the total weight available at each step in the path. This modification eliminated the need to explore all allowed paths instead of limiting the combinations to those with the greatest potential.

Algorithm 2 Weighted Hereditary Depth-First Search (WHDFS)

Require: source = i
Require: maximum weight
  Best Path = [], Target = n, Cutoff = $n - 1$
  Visited = $[i]$, Stack = $[Gen(A_i)]$, $\mathcal{S} = [A_i]$
  while Stack is not empty **do**
    Children = last element of the stack
    c = Next element in Children **or** None **if** Empty
    if c is None **then**
      Drop last element of Stack
      Drop last element of $\mathcal{S}$
      Drop last element of Visited
    else if length(Visited) < cutoff **then**
      if Target in Visited **then**
        continue
      end if
      $\mathcal{S}_c = A_c \cap S_{c-1}$
      Visited += [c]
      Current Weight = weight function (Visited)
      Available Weight = weight function ($\mathcal{S}_c$)
      if c = Target & Current Weight > Max weight **then**
        Max weight = Current Weight
        Best Path = Visited
      end if
      if (Current Weight + Remaining Weight) > Max weight **then**
        Stack += [Gen(j: **for** index in A_c **if** index $\in \mathcal{S}_c$)]
      else if **then**
        Drop last element of Visited
      end if
    else if length(Visited) = cutoff **then**
      Drop last element of Stack
      Drop last element of $\mathcal{S}$
      Drop last element of Visited
    end if
  end while
  return Best Path

Appendix B.

Overlap matrices

A. TACO overlap matrices

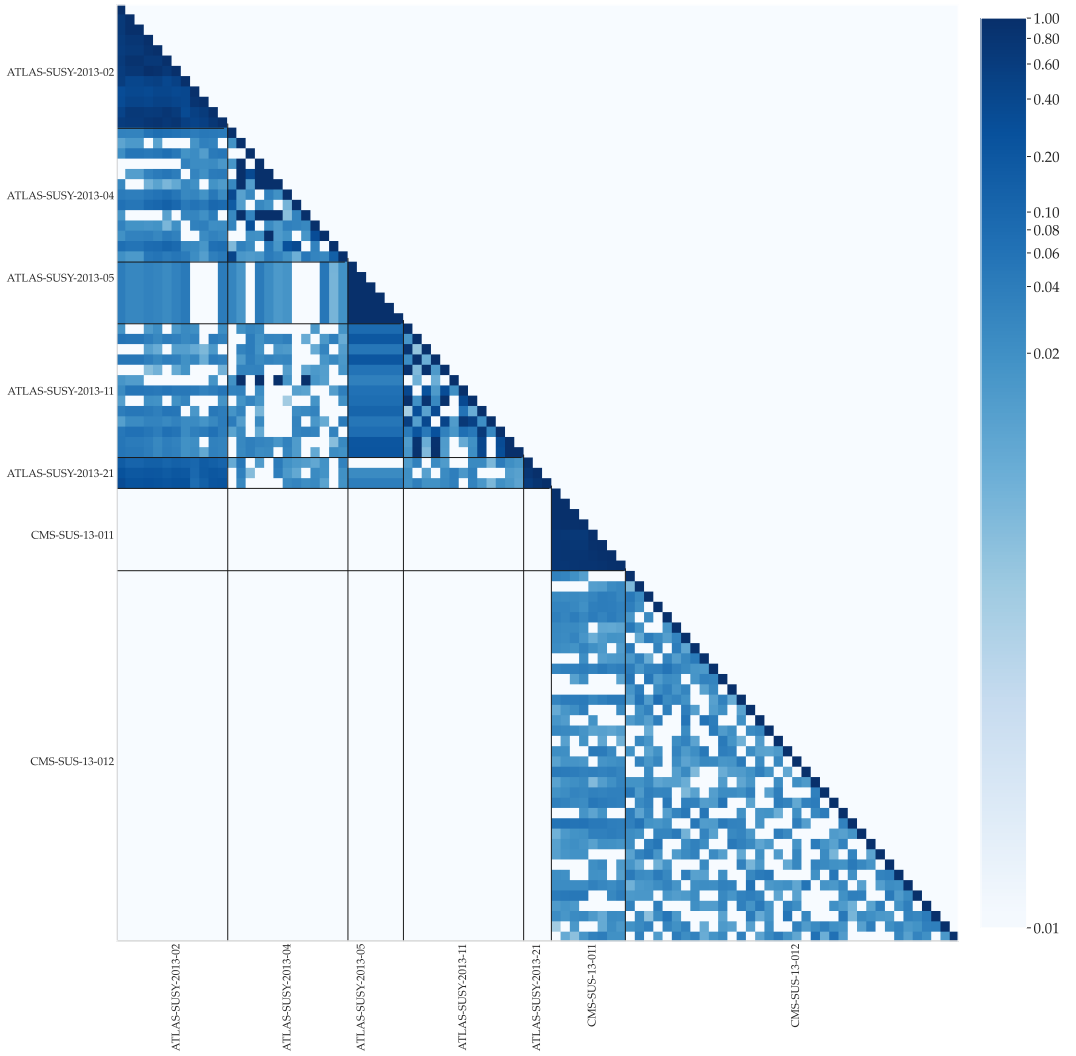


Figure B.1.: The overlap matrix ρ_{ij} obtained from the TACO sampling procedure between all LHC BSM searches at 8 TeV commonly implemented in SModelS and MadAnalysis 5. Non-overlap between ATLAS and CMS analyses is manually imposed, as analyses from both experiments could not accept the same proton collisions, regardless of the MC overlaps. Similarly, overlaps between 8 and 13 TeV analyses must be zero regardless of final-state acceptances. The set of SR pairs considered sufficiently independent in the analyses of Sections 7.3 and 7.4, with $T < 0.01$, are shown in white.

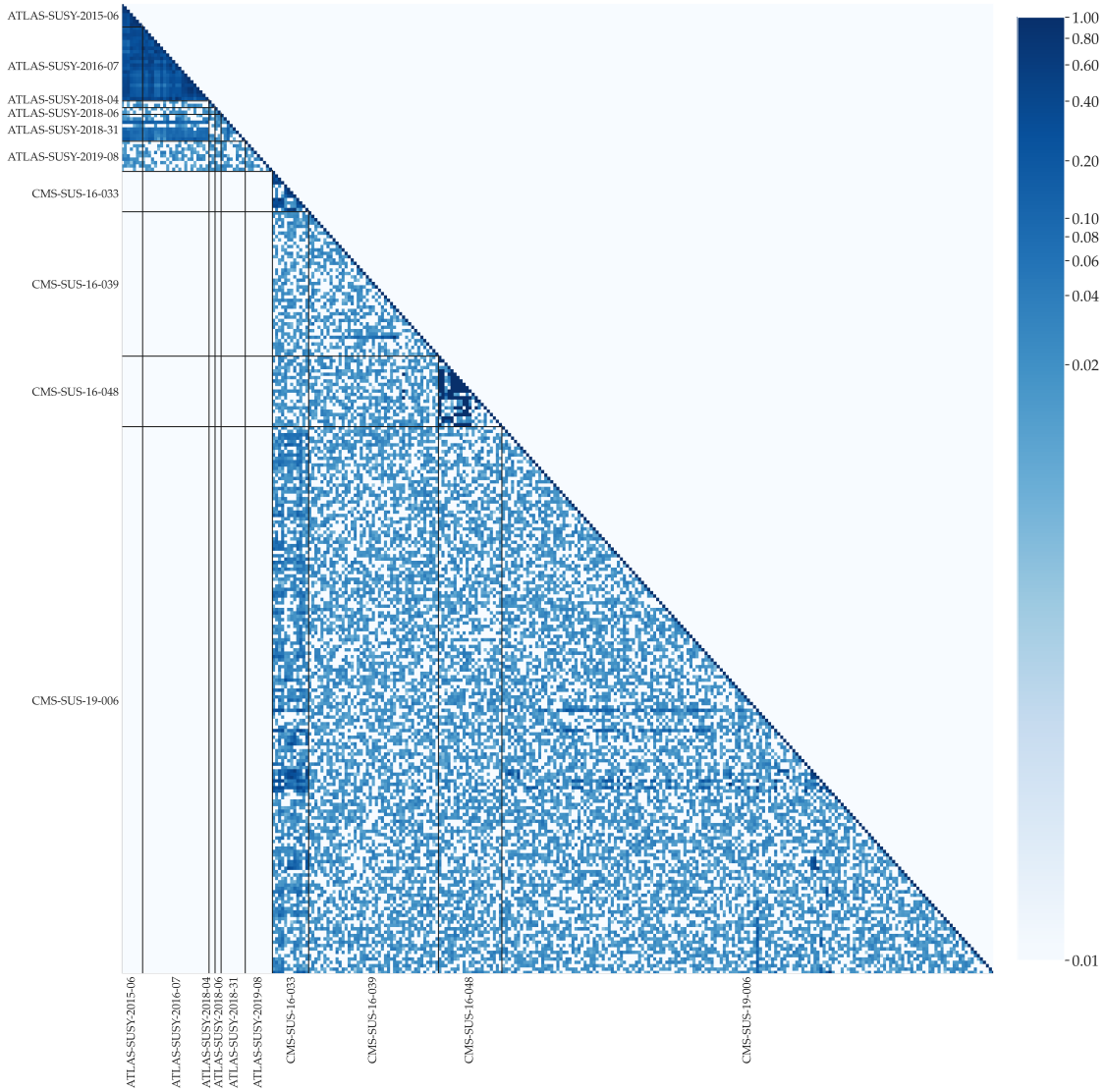


Figure B.2.: The overlap matrix ρ_{ij} obtained from the TACO sampling procedure between all LHC BSM searches at 13 TeV commonly implemented in SModelS and MadAnalysis 5. Non-overlap between ATLAS and CMS analyses is manually imposed, as analyses from both experiments could not accept the same proton collisions, regardless of the MC overlaps. Similarly, overlaps between 8 and 13 TeV analyses must be zero regardless of final-state acceptances. The set of SR pairs considered sufficiently independent in the analyses of Sections 7.3 and 7.4, with $T < 0.01$, are shown in white.

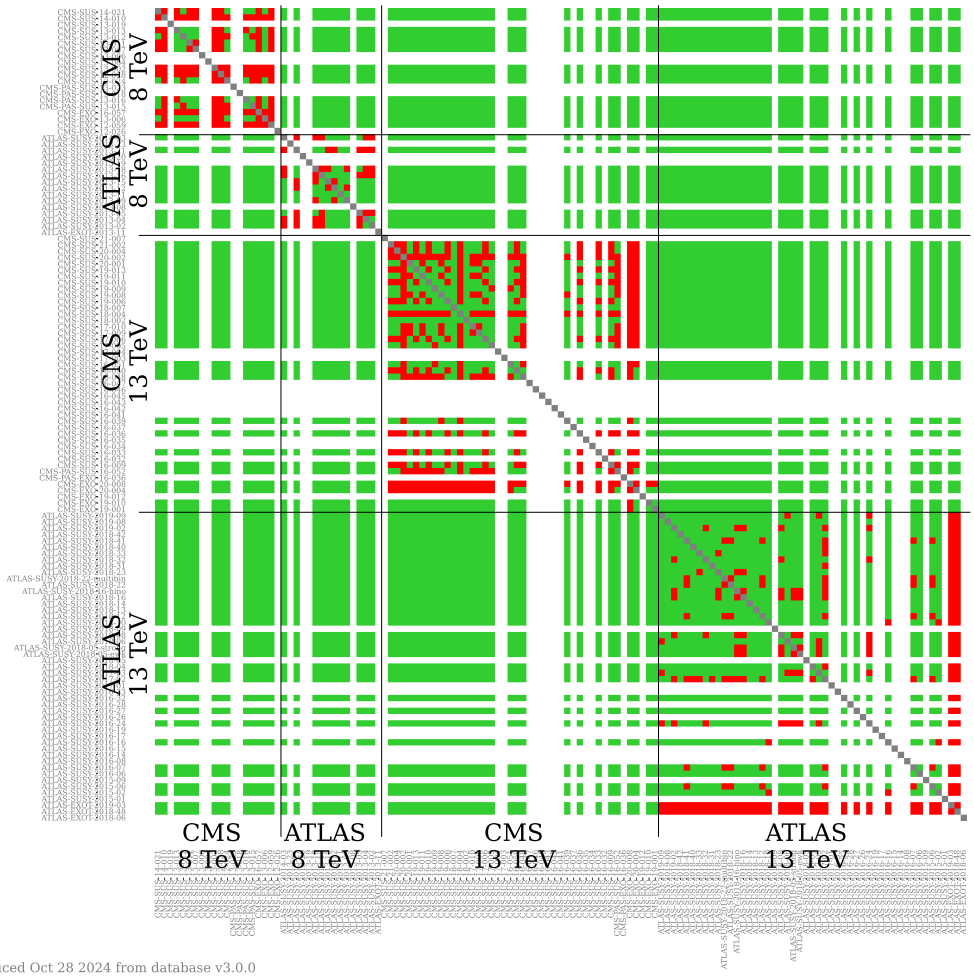


Figure B.3.: The BAM used for the proto-models v2 project. The binary combination condition was evaluated by had for analyses that share experiment and $\sqrt{s}$.

Colophon

This thesis was made in LaTeX 2_ε using the "hepthesis" class [271].

Spelling and grammar were checked using the AI tool Grammarly [272].

Literature searches were aided using the AI tool ChatGPT from OpenAI [273]

Bibliography

[1] M. Edward Peskin and D. V. Schroeder. "An Introduction To Quantum Field Theory". In: Westview Press, 1995, pp. 416–418, 425–427, 496–500, 689–691.

[2] E. Noether. "Invariante Variationsprobleme". ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: http://eudml.org/doc/59024.

[3] P. W. Higgs. "Broken Symmetries And The Masses Of Gauge Bosons". In: *Physical Review Letters* 13.16 (1964), pp. 508–509. ISSN: 0031-9007. DOI: 10.1103/physrevlett.13.508.

[4] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[5] G. McCabe. "The Structure And Interpretation Of The Standard Model". In: vol. 2.2. 1871-1774. Elsevier, 2007, pp. 133–135, 158–164, 169–172. ISBN: 978-0-444-53112-4.

[6] A. A. Kirillov. "An Introduction To Lie Groups And Lie Algebras". English. In: Cambridge, UK;New York; Cambridge University Press, 2008, pp. 7–10.

[7] A. Buckley, C. White, and M. White. "Practical Collider Physics," in: 2053-2563. IOP Publishing, 2021, pp. 24–26, 41–44, 47–50, 370–375. ISBN: 978-0-7503-2444-1. DOI: 10.1088/978-0-7503-2444-1. URL: https://dx.doi.org/10.1088/978-0-7503-2444-1.

[8] T. Nakano and K. Nishijima. "Charge Independence Forv-Particles". In: *Progress of Theoretical Physics* 10.5 (1953), pp. 581–582. ISSN: 0033-068X. DOI: 10.1143/ptp.10.581.

[9] G. Arnison et al. "Search for massive $e\nu\gamma$ and $\mu\nu\gamma$ final states at the CERN super proton synchrotron collider". In: *Physics Letters B* 135.1 (1984), pp. 250–254. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)90491-X. URL: https://www.sciencedirect.com/science/article/pii/037026938490491X.

[10] J. Goldstone. "Field theories with « Superconductor » solutions". In: *Il Nuovo Cimento* 19.1 (1961), pp. 154–164. ISSN: 0029-6341. DOI: 10.1007/bf02812722.

[11] Franco Strocchi. "The Goldstone Theorem". In: *Lecture Notes in Physics*. Lecture Notes in Physics, 2008, pp. 45–49. DOI: 10.1007/978-3-540-73593-9_9.

[12] S. Dawson. "Introduction To Electroweak Symmetry Breaking". In: *AIP conference proceedings* 1116.1 (2009). DOI: 10.1063/1.3131544.

[13] ATLAS Collaboration. "Observation Of A New Particle In The Search For The Standard Model Higgs Boson With The Atlas Detector At The Lhc". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[14] CSM Collaboration. "Observation Of A New Boson At A Mass Of 125 Gev With The CMS Experiment At The Lhc". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[15] M. D. Schwartz. "Quantum Field Theory And The Standard Model". In: Cambridge University Press, 2014, pp. 526–528, 584–588, 685–686, 60–65.

[16] Hideki Yukawa. "On the Interaction of Elementary Particles. I". In: *Progress of Theoretical Physics Supplement* 1 (Jan. 1935), pp. 1–10. ISSN: 0375-9687. DOI: 10.1143/PTPS.1.1. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf. URL: https://doi.org/10.1143/PTPS.1.1.

[17] Hideki Yukawa and Shoichi Sakata. "On the Interaction of Elementary Particles. II". In: *Progress of Theoretical Physics Supplement* 1 (Jan. 1937), pp. 14–23. ISSN: 0375-9687. DOI: 10.1143/PTPS.1.14. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.14/5311265/1-14.pdf. URL: https://doi.org/10.1143/PTPS.1.14.

[18] Andreas Bally, Yi Chung, and Florian Goertz. "Hierarchy problem and the top Yukawa coupling: An alternative to top partner solutions". In: *Physical Review D* 108.5 (2023). ISSN: 2470-0010. DOI: 10.1103/physrevd.108.055008.

[19] D. J. Gross and F. Wilczek. "Ultraviolet Behavior Of Non-Abelian Gauge Theories". In: *Physical Review Letters* 30.26 (1973), pp. 1343–1346. ISSN: 0031-9007. DOI: 10.1103/physrevlett.30.1343.

[20] H. David Politzer. "Reliable Perturbative Results For Strong Interactions?" In: *Physical Review Letters* 30.26 (1973), pp. 1346–1349. ISSN: 0031-9007. DOI: 10.1103/physrevlett.30.1346.

[21] R. K. Ellis, W. J. Stirling, and B. R. Webber. "QCD And Collider Physics". English. In: Cambridge: Cambridge University Press, 1996, pp. 22–33, 400–415. ISBN: 9780511628788.

[22] M. Shifman, A. Vainshtein, and V. Zakharov. "Qcd And Resonance Physics. Theoretical Foundations". In: *Nuclear Physics B* 147.5 (1979), pp. 385–447. ISSN: 0550-3213. DOI: 10.1016/0550-3213(79)90022-1. URL: https://www.sciencedirect.com/science/article/pii/0550321379900221.

[23] M. Gell-Mann. "A Schematic Model Of Baryons And Mesons". English. In: *Physics letters* 8.3 (1964), pp. 214–215.

[24] P. Coles and P. Lucchin. "Cosmology: The Origin And Evolution Of Cosmic Structure". In: Wiley, 2003, pp. 165–168. ISBN: 9780471489092.

[25] L. D Landau, A. A Abrikosov, and I. M Khalatnikov. "On The Elimination Of Infinity In Quantum Electrodynamics". In: *Doklady Akad. Nauk S.S.S.R.* (Mar. 1954). URL: https://www.osti.gov/biblio/4412940.

[26] U. C. Täuber. "Renormalization Group: Applications In Statistical Physics". In: *Nuclear Physics B - Proceedings Supplements* 228 (2012), pp. 7–34. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2012.06.002. URL: https://dx.doi.org/10.1016/j.nuclphysbps.2012.06.002.

[27] R. P. Feynman. "The Theory of Positrons". In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 749–759. DOI: 10.1103/PhysRev.76.749. URL: https://link.aps.org/doi/10.1103/PhysRev.76.749.

[28] Steven Weinberg. "Feynman Rules for Any Spin". In: *Physical Review* 133.5B (1964), B1318–B1332. ISSN: 0031-899X. DOI: 10.1103/physrev.133.b1318.

[29] M. Kobayashi and T. Maskawa. "Cp-Violation In The Renormalizable Theory Of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/ptp.49.652.

[30] S. Weinberg. *Gravitation And Cosmology: Principles And Applications Of The General Theory Of Relativity*. English. Wiley, 1972. ISBN: 9780471925675.

[31] S. Weinberg. "Baryon- And Lepton-Nonconserving Processes". In: *Physical Review Letters* 43.21 (1979), pp. 1566–1570. ISSN: 0031-9007. DOI: 10.1103/physrevlett.43.1566.

[32] J. F. Donoghue. "Introduction To The Effective Field Theory Description Of Gravity". In: *Advanced School on Effective Theories*. June 1995. arXiv: gr-qc/9512024.

[33] Y. Fukuda et al. "Evidence For Oscillation Of Atmospheric Neutrinos". In: *Physical Review Letters* 81.8 (1998), pp. 1562–1567. ISSN: 0031-9007. DOI: 10.1103/physrevlett.81.1562.

[34] R. N. Mohapatra and G. Senjanović. "Neutrino Mass And Spontaneous Parity Nonconservation". In: *Physical Review Letters* 44.14 (1980), pp. 912–915. ISSN: 0031-9007. DOI: 10.1103/physrevlett.44.912.

[35] P. A Zyla et al. "Review Of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020). ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104.

[36] ATLAS Collaboration. "Combined Measurement Of The Higgs Boson Mass From The $H \to \gamma\gamma$ And $H \to ZZ^* \to 4L$ Decay Channels With The Atlas Detector Using $\sqrt{7}$, 8, And 13 Tev pp Collision Data". In: *Physical Review Letters* 131.25 (2023). ISSN: 0031-9007. DOI: 10.1103/physrevlett.131.251802.

[37] G. Francesco Giudice. "Naturally Speaking: The Naturalness Criterion And Physics At The Lhc". In: WorldScientific, 2008, pp. 155–178. DOI: 10.1142/9789812779762_0010. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812779762%5C_0010.

[38] N. Aghanim et al. "Planck2018 Results". In: *Astronomy & Astrophysics* 641 (2020), A6. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201833910.

[39] S. Coleman and J. Mandula. "All Possible Symmetries Of The s-matrix". In: *Physical Review* 159.5 (1967), pp. 1251–1256. ISSN: 0031-899X. DOI: 10.1103/physrev.159.1251.

[40] P. Nath. "Grand Unification". In: *Supersymmetry, Supergravity, and Unification*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2016, pp. 14–19, 139–183, 184–235.

[41] J. Wess and B. Zumino. "Supergauge Transformations In Four Dimensions". In: *Nuclear Physics B* 70.1 (1974), pp. 39–50. ISSN: 0550-3213. DOI: 10.1016/0550-3213(74)90355-1.

[42] R. Haag, J. T. Łopuszański, and M. Sohnius. "All Possible Generators Of Supersymmetries Of The S-Matrix". In: *Nuclear Physics B* 88.2 (1975), pp. 257–274. ISSN: 0550-3213. DOI: 10.1016/0550-3213(75)90279-5. URL: https://www.sciencedirect.com/science/article/pii/0550321375902795.

[43] S. P. Martin. "A Supersymmetry Primer". In: *The High Luminosity Large Hadron Collider*. The High Luminosity Large Hadron Collider, 1998, pp. 1–5, 17–22, 53–54, 54–60. DOI: 10.1142/9789812839657_0001.

[44] D. Tong. "Supersymmetric Field Theory". In: *Online source* (2021), pp. 8–12, 99–102.

[45] S. Weinberg. "The Quantum Theory Of Fields, Volume 3: Supersymmetry". In: Cambridge University Press, 2000, pp. 82–85.

[46] J. Wess and J. Bagger. "Supersymmetry And Supergravity". English. In: 2nd rev. and expand. Princeton, N.J: Princeton University Press, 1992, pp. 174–185.

[47] H. Baer and X. Tata. *Weak Scale Supersymmetry: From Superfields To Scattering Events*. English. Cambridge: Cambridge University Press, 2006, pp. 23–28, 69–74. ISBN: 9780511617270.

[48] E. Dudas. "Holomorphy And Dynamical Scales In Supersymmetric Gauge Theories". In: *Physics Letters B* 375.1-4 (1996), pp. 189–195. ISSN: 0370-2693. DOI: 10.1016/0370-2693(96)00247-x.

[49] G. Honecker. "Kähler Metrics And Gauge Kinetic Functions For Intersecting D6-Branes On Toroidal Orbifolds – The Complete Perturbative Story". In: *Fortschritte der Physik* 60.3 (2012), pp. 243–326. ISSN: 0015-8208. DOI: 10.1002/prop.201100087. URL: https://dx.doi.org/10.1002/prop.201100087.

[50] N. Arkani-Hamed and H. Murayama. "Holomorphy, Rescaling Anomalies And Exact B Functions In Supersymmetric Gauge Theories". In: *Journal of High Energy Physics* 2000.06 (June 2000), pp. 030–030. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2000/06/030. URL: http://dx.doi.org/10.1088/1126-6708/2000/06/030.

[51] R. Delbourgo, P. D. Jarvis, and R. C. Warner. "Grassmann Coordinates And Lie Algebras For Unified Models". English. In: *Journal of mathematical physics* 34.8 (1993), pp. 3616–3641.

[52] M. E. Peskin. "Duality In Supersymmetric Yang-Mills Theory". In: *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 96): Fields, Strings, and Duality*. Feb. 1997, pp. 729–809. arXiv: hep-th/9702094.

[53] H. Nilles. "Supersymmetry, Supergravity And Particle Physics". In: *Physics Reports* 110.1 (1984), pp. 1–162. ISSN: 0370-1573. DOI: 10.1016/0370-1573(84)90008-5. URL: https://www.sciencedirect.com/science/article/pii/0370157384900085.

[54] Gian F. Giudice et al. "Gaugino mass without singlets". In: *Journal of High Energy Physics* 1998.12 (Apr. 1999), p. 027. DOI: 10.1088/1126-6708/1998/12/027. URL: https://dx.doi.org/10.1088/1126-6708/1998/12/027.

[55] A. H. Chamseddine, R. Arnowitt, and P. Nath. "Locally Supersymmetric Grand Unification". In: *Physical Review Letters* 49.14 (1982), pp. 970–974. ISSN: 0031-9007. DOI: 10.1103/physrevlett.49.970.

[56] S. K. Soni and H. Weldon. "Analysis Of The Supersymmetry Breaking Induced By N = 1 Supergravity Theories". In: *Physics Letters B* 126.3 (1983), pp. 215–219. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)90593-2. URL: https://www.sciencedirect.com/science/article/pii/0370269383905932.

[57] L. Hall, J. Lykken, and S. Weinberg. "Supergravity As The Messenger Of Supersymmetry Breaking". In: *Physical Review D* 27.10 (1983), pp. 2359–2378. ISSN: 0556-2821. DOI: 10.1103/physrevd.27.2359.

[58] E. Cremmer et al. "Super-Higgs Effect In Supergravity With General Scalar Interactions". In: *Physics Letters B* 79.3 (1978), pp. 231–234. ISSN: 0370-2693. DOI: 10.1016/0370-2693(78)90230-7. URL: https://www.sciencedirect.com/science/article/pii/0370269378902307.

[59] M. Dine and A. E. Nelson. "Dynamical Supersymmetry Breaking At Low Energies". In: *Physical Review D* 48.3 (1993), pp. 1277–1287. ISSN: 0556-2821. DOI: 10.1103/physrevd.48.1277.

[60] M. Dine et al. "New Tools For Low Energy Dynamical Supersymmetry Breaking". In: *Physical Review D* 53.5 (1996), pp. 2658–2669. ISSN: 0556-2821. DOI: 10.1103/physrevd.53.2658.

[61] N. Arkani-Hamed, S. Dimopoulos, and J. March-Russell. "Stabilization Of Sub-millimeter Dimensions: The New Guise Of The Hierarchy Problem". In: *Physical Review D* 63.6 (2001). ISSN: 0556-2821. DOI: 10.1103/physrevd.63.064020.

[62] L. Randall and R. Sundrum. "Out Of This World Supersymmetry Breaking". In: *Nuclear Physics B* 557.1 (1999), pp. 79–118. ISSN: 0550-3213. DOI: 10.1016/S0550-3213(99)00359-4. URL: https://www.sciencedirect.com/science/article/pii/S0550321399003594.

[63] R. Bauerschmidt, D. C. Brydges, and G. Slade. "Introduction To A Renormalisation Group Method". In: Springer Singapore, 2019, pp. 83–90. ISBN: 9789813295933. DOI: 10.1007/978-981-32-9593-3. URL: http://dx.doi.org/10.1007/978-981-32-9593-3.

[64] H. Baer, V. Barger, and D. Sengupta. "Anomaly-Mediated Susy Breaking Model Retrofitted For Naturalness". In: *Physical Review D* 98.1 (2018). ISSN: 2470-0010. DOI: 10.1103/physrevd.98.015039.

[65] G. F. Giudice et al. "Gaugino Mass Without Singlets". In: *Journal of High Energy Physics* 1998.12 (Apr. 1999), p. 027. DOI: 10.1088/1126-6708/1998/12/027. URL: https://dx.doi.org/10.1088/1126-6708/1998/12/027.

[66] T. Ohlsson. "Proton Decay". In: *Nuclear Physics B* 993 (2023), p. 116268. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2023.116268. URL: https://www.sciencedirect.com/science/article/pii/S0550321323001979.

[67] *Nucleon Decay: Theory And Experimental Overview*. Zenodo, Jan. 2024. DOI: 10.5281/zenodo.10493165. URL: https://doi.org/10.5281/zenodo.10493165.

[68] H. Baer and X. Tata. "The Minimal Supersymmetric Standard Model". In: *Weak Scale Supersymmetry: From Superfields to Scattering Events*. Cambridge University Press, 2006, pp. 127–140.

[69] R. Dermíšek and N. Mcginnis. "Top-Bottom-Tau Yukawa Coupling Unification In The Mssm Plus One Vectorlike Family And Fermion Masses As Ir Fixed Points". In: *Physical Review D* 99.3 (2019). ISSN: 2470-0010. DOI: 10.1103/physrevd.99.035033.

[70] S. Dimopoulos and H. Georgi. "Softly Broken Supersymmetry And Su(5)". In: *Nuclear Physics B* 193.1 (1981), pp. 150–162. ISSN: 0550-3213. DOI: 10.1016/0550-3213(81)90522-8.

[71] D. Ghosh et al. "How Constrained Is The Constrained Mssm?" In: *Physical Review D* 86.5 (2012). ISSN: 1550-7998. DOI: 10.1103/physrevd.86.055007.

[72] L. Hall, V. Kostelecky, and S. Raby. "New Flavor Violations In Supergravity Models". In: *Nuclear Physics B* 267.2 (1986), pp. 415–432. ISSN: 0550-3213. DOI: 10.1016/0550-3213(86)90397-4. URL: https://sciencedirect.com/science/article/pii/0550321386903974.

[73] A. Djouadi et al. *The Minimal Supersymmetric Standard Model: Group Summary Report*. 1999. arXiv: hep-ph/9901246 [hep-ph]. URL: https://arxiv.org/abs/hep-ph/9901246.

[74] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114. eprint: https://pubs.aip.org/aip/jcp/article-pdf/21/6/1087/18802390/1087_1_online.pdf. URL: https://doi.org/10.1063/1.1699114.

[75] John Skilling. "Nested sampling for general Bayesian computation". In: *Bayesian Analysis* 1.4 (2006), pp. 833–859. DOI: 10.1214/06-BA127. URL: https://doi.org/10.1214/06-BA127.

[76] S. Kraml et al. "Smodels: A Tool For Interpreting Simplified-Model Results From The Lhc And Its Application To Supersymmetry". In: *The European Physical Journal C* 74.5 (2014). ISSN: 1434-6044. DOI: 10.1140/epjc/s10052-014-2868-5. URL: https://dx.doi.org/10.1140/epjc/s10052-014-2868-5.

[77] C. Gütschow and Z. Marshall. "Setting Limits On Supersymmetry Using Simplified Models". In: *Journal of Visualized Experiments* 81 (2013). ISSN: 1940-087X. DOI: 10.3791/50419. URL: https://dx.doi.org/10.3791/50419.

[78] Johan Alwall, Philip C. Schuster, and Natalia Toro. "Simplified models for a first characterization of new physics at the LHC". In: *Physical Review D* 79.7 (2009). ISSN: 1550-7998. DOI: 10.1103/physrevd.79.075020.

[79] Daniele Alves et al. "Simplified models for LHC new physics searches". In: *Journal of Physics G: Nuclear and Particle Physics* 39.10 (2012), p. 105005. ISSN: 0954-3899. DOI: 10.1088/0954-3899/39/10/105005. URL: https://dx.doi.org/10.1088/0954-3899/39/10/105005.

[80] S. Chatrchyan et al. "Interpretation of searches for supersymmetry with simplified models". In: *Physical Review D* 88.5 (2013). ISSN: 1550-7998. DOI: 10.1103/physrevd.88.052017.

[81] G. Alguero et al. "Constraining New Physics With Smodels Version 2". In: *Journal of High Energy Physics* 2022.8 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep08(2022)068.

[82] M. Papucci et al. "Fastlim : A Fast Lhc Limit Calculator". In: *European Physical Journal C* 74.11 (2014), pp. 3163–3163. DOI: 10.1140/EPJC/S10052-014-3163-1.

[83] H. Okawa. *Interpretations Of Susy Searches In Atlas With Simplified Models. Supersymmetry*. Geneva, 2011. arXiv: 1110.0282. URL: http://cds.cern.ch/record/1386689.

[84] G. Hooft. "Naturalness, Chiral Symmetry, And Spontaneous Chiral Symmetry Breaking". In: *Recent Developments in Gauge Theories*. Springer US, 1980, pp. 135–157. ISBN: 978-1-4684-7571-5. DOI: 10.1007/978-1-4684-7571-5_9. URL: https://doi.org/10.1007/978-1-4684-7571-5_9.

[85] J. Ellis and K. A. Olive. *Supersymmetric Dark Matter Candidates*. 2010. arXiv: 1001.3651 [astro-ph.CO]. URL: https://arxiv.org/abs/1001.3651.

[86] K. Okumura. "Relic Abundance And Detection Prospects Of Neutralino Dark Matter In Mirage Mediation". In: *AIP Conference Proceedings*. AIP Conference Proceedings, 2006, pp. 67–73. DOI: 10.1063/1.2409069.

[87] K. Choi et al. "Neutralino Dark Matter In Mirage Mediation". In: *Journal of Cosmology and Astroparticle Physics* 2006.12 (2006). ISSN: 1475-7516. DOI: 10.1088/1475-7516/2006/12/017. URL: http://dx.doi.org/10.1088/1475-7516/2006/12/017.

[88] A. Arbey and F. Mahmoudi. "Dark Matter And The Early Universe: A Review". In: *Progress in Particle and Nuclear Physics* 119 (2021), p. 103865. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2021.103865. URL: https://www.sciencedirect.com/science/article/pii/S0146641021000193.

[89] Y. Gu et al. "Light Gravitino Dark Matter: Lhc Searches And The Hubble Tension". In: *Physical Review D* 102.11 (2020). ISSN: 2470-0010. DOI: 10.1103/physrevd.102.115005.

[90] G. Jungman, M. Kamionkowski, and K. Griest. "Supersymmetric Dark Matter". In: *Physics Reports* 267.5 (1996), pp. 195–373. ISSN: 0370-1573. DOI: 10.1016/0370-1573(95)00058-5. URL: https://www.sciencedirect.com/science/article/pii/0370157395000585.

[91] J. Schwichtenberg. "Gauge Coupling Unification Without Supersymmetry". In: *The European Physical Journal C* 79.4 (2019). ISSN: 1434-6044. DOI: 10.1140/epjc/s10052-019-6878-1.

[92] ATLAS Collaboration. "Summary Of The Atlas Experiment'S Sensitivity To Supersymmetry After Lhc Run 1 — Interpreted In The Phenomenological Mssm". In: *Journal of High Energy Physics* 2015.10 (2015). ISSN: 1029-8479. DOI: 10.1007/jhep10(2015)134.

[93] ATLAS Collaboration. "Search For Squarks And Gluinos With The Atlas Detector In Final States With Jets And Missing Transverse Momentum Using $\sqrt{s} = 8$ Tev Proton-Proton Collision Data". In: *Journal of High Energy Physics* 2014.9 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)176.

[94] ATLAS Collaboration. "Search For New Phenomena In Final States With Large Jet Multiplicities And Missing Transverse Momentum At $\sqrt{s} = 8$ Tev Proton-Proton Collisions Using The Atlas Experiment". In: *Journal of High Energy Physics* 2013.10 (2013). ISSN: 1029-8479. DOI: 10.1007/jhep10(2013)130.

[95] ATLAS Collaboration. "Search For Squarks And Gluinos In Events With Isolated Leptons, Jets And Missing Transverse Momentum At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2015.4 (2015). ISSN: 1029-8479. DOI: 10.1007/jhep04(2015)116.

[96] ATLAS Collaboration. "Search For Supersymmetry In Events With Large Missing Transverse Momentum, Jets, And At Least One Tau Lepton In 20fb^{-}1 Of $\sqrt{s} = 8$ Tev Proton-Proton Collision Data With The Atlas Detector". In: *Journal of High Energy Physics* 2014.9 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)103.

[97] ATLAS Collaboration. "Search For Supersymmetry At $\sqrt{s} = 8$ Tev In Final States With Jets And Two Same-Sign Leptons Or Three Leptons With The Atlas Detector". In: *Journal of High Energy Physics* 2014.6 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep06(2014)035.

[98] ATLAS Collaboration. "Search For Strong Production Of Supersymmetric Particles In Final States With Missing Transverse Momentum And At Least Three B-Jets At $\sqrt{s} = 8$ Tev Proton-Proton Collisions With The Atlas Detector". In: *Journal of High Energy Physics* 2014.10 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep10(2014)024.

[99] ATLAS Collaboration. "Search For New Phenomena In Final States With An Energetic Jet And Large Missing Transverse Momentum In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *The European Physical Journal C* 75.7 (July 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3517-3. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3517-3.

[100] ATLAS Collaboration. "Search For Direct Pair Production Of The Top Squark In All-Hadronic Final States In Proton-Proton Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2014.9 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)015.

[101] ATLAS Collaboration. "Search For Top Squark Pair Production In Final States With One Isolated Lepton, Jets, And Missing Transverse Momentum In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *Journal of High Energy Physics* 2014.11 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)118.

[102] ATLAS Collaboration. "Search For Direct Top-Squark Pair Production In Final States With Two Leptons In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2014.6 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep06(2014)124.

[103] ATLAS Collaboration. "Search For Pair-Produced Third-Generation Squarks Decaying Via Charm Quarks Or In Compressed Supersymmetric Scenarios In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Physical Review D* 90.5 (2014). ISSN: 1550-7998. DOI: 10.1103/physrevd.90.052008.

[104] ATLAS Collaboration. "Search For Direct Top Squark Pair Production In Events With A Z Boson, B-Jets And Missing Transverse Momentum In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *The European Physical Journal C* 74.6 (June 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-2883-6. URL: http://dx.doi.org/10.1140/epjc/s10052-014-2883-6.

[105] ATLAS Collaboration. "Search For Direct Third-Generation Squark Pair Production In Final States With Missing Transverse Momentum And Two B-Jets In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *Journal of High Energy Physics* 2013.10 (2013). ISSN: 1029-8479. DOI: 10.1007/jhep10(2013)189.

[106] ATLAS Collaboration. "Atlas Run 1 Searches For Direct Pair Production Of Third-Generation Squarks At The Large Hadron Collider". In: *The European Physical Journal C* 75.10 (2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3726-9. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3726-9.

[107] ATLAS Collaboration. "Search For Direct Pair Production Of A Chargino And A Neutralino Decaying To The 125 Gev Higgs Boson In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *The European Physical Journal C* 75.5 (2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3408-7. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3408-7.

[108] ATLAS Collaboration. "Search For Direct Production Of Charginos, Neutralinos And Sleptons In Final States With Two Leptons And Missing Transverse Momentum In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High*

Energy Physics 2014.5 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep05(2014)071. URL: http://dx.doi.org/10.1007/JHEP05(2014)071.

[109] ATLAS Collaboration. "Search For Direct Production Of Charginos And Neutralinos In Events With Three Leptons And Missing Transverse Momentum In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *Journal of High Energy Physics* 2014.4 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep04(2014)169.

[110] ATLAS Collaboration. "Search For The Direct Production Of Charginos, Neutralinos And Staus In Final States With At Least Two Hadronically Decaying Taus And Missing Transverse Momentum In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2014.10 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep10(2014)096.

[111] ATLAS Collaboration. "Search For Supersymmetry In Events With Four Or More Leptons In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *Physical Review D* 90.5 (2014). ISSN: 1550-7998. DOI: 10.1103/physrevd.90.052001.

[112] ATLAS Collaboration. "Search For Charginos Nearly Mass Degenerate With The Lightest Neutralino Based On A Disappearing-Track Signature In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Physical Review D* 88.11 (2013). ISSN: 1550-7998. DOI: 10.1103/physrevd.88.112006.

[113] ATLAS Collaboration. "Search For Neutral Higgs Bosons Of The Minimal Supersymmetric Standard Model In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2014.11 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)056.

[114] ATLAS Collaboration. "Searches For Heavy Long-Lived Charged Particles With The Atlas Detector In Proton-Proton Collisions At $\sqrt{s} = 8$ Tev". In: *Journal of High Energy Physics* 2015.1 (2015). ISSN: 1029-8479. DOI: 10.1007/jhep01(2015)068.

[115] ATLAS Collaboration. "Search For Neutral Higgs Bosons Of The Minimal Supersymmetric Standard Model In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Journal of High Energy Physics* 2014.11 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)056.

[116] W. Adam and I. Vivarelli. "Status Of Searches For Electroweak-Scale Supersymmetry After Lhc Run 2". In: *International Journal of Modern Physics A* 37.02 (2022). ISSN: 0217-751X. DOI: 10.1142/s0217751x21300222.

[117] ATLAS Collaboration. *SUSY August 2023 Summary Plot Update*. Geneva, 2023. URL: http://cds.cern.ch/record/2871728.

[118] ATLAS Collaboration. "Search for electroweak production of supersymmetric states in scenarios with compressed mass spectra at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physical Review D* 97.5 (2018). ISSN: 2470-0010. DOI: 10.1103/physrevd.97.052010.

[119] A. Nikolaevich Kolmogorov. *Foundations Of The Theory Of Probability*. Chelsea Pub. Co, 1933.

[120] J. M. Bernardo and A. F. M. Smith. "Bayesian Theory". English. In: New York;Chichester, Eng; Wiley, 2000, pp. 40–45, 60–75, 106–110. ISBN: 9780470316870.

[121] F. J. Samaniego. "A Comparison Of The Bayesian And Frequentist Approaches To Estimation". English. In: New York: Springer, 2010, pp. 46–50, 59–60, 101–103. ISBN: 9781441959416;

[122] Sir Jeffreys Harold. "Theory of probability". English. In: Oxford: Clarendon Press, 1939, pp. 27–30.

[123] Robert E. Kass. "Bayes Factors in Practice". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 42.5 (1993), p. 551. ISSN: 0039-0526. DOI: 10.2307/2348679.

[124] K. I. Park and S. (Online service). *Fundamentals Of Probability And Stochastic Processes With Applications To Communications*. English. Cham: Springer International Publishing, 2018, pp. 110–115. ISBN: 9783319680750.

[125] A. Papoulis and S. U. Pillai. "Probability, Random Variables, And Stochastic Processes". English. In: 4th. London;New York; McGraw-Hill, 2002, pp. 109–120. ISBN: 9780073660110.

[126] S. Kullback and R. A. Leibler. "On Information And Sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729694.

[127] I. Csiszar. "Divergence Geometry Of Probability Distributions And Minimization Problems". In: *The Annals of Probability* 3.1 (1975), pp. 146–158. ISSN: 0091-1798. DOI: 10.1214/aop/1176996454.

[128] M. S. Alencar and R. T. d. Alencar. "Set, Measure And Probability Theory". English. In: Milton: River Publishers, 2023, pp. 222–230. ISBN: 9781003808954.

[129] K. P. Choi. "On The Medians Of Gamma Distributions And An Equation Of Ramanujan". In: *Proceedings of the American Mathematical Society* 121.1 (1994), pp. 245–251. ISSN: 0002-9939. DOI: 10.1090/s0002-9939-1994-1195477-8.

[130] Andrew Gelman. "Bayesian data analysis". English. In: Third. Boca Raton, Florida: CRC Press, 2013, pp. 30–40. ISBN: 9781439898208.

[131] Kevin P. Murphy and Massachusetts Institute of Technology. "Probabilistic machine learning: an introduction". English. In: Cambridge, Massachusetts: The MIT Press, 2022, pp. 97, 105–106, 156. ISBN: 0262369303.

[132] M. Ahsanullah and S. (Online service). "Characterizations Of Univariate Continuous Distributions". English. In: Paris: Atlantis Press, 2017, pp. 17–22. ISBN: 9789462391390;

[133] C. Lee, F. Famoye, and A. Y. Alzaatreh. "Methods For Generating Families Of Univariate Continuous Distributions In The Recent Decades". In: *WIREs Computational Statistics* 5.3 (2013), pp. 219–238. ISSN: 1939-5108. DOI: 10.1002/wics.1255.

[134] M. Sugiyama. "Introduction To Statistical Machine Learning". In: Boston: Morgan Kaufmann, 2015, pp. 485–490. ISBN: 978-0-12-802121-7. DOI: 10.1016/B978-0-12-802121-7.00051-0. URL: https://www.sciencedirect.com/science/article/pii/B9780128021217000510.

[135] D. J. C. MacKay. *Information Theory, Inference, And Learning Algorithms*. English. Cambridge: Cambridge University Press, 2003, pp. 319–330. ISBN: 9780521642989.

[136] MIT Open Learning. *MIT Open Learning: Statistics For Applications*. https://openlearninglibrary.mit.edu/asset-v1:MITx+18.05r_10+2022_Summer+type@asset+block/pdf_mit18_05_s22_class15-prep-a.pdf. Licence: https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode, Accessed: 2024-09-09. 2022.

[137] L. Wasserman. "All Of Statistics: A Concise Course In Statistical Inference". English. In: New York: Springer, 2004, pp. 87–88, 61–65, 124–130. ISBN: 9780387402727.

[138] R. A. Fisher. "On The Mathematical Foundations Of Theoretical Statistics". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222.594-604 (1922), pp. 309–368. ISSN: 0264-3952. DOI: 10.1098/rsta.1922.0009. URL: https://dx.doi.org/10.1098/rsta.1922.0009.

[139] R. A. Fisher. "Two New Properties Of Mathematical Likelihood". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 144.852 (1934), pp. 285–307. ISSN: 0950-1207. DOI: 10.1098/rspa.1934.0050.

[140] R. A. Fisher. "Theory Of Statistical Estimation". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 22.5 (1925), pp. 700–725. ISSN: 0305-0041. DOI: 10.1017/s0305004100009580.

[141] G. Barnard. "Statistical Inference". In: *Journal of the Royal Statistical Society series B-Statistical Methodology* 11.2 (1949), pp. 115–149. ISSN: 1369-7412.

[142] J. Berger and R. Wolpert. "The Likelihood Principle". In: Institute of Mathematical Statistics. Lecture. Institute of Mathematical Statistics, 1988, pp. 66, 182. ISBN: 9780940600133. URL: https://books.google.co.uk/books?id=7fz8JGLmWbgC.

[143] Kevin P. Murphy. *Machine learning: a probabilistic perspective*. English. Cambridge, MA: MIT Press, 2012, pp. 220–226.

[144] A. W. F. Edwards. "Likelihood". English. In: Expand. London;Baltimore; Johns Hopkins University Press, 1992, pp. 1–36. ISBN: 9780801844454.

[145] D. R. Cox. "Principles Of Statistical Inference". English. In: Cambridge: Cambridge University Press, 2006, pp. 55–70, 108–109. ISBN: 9780511813559.

[146] L. Le Cam. "Maximum Likelihood: An Introduction". In: *International Statistical Review / Revue Internationale de Statistique* 58.2 (1990), pp. 153–171. ISSN: 03067734, 17515823. URL: http://www.jstor.org/stable/1403464 (visited on Aug. 20, 2024).

[147] Y. Pawitan. "In All Likelihood: Statistical Modelling And Inference Using Likelihood". English. In: New York;Oxford; Clarendon Press, 2001, pp. 12–29, 77–85. ISBN: 9780191650574.

[148] J. Neyman and E. S. Pearson. "Ix. On The Problem Of The Most Efficient Tests Of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706 (1933), pp. 289–337. ISSN: 0264-3952. DOI: 10.1098/rsta.1933.0009.

[149] J. Fan, C. Zhang, and J. Zhang. "Generalized Likelihood Ratio Statistics And Wilks Phenomenon". In: *The Annals of Statistics* 29.1 (2001), pp. 153–193. ISSN: 0090-5364. DOI: 10.1214/aos/996986505.

[150] S. S. Wilks. "The Large-Sample Distribution Of The Likelihood Ratio For Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. ISSN: 0003-4851. DOI: 10.1214/aoms/1177732360.

[151] E. L. Lehmann, J. P. Romano, and S. (Online service). "Testing Statistical Hypotheses". English. In: 4th 2022. Cham: Springer International Publishing, 2022, pp. 65–70. ISBN: 9783030705787.

[152] T. Tokhadze. "Likelihoodism And Guidance For Belief". In: *Journal for General Philosophy of Science* 53.4 (2022), pp. 501–517. ISSN: 0925-4560. DOI: 10.1007/s10838-022-09608-3.

[153] R. Richard M. "Statistical Evidence: A Likelihood Paradigm". English. In: Boca Raton, Florida: Chapman & Hall/CRC, 2000, pp. 35–40, 41–50, 63–65. ISBN: 9780203738665.

[154] G. Casella and R. L. Berger. "Statistical Inference". English. In: 2nd. London;Pacific Grove, CA; Duxbury, 2002, pp. 373–380. ISBN: 9780534243128;

[155] M. P. Fay and E. H. Brittain. "Statistical Hypothesis Testing In Context: Reproducibility, Inference, And Science". In: Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2022.

[156] L. Lyons. *Discovering The Significance Of 5 Sigma*. 2013. arXiv: 1310.1284 [physics.data-an]. URL: https://arxiv.org/abs/1310.1284.

[157] P. Embrechts and M. Hofert. "A Note On Generalized Inverses". In: *Mathematical Methods of Operations Research* 77.3 (2013), pp. 423–432. ISSN: 1432-2994. DOI: 10.1007/s00186-013-0436-7. URL: https://dx.doi.org/10.1007/s00186-013-0436-7.

[158] D. J. Murdoch, Y. Tsai, and J. Adcock. "P-Values Are Random Variables". In: *The American Statistician* 62.3 (2008), pp. 242–245. ISSN: 00031305. URL: http://www.jstor.org/stable/27644033 (visited on Aug. 28, 2024).

[159] J. Neyman. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability". In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236.767 (1937), pp. 333–380. ISSN: 0080-4614. DOI: 10.1098/rsta.1937.0005.

[160] E. L. Lehmann. "Theory Of Point Estimation". English. In: New York, NY: Wiley, 1983, pp. 300–312, 185–187. ISBN: 9780471058496;

[161] K CRANMER. *STATISTICAL CHALLENGES FOR SEARCHES FOR NEW PHYSICS AT THE LHC*. Brookhaven National Lab. (BNL), Upton, NY (United States), Sept. 2005. URL: https://www.osti.gov/biblio/861328.

[162] Kyle Cranmer. *Practical Statistics for the LHC*. 2015. arXiv: 1503.07622. URL: https://arxiv.org/abs/1503.07622.

[163] T. B. Berrett and R. J. Samworth. "Usp: An Independence Test That Improves On Pearson'S Chi-Squared And The G -Test". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477.2256 (2021). ISSN: 1364-5021. DOI: 10.1098/rspa.2021.0549.

[164] R. A. Fisher. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P". In: *Journal of the Royal Statistical Society* 85.1 (1922), p. 87. ISSN: 0952-8385. DOI: 10.2307/2340521.

[165] F. Gao and L. Han. "Implementing The Nelder-Mead Simplex Algorithm With Adaptive Parameters". In: *Computational Optimization and Applications* 51.1 (2012), pp. 259–277. ISSN: 0926-6003. DOI: 10.1007/s10589-010-9329-3.

[166] X. Meng and D. Van Dyk. "The Em Algorithm—An Old Folk-Song Sung To A Fast New Tune". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 59.3 (Jan. 2002), pp. 511–567. ISSN: 0035-9246. DOI: 10.1111/1467-9868.00082. eprint: https://academic.oup.com/jrsssb/article-pdf/59/3/511/49588939/jrsssb_59_3_511.pdf. URL: https://doi.org/10.1111/1467-9868.00082.

[167] R. Fletcher. "Practical Methods Of Optimization". English. In: 2nd. New York: Wiley, 1987, pp. 50–60. ISBN: 9781118723203.

[168] A. Gelman. "Prior Distributions For Variance Parameters In Hierarchical Models (Comment On Article By Browne And Draper)". In: *Bayesian Analysis* 1 (2004), pp. 515–534. URL: https://api.semanticscholar.org/CorpusID:15141558.

[169] D. J. Venzon and S. H. Moolgavkar. "A Method For Computing Profile-Likelihood-Based Confidence Intervals". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 37.1 (1988), p. 87. ISSN: 0035-9254. DOI: 10.2307/2347496.

[170] J. Alwall et al. "The Automated Computation Of Tree-Level And Next-To-Leading Order Differential Cross Sections, And Their Matching To Parton Shower Simulations". In: *Journal of High Energy Physics* 2014.7 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: https://dx.doi.org/10.1007/jhep07(2014)079.

[171] T. Gleisberg et al. "Sherpa 1, A Proof-Of-Concept Version". In: *Journal of High Energy Physics* 2004.02 (Feb. 2004), pp. 056–056. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2004/02/056. URL: http://dx.doi.org/10.1088/1126-6708/2004/02/056.

[172] C. Bierlich et al. "A Comprehensive Guide To The Physics And Usage Of Pythia 8.3". In: *scipost* (2022). DOI: 10.21468/scipostphyscodeb.8.

[173] E. Conte, B. Fuks, and G. Serret. "Madanalysis 5, A User-Friendly Framework For Collider Phenomenology". In: *Comput. Phys. Commun.* 184 (2013), pp. 222–256. DOI: 10.1016/j.cpc.2012.09.009. arXiv: 1206.1599 [hep-ph].

[174] A. Buckley et al. "Rivet User Manual". In: *Comput. Phys. Commun.* 184 (2013), pp. 2803–2819. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

[175] J. M. Butterworth et al. "Constraining New Physics With Collider Measurements Of Standard Model Signatures". In: *Journal of High Energy Physics* 2017.3 (2017). ISSN: 1029-8479. DOI: 10.1007/jhep03(2017)078.

[176] Christian Bierlich et al. "Robust independent validation of experiment and theory: Rivet version 4 release note". In: *SciPost Physics Codebases* (2024). ISSN: 2949-804X. DOI: 10.21468/scipostphyscodeb.36.

[177] Andy Buckley et al. "The HepMC3 event record library for Monte Carlo event generators". In: *Computer Physics Communications* 260 (2021), p. 107310. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107310.

[178] Eamonn Maguire, Lukas Heinrich, and Graeme Watt. "HEPData: a repository for high energy physics data". In: *J. Phys. Conf. Ser.* 898.10 (2017). Ed. by Richard Mount and Craig Tull, p. 102006. DOI: 10.1088/1742-6596/898/10/102006. arXiv: 1704.05473 [hep-ex].

[179] Andy Buckley et al. "Consistent, multidimensional differential histogramming and summary statistics with YODA 2". In: *SciPost Phys.* (Dec. 2023). DOI: 10.21468/SciPostPhysCodeb.45. arXiv: 2312.15070 [hep-ph].

[180] Jack Y. Araz et al. *Les Houches guide to reusable ML models in LHC analyses*. 2024. arXiv: 2312.14575 [hep-ph]. URL: https://arxiv.org/abs/2312.14575.

[181] J. D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[182] S. Mrenna and P. Richardson. "Matching Matrix Elements And Parton Showers With Herwig And Pythia". In: *Journal of High Energy Physics* 2004.05 (June 2004), p. 040. DOI: 10.1088/1126-6708/2004/05/040. URL: https://dx.doi.org/10.1088/1126-6708/2004/05/040.

[183] S. Ovyn, X. Rouby, and V. Lemaitre. *Delphes, A Framework For Fast Simulation Of A Generic Collider Experiment*. 2010. arXiv: 0903.2225 [hep-ph]. URL: https://arxiv.org/abs/0903.2225.

[184] M. Mahdi Altakach et al. *SModelS V3: Going Beyond Z_2 Topologies*. 2024. arXiv: 2409.12942 [hep-ph]. URL: https://arxiv.org/abs/2409.12942.

[185] Nicolas Berger et al. *Simplified Template Cross Sections - Stage 1.1*. 2019. arXiv: 1906.02754 [hep-ph]. URL: https://arxiv.org/abs/1906.02754.

[186] Lorenzo Moneta et al. *The RooStats Project*. 2011. arXiv: 1009.1003 [physics.data-an]. URL: https://arxiv.org/abs/1009.1003.

[187] M. Feickert, L. Heinrich, and G. Stark. *PYHF: Pure-Python Implementation Of HistFactory With Tensors And Automatic Differentiation*. 2022. arXiv: 2211.15838 [hep-ex]. URL: https://arxiv.org/abs/2211.15838.

[188] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. New York, 2012. DOI: 10.17181/CERN-OPEN-2012-016. URL: http://cds.cern.ch/record/1456844.

[189] A. Hofler et al. "Innovative Applications Of Genetic Algorithms To Problems In Accelerator Physics". In: *Physical Review Special Topics - Accelerators and Beams* 16.1 (2013). ISSN: 1098-4402. DOI: 10.1103/physrevstab.16.010101.

[190] P. Mitra, C. Murthy, and S. Pal. "Unsupervised Feature Selection Using Feature Similarity". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.3 (2002), pp. 301–312. ISSN: 0162-8828. DOI: 10.1109/34.990133.

[191] S. Lin et al. "Parameter Determination Of Support Vector Machine And Feature Selection Using Simulated Annealing Approach". In: *Applied Soft Computing* 8.4 (2008). Soft Computing for Dynamic Data Mining, pp. 1505–1512. ISSN: 1568-4946. DOI: 10.1016/j.asoc.2007.10.012. URL: https://www.sciencedirect.com/science/article/pii/S156849460700141X.

[192] R. Sedgewick. *Algorithms In C: Graph Algorithms*. 3rd. Addison-Wesley, 2002. ISBN: 9780201316636.

[193] A. A. Hagberg, D. A. Schult, and P. J. Swart. "Exploring Network Structure, Dynamics, And Function Using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.

[194] J. Yellen et al. "Strength In Numbers: Optimal And Scalable Combination Of Lhc New-Physics Searches". In: *SciPost Phys.* 14 (2023), p. 077. DOI: 10.21468/SciPostPhys.14.4.077. URL: https://scipost.org/10.21468/SciPostPhys.14.4.077.

[195] ATLAS Collaboration. "Search For Squarks And Gluinos With The Atlas Detector In Final States With Jets And Missing Transverse Momentum Using $\sqrt{s} = 8$ Tev Proton–Proton Collision Data". In: *JHEP* 09 (2014), p. 176. DOI: 10.1007/JHEP09(2014)176. arXiv: 1405.7875 [hep-ex].

[196] ATLAS Collaboration. "Search For New Phenomena In Final States With Large Jet Multiplicities And Missing Transverse Momentum At $\sqrt{s} = 8$ Tev Proton-Proton Collisions Using The Atlas Experiment". In: *JHEP* 10 (2013). [Erratum: JHEP 01, 109 (2014)], p. 130. DOI: 10.1007/JHEP10(2013)130. arXiv: 1308.1841 [hep-ex].

[197] ATLAS Collaboration. "Search For Direct Third-Generation Squark Pair Production In Final States With Missing Transverse Momentum And Two B-Jets In $\sqrt{s} = 8$ Tev pp Collisions With The Atlas Detector". In: *JHEP* 10 (2013), p. 189. DOI: 10.1007/JHEP10(2013)189. arXiv: 1308.2631 [hep-ex].

[198] ATLAS Collaboration. "Search For Direct Production Of Charginos, Neutralinos And Sleptons In Final States With Two Leptons And Missing Transverse Momentum In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *JHEP* 05 (2014), p. 071. DOI: 10.1007/JHEP05(2014)071. arXiv: 1403.5294 [hep-ex].

[199] ATLAS Collaboration. "Search For Pair-Produced Third-Generation Squarks Decaying Via Charm Quarks Or In Compressed Supersymmetric Scenarios In pp Collisions At $\sqrt{s} = 8$ Tev With The Atlas Detector". In: *Phys. Rev. D* 90.5 (2014), p. 052008. DOI: 10.1103/PhysRevD.90.052008. arXiv: 1407.0608 [hep-ex].

[200] ATLAS Collaboration. "Search For Squarks And Gluinos In Final States With Jets And Missing Transverse Momentum At $\sqrt{s} = 13$ Tev With The Atlas Detector". In: *Eur. Phys. J. C* 76.7 (2016), p. 392. DOI: 10.1140/epjc/s10052-016-4184-8. arXiv: 1605.03814 [hep-ex].

[201] ATLAS Collaboration. "Search For Squarks And Gluinos In Final States With Jets And Missing Transverse Momentum Using 36fb^{-1} Of $\sqrt{s} = 13$ Tev pp Collision Data With The Atlas Detector". In: *Phys. Rev. D* 97.11 (2018), p. 112001. DOI: 10.1103/PhysRevD.97.112001. arXiv: 1712.02332 [hep-ex].

[202] ATLAS Collaboration. "Search For Direct Stau Production In Events With Two Hadronic τ-Leptons In $\sqrt{s} = 13$ Tev pp Collisions With The Atlas Detector". In: *Phys. Rev. D* 101.3 (2020), p. 032009. DOI: 10.1103/PhysRevD.101.032009. arXiv: 1911.06660 [hep-ex].

[203] ATLAS Collaboration. "Search For Chargino-Neutralino Production With Mass Splittings Near The Electroweak Scale In Three-Lepton Final States In $\sqrt{s} = 13$ Tev pp Collisions With The Atlas Detector". In: *Phys. Rev. D* 101.7 (2020), p. 072001. DOI: 10.1103/PhysRevD.101.072001. arXiv: 1912.08479 [hep-ex].

[204] ATLAS Collaboration. "Search For Bottom-Squark Pair Production With The Atlas Detector In Final States Containing Higgs Bosons, B-Jets And Missing Transverse Momentum". In: *JHEP* 12 (2019), p. 060. DOI: 10.1007/JHEP12(2019)060. arXiv: 1908.03122 [hep-ex].

[205] ATLAS Collaboration. "Search For Electroweak Production Of Charginos And Sleptons Decaying Into Final States With Two Leptons And Missing Transverse Momentum In $\sqrt{s} = 13$ Tev pp Collisions Using The Atlas Detector". In: *Eur. Phys. J. C* 80.2 (2020), p. 123. DOI: 10.1140/epjc/s10052-019-7594-6. arXiv: 1908.08215 [hep-ex].

[206] ATLAS Collaboration. "Search For Direct Production Of Electroweakinos In Final States With One Lepton, Missing Transverse Momentum And A Higgs Boson Decaying Into Two B-Jets In pp Collisions At $\sqrt{s} = 13$ Tev With The Atlas Detector". In: *Eur. Phys. J. C* 80.8 (2020), p. 691. DOI: 10.1140/epjc/s10052-020-8050-3. arXiv: 1909.09226 [hep-ex].

[207] CSM Collaboration. "Search For Supersymmetry In Final States With Photons And Missing Transverse Momentum In Proton-Proton Collisions At 13 Tev". In: *JHEP* 06 (2019), p. 143. DOI: 10.1007/JHEP06(2019)143. arXiv: 1903.07070 [hep-ex].

[208] CSM Collaboration. "Search For New Physics In The Multijet And Missing Transverse Momentum Final State In Proton-Proton Collisions At $\sqrt{s} = 8$ Tev". In: *JHEP* 06 (2014), p. 055. DOI: 10.1007/JHEP06(2014)055. arXiv: 1402.4770 [hep-ex].

[209] CSM Collaboration. "Search For Supersymmetry In Multijet Events With Missing Transverse Momentum In Proton-Proton Collisions At 13 Tev". In: *Phys. Rev. D* 96.3 (2017), p. 032003. DOI: 10.1103/PhysRevD.96.032003. arXiv: 1704.07781 [hep-ex].

[210] CSM Collaboration. "Search For Electroweak Production Of Charginos And Neutralinos In Multilepton Final States In Proton-Proton Collisions At $\sqrt{s} = 13$ Tev". In: *JHEP* 03 (2018), p. 166. DOI: 10.1007/JHEP03(2018)166. arXiv: 1709.05406 [hep-ex].

[211] CSM Collaboration. "Search For New Physics In Events With Two Soft Oppositely Charged Leptons And Missing Transverse Momentum In Proton-Proton Collisions At $\sqrt{s} = 13$ Tev". In: *Phys. Lett. B* 782 (2018), pp. 440–467. DOI: 10.1016/j.physletb.2018.05.062. arXiv: 1801.01846 [hep-ex].

[212] CSM Collaboration. "Search For Top Squarks And Dark Matter Particles In Opposite-Charge Dilepton Final States At $\sqrt{s} = 13$ Tev". In: *Phys. Rev. D* 97.3 (2018), p. 032009. DOI: 10.1103/PhysRevD.97.032009. arXiv: 1711.00752 [hep-ex].

[213] CSM Collaboration. "Search For Supersymmetry In Proton-Proton Collisions At 13 Tev In Final States With Jets And Missing Transverse Momentum". In: *JHEP* 10 (2019), p. 244. DOI: 10.1007/JHEP10(2019)244. arXiv: 1908.04722 [hep-ex].

[214] A. Kraml Sabine Lessa, H. Reyes-González, and W. Waltenberger. http://smodels.github.io/wiki/SmsDictionary. 2024.

[215] R. Tyrell Rockafellar. "Convex Analysis". In: Princeton: Princeton University Press, 1970, pp. 20–40. ISBN: 9781400873173. DOI: doi:10.1515/9781400873173. URL: https://doi.org/10.1515/9781400873173.

[216] G. P. Villavicencio. "Polytopes And Graphs". English. In: Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2024, pp. 44–80. ISBN: 9781009257794.

[217] A. Kraml Sabine Lessa, H. Reyes-González, and W. Waltenberger. http://smodels.github.io/wiki/ListOfAnalyses. 2024.

[218] J. Alwall et al. "Computing Decay Rates For New Physics Theories With Feynrules And Madgraph 5_amc@NLO". In: *Comput. Phys. Commun.* 197 (2015), pp. 312–323. DOI: 10.1016/j.cpc.2015.08.031. arXiv: 1402.1178 [hep-ph].

[219] NNPDF Collaboration. "Parton Distributions With Qed Corrections". In: *Nucl. Phys. B* 877 (2013), pp. 290–320. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv: 1308.0598 [hep-ph].

[220] A. Buckley et al. "LHApdf6: Parton Density Access In The Lhc Precision Era". In: *Eur. Phys. J.* C75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[221] T. Sjostrand, S. Mrenna, and P. Z. Skands. "A Brief Introduction To Pythia 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[222] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet User Manual". In: *Eur.Phys.J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[223] B. C. Allanach et al. "SUSY Les Houches Accord 2". In: *Comput. Phys. Commun.* 180 (2009), pp. 8–25. DOI: 10.1016/j.cpc.2008.08.004. arXiv: 0801.0045 [hep-ph].

[224] E. Bothmann et al. *A Standard Convention For Particle-Level Monte Carlo Event-Variation Weights*. Mar. 2022. arXiv: 2203.08230 [hep-ph].

[225] C. J. Clopper and E. S. Pearson. "The Use Of Confidence Or Fiducial Limits Illustrated In The Case Of The Binomial". In: *Biometrika* 26.4 (Dec. 1934), pp. 404–413. ISSN: 0006-3444. DOI: 10.1093/biomet/26.4.404. eprint: https://academic.oup.com/biomet/article-pdf/26/4/404/823407/26-4-404.pdf. URL: https://doi.org/10.1093/biomet/26.4.404.

[226] A. Buckley et al. "The Simplified Likelihood Framework". In: *Journal of High Energy Physics* 2019.4 (2019). ISSN: 1029-8479. DOI: 10.1007/jhep04(2019)064.

[227] A. Wald. "Tests Of Statistical Hypotheses Concerning Several Parameters When The Number Of Observations Is Large". In: *Transactions of the American Mathematical Society* 54.3 (1943), p. 426. ISSN: 0002-9947. DOI: 10.2307/1990256.

[228] G. Cowan et al. "Asymptotic Formulae For Likelihood-Based Tests Of New Physics". In: *The European Physical Journal C* 71.2 (2011), pp. 1–19. ISSN: 1434-6044. DOI: 10.1140/epjc/s10052-011-1554-0. URL: https://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[229] P. Skands et al. "Susy Les Houches Accord: Interfacing Susy Spectrum Calculators, Decay Packages, And Event Generators". In: *Journal of High Energy Physics* 2004.07 (July 2004), pp. 036–036. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2004/07/036. URL: http://dx.doi.org/10.1088/1126-6708/2004/07/036.

[230] A. Buckley. *Pyslha: A Pythonic Interface To Susy Les Houches Accord Data.* 2013. arXiv: 1305.4194 [hep-ph].

[231] A. Kraml Sabine Lessa, H. Reyes-González, and W. Waltenberger. http://smodels.github.io. 2024.

[232] A. Kraml Sabine Lessa, H. Reyes-González, and W. Waltenberger. http://smodels.github.io/wiki/Validation. 2024.

[233] W. Waltenberger, A. Lessa, and S. Kraml. "Artificial Proto-Modelling: Building Precursors Of A Next Standard Model From Simplified Model Results". In: *Journal of High Energy Physics* 2021.3 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep03(2021)207.

[234] F. Ambrogi et al. "Smodels V1.2: Long-Lived Particles, Combination Of Signal Regions, And Other Novelties". In: *Computer Physics Communications* 251 (2020), p. 106848. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2019.07.013. URL: https://www.sciencedirect.com/science/article/pii/S0010465519302255.

[235] ATLAS Collaboration. "Summary Of The Atlas Experiment'S Sensitivity To Supersymmetry After Lhc Run 1 — Interpreted In The Phenomenological Mssm". In: *Journal of High Energy Physics* 2015.10 (2015). ISSN: 1029-8479. DOI: 10.1007/jhep10(2015)134.

[236] S. Meyn, R. L. Tweedie, and P. W. Glynn. "Markov Models". In: *Markov Chains and Stochastic Stability.* Cambridge Mathematical Library. Cambridge University Press, 2009, pp. 21–47, 48–74.

[237] C. Arina, B. Fuks, and L. Mantani. "A Universal Framework For T-Channel Dark Matter Models". In: *Eur. Phys. J. C* 80.5 (2020), p. 409. DOI: 10.1140/epjc/s10052-020-7933-7. arXiv: 2001.05024 [hep-ph].

[238] C. Arina et al. "Closing In On T-Channel Simplified Dark Matter Models". In: *Phys. Lett. B* 813 (2021), p. 136038. DOI: 10.1016/j.physletb.2020.136038. arXiv: 2010.07559 [hep-ph].

[239] B. Dumont et al. "Toward A Public Analysis Database For Lhc New Physics Searches Using Madanalysis 5". In: *Eur. Phys. J. C* 75.2 (2015), p. 56. DOI: 10.1140/epjc/s10052-014-3242-3. arXiv: 1407.3278 [hep-ph].

[240] B. Fuks, S. Banerjee, and B. Zaldivar. *Implementation Of A Search For Squarks And Gluinos In The Multi-Jet + Missing Energy Channel (3.2fb^{-1}, 13 Tev, Atlas-Susy-2015-06)*. Version V1. 2021. DOI: 10.14428/DVN/MHPXX4. URL: https://doi.org/10.14428/DVN/MHPXX4.

[241] G. Chalons and H. Reyes-Gonzalez. *Implementation Of A Search For Squarks And Gluinos In The Multi-Jet + Missing Energy Channel (36.1fb^{-1}, 13 Tev, Atlas-Susy-2016-07)*. Version V1. 2021. DOI: 10.14428/DVN/I5IFG8. URL: https://doi.org/10.14428/DVN/I5IFG8.

[242] F. Ambrogi and J. Sonneveld. *Implementation Of A Search For Supersymmetry In The Multi-Jet + Missing Energy Channel (35.9fb^{-1}, 13 Tev, Cms-Sus-16-033)*. Version V1. 2021. DOI: 10.14428/DVN/GBDC91. URL: https://doi.org/10.14428/DVN/GBDC91.

[243] M. Malte, S. Bein, and J. Sonneveld. *Re-Implementation Of A Search For Supersymmetry In The H_T/Missing H_T Channel (137fb^{-1}, Cms-Susy-19-006)*. Version V6. 2020. DOI: 10.14428/DVN/4DEJQM. URL: https://doi.org/10.14428/DVN/4DEJQM.

[244] M. Mrowietz, S. Bein, and J. Sonneveld. "Implementation Of The Cms-Sus-19-006 Analysis In The Madanalysis 5 Framework (Supersymmetry With Large Hadronic Activity And Missing Transverse Energy, 137fb^{-1})". In: *Mod. Phys. Lett. A* 36.01 (2021), p. 2141007. DOI: 10.1142/S0217732321410078.

[245] T. Sjöstrand et al. "An Introduction To Pythia 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[246] W. Buttinger. *Background Estimation With The ABCD Method Featuring The TRoofit Toolkit*. 2018. URL: https://api.semanticscholar.org/CorpusID:235806829.

[247] W. Beenakker. "Squark And Gluino Production At Hadron Colliders". In: *Nuclear Physics B* 492.1-2 (1997), pp. 51–103. ISSN: 0550-3213. DOI: 10.1016/s0550-3213(97)00084-9.

[248] A. Kulesza and L. Motyka. "Threshold Resummation For Squark-Antisquark And Gluino-Pair Production At The Lhc". In: *Physical Review Letters* 102.11 (2009). ISSN: 0031-9007. DOI: 10.1103/physrevlett.102.111802.

[249] A. Kulesza and L. Motyka. "Soft Gluon Resummation For The Production Of Gluino-Gluino And Squark-Antisquark Pairs At The Lhc". In: *Physical Review D* 80.9 (2009). ISSN: 1550-7998. DOI: 10.1103/physrevd.80.095004.

[250] W. Beenakker et al. "Soft-Gluon Resummation For Squark And Gluino Hadroproduction". In: *Journal of High Energy Physics* 2009.12 (Dec. 2009), pp. 041–041. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2009/12/041. URL: http://dx.doi.org/10.1088/1126-6708/2009/12/041.

[251] W. Beenakker et al. "Squark And Gluino Hadroproduction". In: *International Journal of Modern Physics A* 26.16 (2011), pp. 2637–2664. ISSN: 0217-751X. DOI: 10.1142/s0217751x11053560.

[252] W. Beenakker et al. "Stop Production At Hadron Colliders". In: *Nuclear Physics B* 515.1-2 (1998), pp. 3–14. ISSN: 0550-3213. DOI: 10.1016/s0550-3213(98)00014-5.

[253] W. Beenakker et al. "Supersymmetric Top And Bottom Squark Production At Hadron Colliders". In: *Journal of High Energy Physics* 2010.8 (2010). ISSN: 1029-8479. DOI: 10.1007/jhep08(2010)098.

[254] W. Beenakker et al. "Nlo+Nll Squark And Gluino Production Cross Sections With Threshold-Improved Parton Distributions". In: *The European Physical Journal C* 76.2 (Jan. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-3892-4. URL: http://dx.doi.org/10.1140/epjc/s10052-016-3892-4.

[255] CMS Collaboration. *Simplified Likelihood For The Re-Interpretation Of Public Cms Results*. Tech. rep. CMS-NOTE-2017-001. CERN-CMS-NOTE-2017-001. Geneva: CERN, Jan. 2017. URL: https://cds.cern.ch/record/2242860.

[256] H. Akaike. "A New Look At The Statistical Model Identification". English. In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.

[257] CMS Collaboration. "Search For Supersymmetric Partners Of Electrons And Muons In Proton–Proton Collisions At $\sqrt{s} = 13$ Tev". In: *Physics Letters B* 790 (Mar. 2019), pp. 140–166. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.01.005. URL: http://dx.doi.org/10.1016/j.physletb.2019.01.005.

[258] ATLAS Collaboration. "Search For Electroweak Production Of Supersymmetric Particles In Final States With Two Or Three Leptons At $\sqrt{s} = 13$ TeV With The Atlas Detector". In: *The European Physical Journal C* 78.12 (Dec. 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6423-7. URL: http://dx.doi.org/10.1140/epjc/s10052-018-6423-7.

[259] ATLAS Collaboration. "Search For Squarks And Gluinos In Final States With Jets And Missing Transverse Momentum". In: *Physical Review D* 97.11 (2018). ISSN: 2470-0010. DOI: 10.1103/physrevd.97.112001.

[260] A. Tumasyan et al. "Search For Higgsinos Decaying To Two Higgs Bosons And Missing Transverse Momentum In Proton-Proton Collisions At $\sqrt{s} = 13$ Tev". In: *Journal of High Energy Physics* 2022.5 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep05(2022)014.

[261] ATLAS Collaboration. "Search For Top-Squark Pair Production In Final States With One Lepton, Jets, And Missing Transverse Momentum Using 36fb^{-1} Of $\sqrt{s} = 13$ Tev pp Collision Data With The Atlas Detector". In: *JHEP* 06 (2018), p. 108. DOI: 10.1007/JHEP06(2018)108. arXiv: 1711.11520 [hep-ex].

[262] ATLAS Collaboration. "Searches For New Phenomena In Events With Two Leptons, Jets, And Missing Transverse Momentum In 139fb^{-1} Of $\sqrt{s} = 13$ Tev *pp* Collisions With The Atlas Detector". In: *the European Physical Journal C* (Apr. 2022). arXiv: 2204.13072 [hep-ex].

[263] CSM Collaboration. "Search For New Physics In The Multijet And Missing Transverse Momentum Final State In Proton-Proton Collisions At $\sqrt{s} = 8$ Tev". In: *JHEP* 06 (2014), p. 055. DOI: 10.1007/JHEP06(2014)055. arXiv: 1402.4770 [hep-ex].

[264] ATLAS Collaboration. "Search For Squarks And Gluinos With The Atlas Detector In Final States With Jets And Missing Transverse Momentum Using $\sqrt{s} = 8$ Tev Proton–Proton Collision Data". In: *JHEP* 09 (2014), p. 176. DOI: 10.1007/JHEP09(2014)176. arXiv: 1405.7875 [hep-ex].

[265] F. Feruglio. "Extra Dimensions In Particle Physics". In: *The European Physical Journal C* 33.S1 (2004), s114–s128. ISSN: 1434-6044. DOI: 10.1140/epjcd/s2004-03-1699-8. URL: https://dx.doi.org/10.1140/epjcd/s2004-03-1699-8.

[266] V. M. Abazov et al. "Search For Large Extra Dimensions In Themonojet+Et". In: *Physical Review Letters* 90.25 (2003). ISSN: 0031-9007. DOI: 10.1103/physrevlett.90.251802.

[267] E. A. Mirabelli, M. Perelstein, and M. E. Peskin. "Collider Signatures Of New Large Space Dimensions". In: *Physical Review Letters* 82.11 (1999), pp. 2236–2239. ISSN: 0031-9007. DOI: 10.1103/physrevlett.82.2236.

[268] C. Bierlich et al. "Robust Independent Validation Of Experiment And Theory: Rivet Version 3". In: *SciPost Physics* 8.2 (2020). ISSN: 2542-4653. DOI: 10.21468/scipostphys.8.2.026.

[269] ATLAS Collaboration. "Search For New Phenomena In Events With An Energetic Jet And Missing Transverse Momentum In pp Collisions At". In: *Physical Review D* 103.11 (2021). ISSN: 2470-0010. DOI: 10.1103/physrevd.103.112006.

[270] A. Tumasyan et al. "Search For New Particles In Events With Energetic Jets And Large Missing Transverse Momentum In Proton-Proton Collisions At $\sqrt{s} = 13$ Tev". In: *Journal of High Energy Physics* 2021.11 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep11(2021)153.

[271] A. Buckley. *The Hepthesis Latex Class*. 2010.

[272] G. Inc. *Grammarly*. https://www.grammarly.com/. Accessed: 2024-08-05. 2024.

[273] OpenAI. *Chatgpt: Gpt-4 Model*. https://www.openai.com/chatgpt. Accessed: 2024-08-05. 2024.

List of figures

List of tables