Parameter Scans and Machine Learning for beyond Standard Model Physics

Inga Strümke

Thesis for the Degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2019

UNIVERSITY OF BERGEN

Parameter Scans and Machine Learning for beyond Standard Model Physics

Inga Strümke

Thesis for the Degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defence: 02.05.2019

Year: 2019

Title: Parameter Scans and Machine Learning
for beyond Standard Model Physics

Name: Inga Strümke

Print: Skipnes Kommunikasjon / University of Bergen

Abstract

This thesis focuses on different beyond Standard Model theories, and the use of statistical methods to investigate them.

Supersymmetry is considered in different contexts. First, the supersymmetry breaking scheme of gaugino mediation is investigated. These models were previously thought to be ruled out after observation of the lightest Higgs with mass 125 GeV, but we show that non-vanishing trilinear terms can provide a sufficiently large Higgs mass via mixing in the stop sector.

Introducing the concept of machine learning, two models are investigated using different machine learning techniques. A two-Higgs doublet model with mass degenerate neutral Higgses is studied, and deep learning is used to separate the two $\mathcal{CP}$-states whose decay products have very similar kinematics. The challenge of prior dependence associated with problems relying on simulated training data and the probabilistic interpretation of classifier output is discussed. Second, another supersymmetric scenario, this time with a sneutrino as the lightest observable supersymmetric particle, is considered from a collider perspective. A boosted decision tree is used to detect signal events in datasets with small signal mixture parameters.

Later, a supersymmetric scenario with a gravitino as lightest and neutralino as next-to-lightest supersymmetric particle is considered in a cosmological context. The presence of late-decaying neutralinos can potentially come in conflict with constraints from Big Bang Nucleosynthesis, and we investigate a specific region in parameter space where resonant annihilation via a heavy Higgs lowers the neutralino relic abundance. In addition to this constraint, collider searches and the observed dark matter abundance are combined to form a likelihood which guides a scan through the parameter space.

Dark matter is further investigated from a model-independent viewpoint. The potentially sharp gamma-ray features in a signal from dark matter annihilation into Standard Model excited meson states are simulated, and we show how such a signature would stand out from the astrophysical background. The relevant energy range for this would be what has come to be called the “MeV-gap”, since this range is astrophysically relatively unexplored.

Acknowledgments

I would like to thank my supervisor professor Jörn Kersten, for motivating me in our projects and being patient through my various confusions, while also giving me the freedom to have and pursue my own ideas. Second, I would like to thank my co-supervisor professor Are Raklev for hospitality and collaboration at the UiO, and for helpful comments and proofreading. Thanks also to Jan Heisig, for good discussions and collaboration throughout the past four years.

Academia is a funny place. It gives us the opportunity to meet like-minded people and make friends for life — and then makes them move away every three years or so. Despite, or perhaps just because of, that, some of the most valued relations in my life have been made during my years at different universities. This is a warm thanks to everybody I had the pleasure of studying, having fun, getting lost and learning to live with at the NTNU in Trondeim, the Universidad de Valencia, the University of Bergen and the University of Oslo. Odd Rune, Rianne, Simen, Knut, Andreas, Jeriek, Peder, Haakon, Evelyn, Elena, Theresa — the names are of course too numerous to list. I would also like to thank again my first supervisor, professor Jens Oluf Andersen: I think this here is partly your fault.

A special thanks to Nick Murphy for being my favourite goofball, my brother from another mother and partner in SUSY-crime, and to Anders Kvellestad for always, even after long nights of debugging and application writing, being available for discussions about priors, likelihoods, life and the rest of the universe. A warm thanks also to Jonas Glesaaen for invaluable help in physic and programming, and a deep friendship — despite making me feel guilty for my lacking skills in `C++` template meta-programming.

To the women who have inspired me in ways that have surprised and delighted me: Thanks to professor Anna Lipniacka for being so real. So brilliant at what she does, so hard-working and still so naturally taking a week off for an Ironman. Thanks to professor Pilar Hernandez for inspiring and sternly helping me so much during my stay in Valencia that the impression has lasted until now and probably forever will. Thanks to professor Heidi Sandaker for being so open and warm, for the conversations that left me with the feeling that grants can be had and solutions can be found, and for being the first to welcome me at the UiO.

Quiero dar gracias a Pepita, Paquita y Miguel, por haberme adoptado, apoyado y enseñado que las cosas de palacio van despacio.

I also want to thank my mother, father and sister. As far as I have gathered, the current belief is that what we become is determined in equal parts by nature and nurture, meaning mostly family.

On the same note, I want to thank Norway, for a peaceful society and free education — and mountains, fjords and glaciers.

Lastly, a heartfelt thanks to my love Steffen. Little did we know, when you knocked on my office door, that day in May four years ago.

List of publications

Trilinear-augmented gaugino mediation
Co-authors: Jan Heisig, Jörn Kersten and Nick Murphy
Internal reference: [1]

Signal mixture estimation for degenerate heavy Higgses using a deep neural network
Co-authors: Anders Kvellestad and Steffen Maeland
Internal reference:[2]

Smoking gun dark matter signatures in the MeV range in
Science with e-ASTROGAM: A space mission for MeV-GeV gamma-ray astrophysics
Co-authors: Torsten Bringmann, Andre Hryczuk, Are Raklev and Jeriek Van den Abeele
Internal reference:[3]

Projects in progress

An open window for high reheating temperatures in gravitino dark matter scenarios
Co-authors: Jan Heisig, Jörn Kersten and Jeriek Van den Abeele
Internal reference:[4]

Gamma rays from the annihilation of dark matter into heavy mesons
Co-authors: Are Raklev and Jeriek Van den Abeele
Internal reference:[5]

Enabling sneutrino detection in weak signal scenarios using machine learning methods
Co-authors: Daniel Alvestad, Nikolai Fomin, Jörn Kersten
Internal reference:[6]

Code

Trilinear-augmented gaugino mediation
gitlab.com/Strumke/gaugino_mediation

Signal mixture estimation for degenerate heavy Higgses using a deep neural network
gitlab.com/BSML/ml-2hdm

An open window for high reheating temperatures in gravitino dark matter scenarios
gitlab.com/Strumke/gravitino_reheating

Gamma rays from the annihilation of dark matter into heavy mesons
gitlab.com/Strumke/dm_mesons

Enabling sneutrino detection in weak signal scenarios using machine learning methods
gitlab.com/BSML/sneutrinoml

Contents

Chapter 1

The short introduction

The Standard Model of particle physics comprises what we know about force and matter, and within its domain of validity, has been well tested and verified. Unfortunately, said domain of validity is not defined. Since it is not possible to probe arbitrarily short length scales — or extremely high energies — there is no telling what may be hiding there. Venturing to higher energy scales might reveal new particles, unifying forces, restoring symmetries, a physical model containing our Standard Model only as a subset, or something entirely different. The Standard Model of physics is a result of theoretical models invented, tested, discarded, adjusted and improved, of experiments conducted at ever increasing precision, magnitudes and energies, of measurements from laboratories beneath the ground and under ice, and of observations from telescopes pointed at the universe, detecting processes beyond our grasp in distance, time and energy scale. And it is consistent, but it is not complete.

The average photon in the universe is part of the Cosmic Microwave Background (CMB) radiation, and consequently has an energy of around $6.626534 \cdot 10^{-4}\,\mathrm{eV}$, corresponding to a temperature of 2.72 Kelvin [7]. The average place in the universe is thus, on human length scales, completely cold, dark and quiet. Dark because the CMB photons have so little energy that they fall well below the range of visible light. Cold because there is nothing but the CMB there to heat it up. Quiet because sound waves cannot propagate in empty space.

However, looking at the CMB with different eyes, e.g. through a large telescope, reveals 13.77 billion years old quantum fluctuations, indicating that the average place in the universe has not always been dark, cold and quiet. Quite the contrary, it shows that the universe has once been affected, if not dominated, by quantum fluctuations.

Looking closer at the empty space, were it possible, would reveal a busy environment of vibrating and interacting quantum fields. Looking even closer, perhaps past a quantum foam, could reveal the fabric of spacetime, if there is such

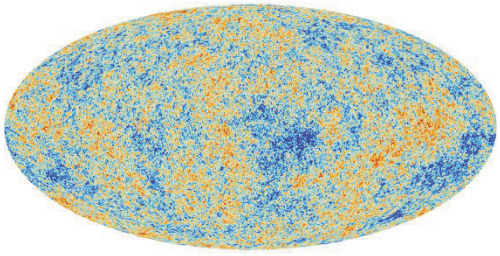


Figure 1.1: Planck data showing the CMB, which was created during the era of recombination, when the universe was around 400 000 years young. The irregularities are called primordial fluctuations, and they are density variations from the early universe. These are probably the origin of all the structures in the universe, and according to the model of inflation stem from quantum fluctuations. Image from sci.esa.int/planck.

a thing. Listening more closely, would sound the rumbling and chirping of gravitational waves.

It is not a given that it is practically possible to look close enough, i.e. to apply enough energy, for the basic building blocks of the universe to reveal themselves. Neither is it a given that there exists an ultimate theory which is not merely the low-energy reduction of yet another model, but rather the apex of theories describing the universe. Fortunately, there is even less reason to believe that we are approaching the limit of how close we can look. We can smash particles, look and listen at the universe, and we can do our best to come up with theories and ways to test them.

1.1 Our universe. . . so far

Many a young physicist has at some point gone to ICTP in Trieste, CERN in Geneva, or one of the many other large and inspiring physics institutions, and come back with a mug displaying some variation of

$$\begin{aligned} \mathscr{L} &= \sqrt{g} \quad \Big\{ R - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &\qquad + i\bar{\psi} \not{D} \psi + \bar{\Psi}_i \mathcal{Y}_{ij} \phi \psi_j + \text{h.c.} \\ &\qquad + |\mathcal{D}_\mu \phi|^2 - V(\phi) \Big\} \\ &\cong \quad \text{Our universe} \ldots \text{so far}. \end{aligned}$$

Although this popular formulation is somewhat simplified, it does encapsulate what is known about the universe, so far. The first term of the expression is perhaps the one that looks the least threatening, but it denotes the curvature of spacetime and represents the theory of gravity, which has not been unified within one quantum field theory together with the three other fundamental forces, so far. The second term contains these fundamental forces; electromagnetism, the weak and the strong nuclear forces. In the Standard Model, these forces are carried by mediating particles all belonging to the family of bosons. Moving on to the second line, there are the fermions, represented by spinors. These have half-integer spins, and constitute all matter particles observed, so far. The next terms contain the only scalar particle in the Standard Model, the Higgs boson. The spontaneously broken symmetry of the Higgs potential, see fig. 1.2, results in a non-zero vacuum expectation value of the Higgs field, which in turn provides all particles which interact with the Higgs, a mass. All in all, these are the particle interactions in the universe which particle physics is able to explain, so far. The Standard Model particle content and interactions are artistically represented in fig. 1.3.

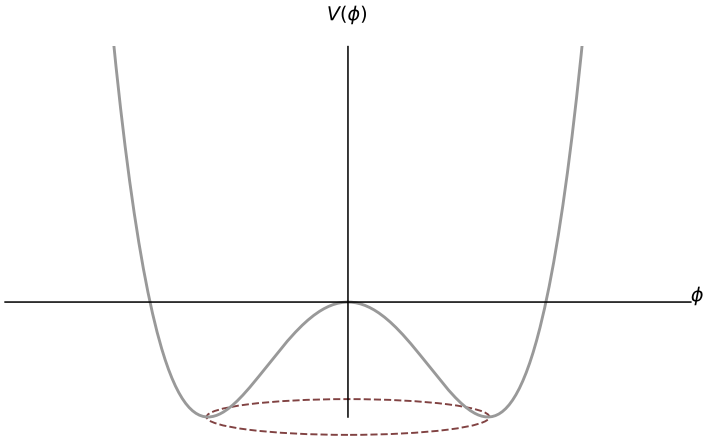


Figure 1.2: A sketch of the Higgs potential $V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$, where the characteristic "Mexican hat" shape, here projected onto two dimensions, arises when $\mu^2 < 0$ and $\lambda > 0$. The dark red circle symbolises a continuum of vacuum states.

1.2 . . . and beyond

There are of course many ways to approach our favourite question **What is beyond the Standard Model?**, and the recurring strategy throughout the present work is

1. Choose a model
2. Make predictions
3. Identify allowed parameters
4. If possible, aid in detection
5. Hope for observation. Have a go at another model in the meantime

Item 1 in this list can for instance be approached by trying to explain observations and phenomena or answer questions currently beyond the Standard Model. At the time of writing, among the most prominent of these concern, in no particular order, dark matter, dark energy, neutrino masses, whether gravity can be quantized and unified with the other fundamental forces, the matter-antimatter asymmetry in the universe, inflation and the hierarchy problem. All these open questions suggest that the Standard Model must be extended, and a candidate model should address at least some of these, while performing as well as the Standard Model within known regimes.

Item 2 involves performing calculations or doing simulations. A model's prediction can for instance be the physical mass spectrum, which can be calculated numerically using programs such as `SPheno` [8], used in *Trilinear-augmented gaugino mediation* [1], `Softsusy` [9], used in *An open window for high reheating temperatures in gravitino dark matter scenarios* [4] and *Enabling sneutrino detection in weak signal scenarios using machine learning methods* [6], and `SuSpect` [10]. For simulating collider events and decays, some of the available tools are `Pythia` [11, 12], used in *Signal mixture estimation for degenerate heavy Higgses using a deep neural network* [2], *Smoking gun dark matter signatures in the MeV range* in [3], and `MadGraph` [13], used in [1] and [4].

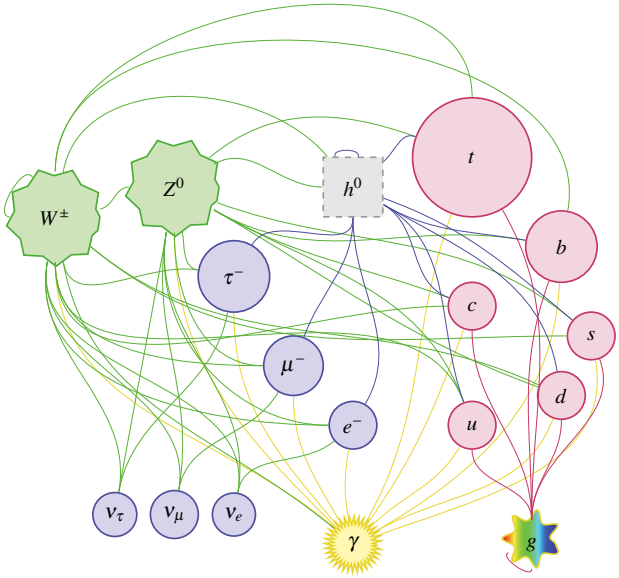


Figure 1.3: The matter particles, Higgs boson and force carriers of the Standard Model. The lines connect interacting particles, and the invisible y-axis as well as the circle diameters indicate roughly the mass hierarchy.

Item 3 can be done by scanning over the free parameters of a model and see how well its predictions fit with observations. This can be done in several ways, the simplest being via a grid scan which divides the parameter space into excluded and allowed regions, as is done in [1] and [2], and more sophisticated methods involving calculating likelihoods and selectively exploring the parameter space, as is done in [4].

Item 4 can amount to many things; in [1], allowed and excluded regions of the model's parameter space are presented; in [2], a machine learning method for better separating mass degenerate $\mathcal{CP}$ states is presented; in [3], astrophysical gamma-ray signals from dark matter annihilation are simulated; in [6], machine learning and statistical methods for detecting signals from supersymmetric processes are proposed.

The following contains an introduction to all the topics covered in the publications, and discussions of the currently unpublished results. The objective of this text is to enable anybody with a degree in physics or mathematics to understand all the results in this work and how to obtain them.

Chapter 2

The long introduction

> The importance of symmetry in modern physics cannot be overstated
>
> *A. Zee [14]*

2.1 Quantum fields

Quantum field theory is the union of quantum mechanics and special relativity into one mathematically powerful, and equally heavy, framework, where the basic building blocks of nature and their mutual interactions are described as fields with different charges.

The dynamics of a D-dimensional physical system are contained in the action S, a functional of the fields, represented as a spacetime integral over the Lagrangian density along the path of the system,

$$S \equiv \int d^D x \, \mathscr{L}(\Phi_i(x), \partial_\mu \Phi_i(x)) \,, \tag{2.1}$$

meaning that all the properties of the i fields in in the system as well as their interactions must be contained within the Lagrangian density. The above equation evidently imposes a restriction on the mass dimension of the Lagrangian density, reciprocal to the dimension of the theory. Furthermore, the Lagrangian density can contain only terms invariant under Lorentz[1] transformations in the D-dimensional spacetime as well as any internal symmetries.

2.1.1 Scalar fields

The Lagrangian is a function of the fields and their derivatives, and the simplest kind of field is a scalar field. A scalar field describes a particle carrying zero intrinsic angular momentum, or spin, and is invariant under Lorentz transformations. The only scalar particle in the Standard Model is the Higgs boson.

Possible terms in the Lagrangian describing a simple theory containing a single scalar field $\phi(x)$ are

$$\mathscr{L} = \partial_\mu \phi \partial^\mu \phi + t\phi + m^2 \phi^2 + g^3_{\text{tri}} \phi^3 + g^4_{\text{quart}} \phi^4 \,. \tag{2.2}$$

Feynman graphs for the processes in this Lagrangian are depicted in fig. 2.1. The first term in eq. (2.2) is a kinetic term, and helps ordain the dimensionality of the scalar field itself, by way of the dimensionality condition of the Lagrangian density, see eq. (2.36). Once the dimension of the field is set, so can the dimensions of the different couplings. In a renormalisable theory, the couplings must be either dimensionless or have positive mass dimension, which is why it

[1] The Lorentz group is the group of all transformations which leave the quantity $\eta_{\mu\nu} x^\mu x^\nu$ invariant, see section 2.2.

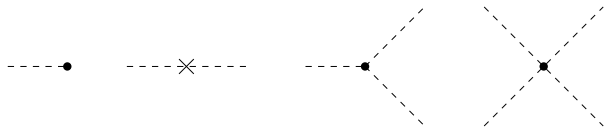


Figure 2.1: The tadpole, mass insertion and three- and four-point interactions of a scalar field.

is not always possible to keep adding higher power terms in the Lagrangian. The second term in eq. (2.2) is called a tadpole term, and is allowed in this form since the scalar field is Lorentz invariant on its own, and provided that it is not charged under any other symmetries.

The third term represents a mass, as indicated by the notation, made obvious by the fact that the parameter m must have mass dimension one. The subsequent terms represent interactions of coupling strengths g, and in this simple example containing only one scalar field, there are of course only self-interactions.

Considering only the free part of the scalar Lagrangian, i.e. the mass and kinetic terms, and calculating the equations of motion by solving the Euler-Lagrange equation

$$\frac{\partial \mathscr{L}}{\partial \phi_i} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial (\partial_\mu \phi_i)} \right) = 0 \,, \tag{2.3}$$

yields the Klein-Gordon equation

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0 \,. \tag{2.4}$$

This equations also holds for pseudoscalar fields, meaning scalar fields which change sign under parity inversion.

Bose-Einstein statistics

Since the scalar field has spin zero, it belongs to the family of bosons. The perhaps most striking property of bosons is that an unlimited number of them can occupy the same place and quantum state — an ability often mistakenly attributed to master students. Collections of bosons are described by Bose-Einstein statistics, which assumes identical and non-interacting particles, and gives the expected number of particles n_i in energy state i as

$$n_i(E_i) = \frac{g_i}{e^{\frac{E_i - \mu}{kT}} - 1} \quad , \quad E_i > \mu \,, \tag{2.5}$$

where E_i is the energy of the i'th state, μ the chemical potential, g_i the degeneracy of energy level i, k the Boltzmann constant and T the temperature.

2.1.2 Vector fields

Leaving behind the scalar field, all bosons in the Standard Model apart from the Higgs boson, are found one step up the spin ladder. They have spin value one, i.e. they are represented by vectors, are therefore called vector bosons, and carry a Lorentz index. For a quick illustration, consider a free photon, a massless vector boson which interacts electromagnetically, described by the simple Lagrangian

$$\mathscr{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,. \tag{2.6}$$

The electromagnetic field strength tensor is defined as $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, and the resulting equations of motion are

$$\partial_\mu F^{\mu\nu} = 0 \,, \tag{2.7}$$

which yield the Maxwell equations when defining $A^\mu = (\phi, \mathbf{A})$.

From here, it is of course possible to keep climbing the spin ladder, and the next step would lead to the graviton, the hypothesised quantum of gravity, which would need to have a spin eigenvalue of two, since gravitational waves are tensor waves in four dimensions — but that topic is (unfortunately!) far beyond the scope of the present work. Also, massless particles of spin higher than one, and massive particles of spin higher than *and including* one, render their theory non-renormalisable. The presence of the massive spin one $W^{\pm}$ and Z^0 in the Standard Model particle zoo thus advertises a longer mathematical story. But first: How vector bosons connect via symmetries to the remaining particles.

2.2 Symmetries

Simply put, a symmetry is an invariance under a set of transformations. Exactly *what* is invariant is certainly a crucial point, and in particle physics it is usually the Lagrangian.

The Standard Model itself is a collection of three symmetry groups, and the following section contains a brief introduction to the relevant group theoretical concepts and an overview of the Standard Model groups.

2.2.1 Symmetry groups

A Lie group is a group whose elements are specified by one or more continuous parameters which vary smoothly. All groups encountered in the present work are matrix Lie groups. These are subgroups of the general linear group $\mathrm{GL}(n;\mathbb{C})$, containing all invertible $n \times n$ matrices with complex entries. Every Lie group has an associated Lie algebra, whose dimension is equal to the dimension of the group itself.

Denoting a basis of the Lie algebra by the $n \times n$ matrices T^a, referred to as the generators of the group, the algebra is classified by the relation

$$[T^a, T^b] = f_c^{ab} T^c \,, \tag{2.8}$$

where the f_c^{ab} are called the structure constants of the Lie algebra. Such a relation is sometimes called a Lie bracket. For non-Abelian groups, meaning groups whose elements do not commute, the generators are normalised such that

$$\mathrm{Tr}\, T^a T^b = \frac{1}{2}\delta^{ab} \,. \tag{2.9}$$

Elements of the Lie group are obtained from elements of the Lie algebra through

$$\begin{array}{ccc} \textbf{Lie algebra} & \xrightarrow{\exp} & \textbf{Lie group} \\ \exp(i\omega_a T^a) & \in & SU(N) \,. \end{array}$$

In physics, a symmetry is an operation or transformation which does not change the laws of the system, and a symmetry group the group of such transformations.

External symmetries

Symmetries involving transformations of spacetime itself are referred to as external symmetries. The aforementioned Lorentz transformations belong to the Lorentz group: The matrix Lie group of linear transformations of the $(1+3)$-dimensional real vector space which leave $\eta_{\mu\nu} x^\mu x^\nu$ invariant, also known as the orthogonal group $O(1,3)$. Together, spacetime translations and the Lorentz transformations, i.e. boosts and rotations,

$$x^\mu \to x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu \,, \tag{2.10}$$

generate the Poincaré group, the symmetry group of special relativity.

If P^μ are the generators of translation, and $M^{\mu\nu}$ the generators of Lorentz transformations, then these together are the generators for the Poincaré group, and the algebra of the group is given by the commutation relations

$$\begin{aligned}
[P^\mu, P^\nu] &= 0 \\
[M^{\mu\nu}, P^\sigma] &= i\left(\eta^{\mu\sigma}P^\nu - \eta^{\nu\sigma}P^\mu\right) \\
[M^{\mu\nu}, M^{\rho\sigma}] &= i\left(\eta^{\mu\rho}M^{\nu\sigma} - \eta^{\mu\sigma}M^{\nu\rho} - \eta^{\nu\rho}M^{\mu\sigma} + \eta^{\nu\sigma}M^{\mu\rho}\right).
\end{aligned} \tag{2.11}$$

Internal symmetries

Perhaps more relevant for particle physics are internal symmetries, which act on the fields themselves. The symmetry groups which give rise to the three fundamental forces in the Standard Model are

$$SU(3)_C \times SU(2)_L \times U(1)_Y . \tag{2.12}$$

Here, $SU(N)$ denotes the special unitary group, containing $n \times n$ unitary matrices with unit determinant. $SU(N)$ has $N^2 - 1$ independent generators and is a subgroup of $U(N)$, which contains all $n \times n$ unitary matrices. These are again subgroups of $\mathrm{GL}(n;\mathbb{C})$, so

$$SU(N) \subset U(N) \subset GL(n;\mathbb{C}) . \tag{2.13}$$

The three symmetry groups of the fundamental particle interactions in the Standard Model are briefly presented in the following.

2.2.2 Interactions

A gauge transformation is a local symmetry transformation which leaves the Lagrangian of a field theory invariant, and the set of such transformations together form a Lie group, in short referred to as the gauge group of the theory. A theory can be represented in different gauges, connected via gauge transformations.

The generators of a gauge group describing a field theory are themselves fields, referred to as gauge fields. These transform under the vector representation of the Lorentz group, meaning that they are vector fields. Hence, upon quantisation the quanta of these fields are bosons with spin quantum number one, and this is where the vector bosons, or gauge bosons, come from.

Quantum electrodynamics

The gauge boson of the electromagnetic force is the photon, which interacts with all particles carrying electric charge Q. The gauge group is that that of local $U(1)$ transformations, meaning phase shifts.

As quantum electrodynamics is unified with the weak force, the weak hypercharge Y is defined, which relates the electric charge and the third component of weak isospin via $T_3 = Q - \frac{Y}{2}$,[2]. In contrast to the symmetry groups for the other fundamental interactions, $U(1)_Y$ is Abelian.

Weak interactions

These interactions are sometimes called quantum flavourdynamics, but for historical reasons the name never really stuck. The charge is flavour, and the weak interaction discriminates between chiralities, see eq. (2.38), in such a way that only left-chiral fields are charged under it, while right-chiral fields are weak singlets.

The gauge bosons are the $W^\pm$ and Z^0. Since the Z^0 is electrically neutral, it couples fermion-antifermion pairs. The exchange of a $W^\pm$ boson changes the flavour of e.g. an electron to an electron neutrino, or a down- to an up-quark. Accordingly, left-chiral fermions form doublets in which the components are treated as different states of the same

[2] Sometimes, the expression is given as $T_3 = Q - Y$, implying that the hypercharge eigenvalues are expressed as $\pm\frac{1}{2}$ and not ± 1.

particle. These states thus have opposite isospin values, and since the number of flavour states available is two, the gauge group of this interaction is $SU(2)_L$. The generators of the group are the $2^2 - 1 = 3$ Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.14}$$

The reason behind the nomer "weak" is that the interaction strength is several orders of magnitude lower than that of the other fundamental forces. This is because the gauge bosons are massive, and decay after a short while. Gauge boson masses violate gauge symmetry, and $SU(2)_L$ is indeed broken at low energies. The symmetry breaking is discussed in section 2.3.

Quantum chromodynamics

Quarks carry a charge commonly referred to as colour charge — hence the "chromo". It can be in one of three states, and by an analogy extended past the familiar positive and negative charges of electromagnetism, the three possible colour states are red, green and blue. The three different states call for a three-dimensional group, and so $SU(3)$ is used to describe these interactions. Particles which carry colour are said to be charged under $SU(3)_C$.

The gauge fields of $SU(3)_C$ are the gluons, here denoted A^a_μ, and there are $3^2 - 1 = 8$ of them. The Lie algebra of the group is spanned by the Gell-Mann matrices λ^a, $a = 1, \ldots, 8$, of which only the first three are listed here

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.15}$$

The gluons are also charged under $SU(3)_C$, which causes loops of virtual gluons in the vacuum to polarise it in such a way that the colour field is augmented. This effect diminishes with proximity to a colour-charged particle, or correspondingly at high energy, which gives rise to the phenomenon of asymptotic freedom. Due to this, QCD is commonly referred to as the strong force; its strength increases over distance, and the result is that unbound coloured states cannot exist. Consequently, the coupling constant g_s of the strong force decreases with energy.

The covariant derivative

The ordinary partial derivative ∂_μ is not gauge invariant, so in the context of a gauge theory, it must be coupled to the gauge fields and made into a vector operator, so that the equations containing it keep their physical properties under gauge transformations.

The gauge covariant derivative $\mathcal{D}_\mu$, which transforms covariantly under the gauge transformations of the Standard Model, is

$$\mathcal{D}_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} A^a_\mu(x) + ig \frac{\sigma_i}{2} W^i_\mu(x) + ig' \frac{Y}{2} B_\mu(x). \tag{2.16}$$

Here, g_s, g and g' are the strong, weak and electromagnetic couplings respectively, A^a_μ are the gluon fields, W^i_μ the weak fields, B_μ the electromagnetic field and $a = 1, \ldots, 8$ and $i = 1, 2, 3$ run over the generators of $SU(3)_C$ and $SU(2)_L$ respectively.

As a side note, the gauge covariant derivative is an analogue to the covariant derivative in general relativity, which does by contrast require a metric. Gauge symmetries do not necessarily come with the concept of a metric tensor, so the gauge covariant derivative should rather be interpreted as an affine connection[3]. Physical events take place at points in spacetime, which is mathematically speaking a smooth manifold, meaning that it locally looks like a smooth deformation of Euclidean space — just like e.g. a differentiable curve locally looks like a deformed line. Vector fields can easily be differentiated in Euclidean space, since the tangent space of vectors at different points are connected via translation. This is not necessarily true for any manifold in general, since nearby tangent spaces are not necessarily easily connected. The affine connection connects these nearby tangent spaces.

[3] It is useful to think of the curvature of the affine connection as the field strength of the gauge potential.

2.3 Spontaneously broken symmetry

Gauge bosons are massless as long as their gauge symmetry is unbroken. This is best demonstrated by directly considering a mass term in the Lagrangian for some gauge field A_μ, which would be on the form

$$\mathscr{L}^A_{mass} = m_A^2 A_\mu A^\mu \,, \tag{2.17}$$

and a general gauge transformation

$$A_\mu \to A_\mu + \partial_\mu \phi \,, \tag{2.18}$$

for some scalar field ϕ. Under this transformation, the mass term becomes

$$m_A^2 A_\mu A^\mu \to m_A^2 A_\mu A^\mu + 2m_A^2 A_\mu \partial^\mu \phi + \partial_\mu \phi \partial^\mu \phi \,, \tag{2.19}$$

i.e. not the same as eq. (2.17), which is consequently not gauge invariant. The conclusion must be that gauge boson mass terms cannot be allowed in the Lagrangian. However, the masses of the electroweak gauge bosons Z^0 and $W^\pm$ is an experimental fact, and must have come about somehow. The answer is the well-known Higgs mechanism, which is briefly reviewed in the following.

2.3.1 The Higgs and what it does

The Higgs boson is the only scalar particle in the Standard Model, and hence the only particle which can acquire a vacuum expectation value, short VEV, meaning a non-zero value of the field in the vacuum. A VEV, being a constant value, is not Lorentz invariant, so only Lorentz scalars — a scalar field or a scalar combination of fields — which are singlets under unbroken gauge groups can have a non-zero VEV.

To see what happens for the Higgs, denote by ϕ the Higgs $SU(2)_L$ doublet,[4] which can be most generally expressed as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ \eta_3(x) + i\eta_4(x) \end{pmatrix}, \tag{2.20}$$

where the η_i are all real fields and each have one degree of freedom. Being an $SU(2)_L$ doublet, the top component has $T_3 = \frac{1}{2}$, while the lower component has $T_3 = -\frac{1}{2}$. As the Standard Model Higgs is observed electrically neutral, one of the components must furthermore have $Q = 0$. Since the whole doublet has hypercharge $Y = \frac{1}{2}$,[5] and using $T_3 = Q - Y$,[6] the only possibility is that the top component have $Q = 1$, i.e. carries a positive electric charge.

Performing an $SU(2)_L$ and a $U(1)_Y$ gauge transformation, this doublet can be transformed into the form

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \sigma(x) \end{pmatrix}, \tag{2.21}$$

where v is a constant and $\sigma(x)$ is again a real field, in the so-called unitary gauge.

The part of the Lagrangian containing the Higgs is

$$\mathscr{L}_{\text{Higgs}} = \left(\mathcal{D}_\mu \phi\right)^\dagger \left(\mathcal{D}^\mu \phi\right) - \mathcal{V}(\phi) \,, \tag{2.22}$$

where the Higgs potential $\mathcal{V}(\phi)$ is required by renormalisability and invariance under $SU(2)_L$ and $U(1)_Y$ to be on the form

$$\mathcal{V}(\phi) = -\mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 , \tag{2.23}$$

with $\lambda > 0$ for vacuum stability. There are two possibilities for the value of μ^2, determining the overall shape of the potential. If $\mu^2 > 0$, the minimum of the potential does not correspond to the minimum value of the field ϕ, which is to

[4] The Higgs field ϕ of course has nothing to do with the expository gauge transformation in the beginning of the section.

[5] This is by construction, for the Higgs to couple in a gauge invariant way to the Standard Model fermions.

[6] Another convention also used in some textbooks is $T_3 = Q - \frac{Y}{2}$, but in that case the half is not included in the hypercharge value itself, i.e. the Higgs doublet has $Y = 1$ in that convention.

say that the Higgs field has a non-zero expectation value in the vacuum; a VEV. Calculating the minimum of $\mathcal{V}(\phi)$ yields an expression for the vacuum expectation value, here expressed in the unitary gauge,

$$\langle\phi\rangle = \frac{1}{2}\begin{pmatrix}0\\ v\end{pmatrix}e^{i\theta}\,; \qquad v = \left(\frac{-\mu^2}{2\lambda}\right)^{1/2}. \tag{2.24}$$

The phase factor $e^{i\theta}$ reflects that the potential has a continuous circle of minima, so as the Higgs field enters the vacuum state, or any state of sufficiently low energy, a direction θ is inadvertently chosen. The value of θ is not measurable, but the vacuum state is no longer symmetric. This is referred to as spontaneous symmetry breaking, and is what happens when the Lagrangian retains its symmetry, while the vacuum state breaks it. Note that only the lower component, which is the electrically neutral one, gets a VEV, while the upper component is zero in the vacuum. This ensures that the vacuum does not break electromagnetism. In group theory language, the breaking scheme is

$$SU(2)_L \otimes U(1)_Y \to U(1)_Q\,. \tag{2.25}$$

The generation of the gauge boson masses is now a straightforward algebraic exercise: Insert the Higgs doublet on the form of eq. (2.21) into the kinetic part of the Higgs Lagrangian — which is always the way to see how a field couples to the gauge fields, since these are contained in the covariant derivative. Using the expression in eq. (2.16) yields

$$\begin{aligned}\left(\mathcal{D}_\mu\phi\right)^\dagger\left(\mathcal{D}^\mu\phi\right) &= \frac{1}{2}\left|\left(\partial_\mu + ig\frac{\sigma_i}{2}W_\mu^i(x) + ig'\frac{1}{2}B_\mu(x)\right)\begin{pmatrix}0\\ v\end{pmatrix}\right|^2 \\ &= \frac{v^2}{8}\left[g^2\left(\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2\right) + \left(gW_\mu^3 - g'B_\mu\right)^2\right],\end{aligned} \tag{2.26}$$

having omitted the coloured part and substituted $Y = \frac{1}{2}$. The mass eigenstates and the corresponding masses of the electroweak gauge bosons are thus

$$\begin{aligned}W_\mu^\pm &\equiv \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp W_\mu^2\right), \quad m_W = \frac{gv}{2} \\ Z_\mu &\equiv \frac{1}{\sqrt{g^2+g'^2}}\left(gW_\mu^3 - g'B_\mu\right), \quad m_Z = \frac{v}{2}\sqrt{g^2+g'^2} \\ A_\mu &\equiv \frac{1}{\sqrt{g^2+g'^2}}\left(g'W_\mu^3 + gB_\mu\right), \quad m_A = 0\,.\end{aligned} \tag{2.27}$$

In the beginning of the present section, the doublet in eq. (2.20) had four degrees of freedom, while the four electroweak gauge bosons B and W^i were massless and thus had two degrees of freedom each. After the spontaneous symmetry breaking, the real scalar Higgs field has one degree of freedom, the three massive gauge bosons Z^0 and $W^\pm$ have three degrees of freedom each, and the photon, which remains massless, still has its two degrees of freedom. In total, there are twelve degrees of freedom before as well as after the symmetry breaking.

Lastly, a quick demonstration of how the fermion masses come about. The fermions interact with the Higgs in so-called Yukawa terms, which are generally speaking terms which couple Dirac and scalar fields. Compactly expressed, the Yukawa part of the Standard Model Lagrangian is

$$\mathscr{L}_{\text{Yukawa}} = \bar{\Psi}^f_{iL}\mathcal{Y}^f_{ij}\phi\,\psi^f_{jR} + \text{h.c.}\,, \tag{2.28}$$

where L and R denote left- and right-handed fields, see section 2.4 for details, f denotes fermion type, and i, j run over the three generations. The capital Ψ^f_{iL} represent $SU(2)_L$ doublets, which are in the Standard Model

$$\begin{aligned}\Psi^L_{iL} &= \begin{pmatrix}\nu_{eL}\\ e_L\end{pmatrix}, \begin{pmatrix}\nu_{\mu L}\\ \mu_L\end{pmatrix}, \begin{pmatrix}\nu_{\tau L}\\ \tau_L\end{pmatrix} \quad \text{Left-handed leptons} \\ \Psi^Q_{iL} &= \begin{pmatrix}u_L\\ d_L\end{pmatrix}, \begin{pmatrix}c_L\\ s_L\end{pmatrix}, \begin{pmatrix}b_L\\ b_L\end{pmatrix} \quad \text{Left-handed quarks}\,,\end{aligned} \tag{2.29}$$

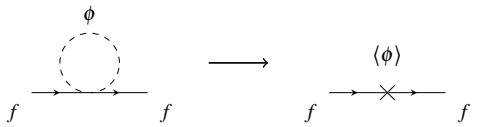


Figure 2.2: A Feynman diagram depiction of the result of spontaneous symmetry breaking: Before, the Higgs couples to the fermions through the term in eq. (2.28). After, the Higgs field has a VEV, which may be interpreted as a mass insertion.

while the ψ^f_{iR} represent $SU(2)_L$ singlets, which are

$$\begin{aligned} \psi^e_{iR} &= e_R, \mu_R, \tau_R && \text{Right-handed leptons} \\ \psi^u_{iR} &= u_R, c_R, t_R && \text{Right-handed up-quarks} \\ \psi^d_{iR} &= d_R, s_R, b_R && \text{Right-handed down-quarks}. \end{aligned} \tag{2.30}$$

Note that this does, at the time of writing, not include neutrinos, as right-handed neutrinos are not yet a part of the Standard Model, although neutrino oscillations have been observed, indicating that they do have a mass, and thus a right-handed component. Though, for the time being they are not included, do not have a right-handed spinor representing them in the Standard Model, and thus no term like eq. (2.28).

The charges of the Standard Model fields are listed in table 2.1. The Higgs doublet, a left-handed spinor and the corresponding right-handed singlet can be combined to terms with zero charge for the down-type fermions. For the up-type fermions, the Higgs must couple in an $SU(2)_L$ invariant way, which can be achieved by defining

$$\tilde{\phi} \equiv -i\left(\phi^\dagger \sigma_2\right)^T; \quad \langle\tilde{\phi}\rangle = \frac{1}{2}\begin{pmatrix} v \\ 0 \end{pmatrix}. \tag{2.31}$$

The term in eq. (2.28) now gives rise to interaction terms between the Higgs and fermions, and fermion mass terms containing v. These are all on the form

$$m^f_i = \mathcal{Y}^f_i \frac{v}{\sqrt{2}}. \tag{2.32}$$

At high energy scales, above the electroweak scale at 264 GeV, electroweak symmetry is restored and $v \to 0$, implying from the above equation and eq. (2.27) that the fermions as well as the electroweak gauge bosons are massless.

The Yukawa matrices have dimension 3×3, can in general be complex and are not necessarily hermitian. Were it not for the gauge interactions, these matrices could be diagonalized and the masses would be the only free parameters. In that basis, referred to as the mass basis, the kinetic terms are however non-diagonal, while they are diagonal in the interaction basis. The interaction and mass eigenstates thus do not have the same eigenbasis, which causes the fermions to mix. The mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which has three angles and one phase which cannot be eliminated by rotations, and are thus physically measurable. The angles, and thus also the off-diagonal terms, are quite small, and the mass and interaction eigenstates very close.

2.4 Fermions matter

It is now more or less clear how the fermions acquire their masses, but that still leaves more than half their tale untold. The following section first describes fermions from the perspective of their history in physics, before taking a more fundamental perspective, relating spinors to the group theory aspect.

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
e_{iL}	0	-$\frac{1}{2}$	-$\frac{1}{2}$
ν_{iL}	0	$\frac{1}{2}$	-$\frac{1}{2}$
$\bar{e}_{iR}$	0	0	-1
u_{iL}	$\frac{1}{2}$ (1,-1,0)	$\frac{1}{2}$	$\frac{1}{6}$
d_{iL}	$\frac{1}{2}$ (1,-1,0)	-$\frac{1}{2}$	$\frac{1}{6}$
$\bar{u}_{iR}$	$\frac{1}{2}$ (1,-1,0)	0	$\frac{2}{3}$
$\bar{d}_{iR}$	$\frac{1}{2}$ (1,-1,0)	0	$-\frac{1}{3}$
H_u	0	$\frac{1}{2}$	$\frac{1}{2}$
H_d	0	-$\frac{1}{2}$	$\frac{1}{2}$

Table 2.1: The charges of the Standard Model fermion and Higgs fields under the gauge groups. The fermion generation index $i = 1,2,3$ runs over (up, strange, top) and (down, charm, bottom) quarks, and (electron, muon, tau) leptons. The colour charges correspond to $\frac{1}{2}$ the eigenvalues of λ_3 in eq. (2.15), but it would mess up the table to write the coloured quarks separately. The spin quantum number corresponds to $\frac{1}{2}$ the eigenvalues of σ_3 in eq. (2.14). The hypercharge is defined as $Y = Q - L$.

2.4.1 Dirac, Weyl and Majorana

In his 1928 paper "The Quantum Theory of the Electron" [15], Dirac first described massive electrons and their antiparticles, and thus laid the foundation for describing fermions in a physical theory. The Lagrangian of a free, massive Dirac field is

$$\mathscr{L} = \bar{\Psi}\left(i\partial\!\!\!/ - m\right)\Psi\,, \tag{2.33}$$

and the equation of motion is

$$\left(i\partial\!\!\!/ - m\right)\Psi = 0\,, \tag{2.34}$$

know as the Dirac equation. The "Feynman slash" notation is shorthand for $\partial\!\!\!/ \equiv \gamma^\mu \partial_\mu$, and the γ's are 4×4 matrices which satisfy

$$\begin{aligned} [\gamma^\mu, \gamma^\nu] &= 2g^{\mu\nu} \\ \gamma_0\gamma_\mu\gamma_0 &= \gamma_\mu^\dagger\,, \end{aligned} \tag{2.35}$$

where the first one is necessary for the Dirac equation to comply with the energy-momentum relation, and the second one for the Hamiltonian implied by the Dirac Lagrangian to be real. There are of course a number of matrices which together obey the above relations, and they are related to each other by change of basis. The objects represented by Ψ are referred to as spinors, obey $\bar{\Psi} \equiv \Psi^\dagger\gamma^0$, so they have vector structure with four entries, and when multiplied by a plane wave are solutions to the Dirac equation. Using again the dimensionality argument, the first term in the Dirac Lagrangian is required canonically normalised, i.e. it must obey

$$[\partial_\mu\phi\,\partial^\mu\phi] = [\bar{\psi}\partial\!\!\!/\psi] = [\mathscr{L}] = D\,, \tag{2.36}$$

which explains why the mass parameter in eq. (2.33) is not squared, as opposed to in the scalar case.

Two years after Dirac's publication, Weyl published "Gravitation and the electron" [16], where he stated that Dirac's equation

> " . . . contains the mass m of the electron as a factor. But mass is a gravitational effect: it is the flux of the gravitational field through a surface enclosing the particle in the same sense that charge is the flux of the electric field. In a satisfactory theory it must therefore be as impossible to introduce a non-vanishing mass without the gravitational field as it is to introduce charge without electromagnetic field. "

Motivated by this, he showed that a simpler equation containing two-component fields suffices to describe massless fermions. A Dirac spinor Ψ_D can be written in terms of two Weyl spinors ξ and χ as

$$\Psi_D = \begin{pmatrix} \xi \\ \chi \end{pmatrix}, \tag{2.37}$$

and using the left- and right-handed projection operators

$$P_L = \frac{1-\gamma_5}{2} \qquad P_R = \frac{1+\gamma_5}{2}, \tag{2.38}$$

the upper and lower, or left- and right-handed, components of the Dirac spinor can be projected out, as

$$P_L\Psi_D \equiv \Psi_L = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \qquad P_R\Psi_D \equiv \Psi_R = \begin{pmatrix} 0 \\ \chi \end{pmatrix}. \tag{2.39}$$

Acting on the Dirac equation with either projection operator yields

$$\begin{aligned} \left(i\partial\!\!\!/ - m\right)\Psi_L &= m\Psi_R \\ \left(i\partial\!\!\!/ - m\right)\Psi_R &= m\Psi_L \end{aligned} \tag{2.40}$$

indicating that the mass term mixes left- and right-handed states, while the equations decouple in the massless case.

The concept of "handedness" is more formally known as helicity, and defined as the sign of the projection of the spin vector onto the momentum vector, and by convention, left is negative while right is positive. Since the direction of propagation depends on the frame of reference, and may be changed by a Lorentz boost, helicity is not an intrinsic property of a massive particle. This is contrary to the case of massless particles, which propagate at the speed of light and for which helicity and chirality are the same. Chirality is a more abstract concept with a group-theoretical description, and the easiest way to make immediate sense of it is perhaps by realising that it is merely the eigenvalue of a spinor when acted upon with γ_5, and as such also invariant under Lorentz transformations, since $[\gamma_5, \sigma_{\mu\nu}]$.

As an explanation for the continuous energy spectrum of electrons coming out of beta decay, Pauli proposed the neutrinos at a conference in 1930 with a letter, now named after its opening "Liebe Radioaktive Damen und Herren" [17]. The neutrinos were at that time believed to be massless, and hence described by Weyl spinors, in addition to their zero electric charge which could imply that they are indeed their own antiparticles. The description of such fermions was first presented by Majorana in the 1937 paper "Teoria simmetrica dell'elettrone e del positrone" [18], but left with little attention until the early 1960s, when the discussion of whether neutrinos were in fact Majorana or Dirac fermions opened up.

Majorana fermions are their own anti-particles, and must thus fulfil

$$\Psi_M = \Psi_M^*. \tag{2.41}$$

These spinors are solutions to the Dirac equation, which becomes real in a basis where all the γ matrices are purely imaginary, known as the Majorana basis.

Fermi-Dirac statistics

Fermi-Dirac statistics describe the energy distribution of identical fermions with negligible mutual interaction, in a state of thermodynamic equilibrium. The Pauli exclusion principle states that no two fermions can occupy the same state,

which of course has a significant effect on the properties of systems consisting of fermions. The average number of particles n_i in a single-particle state i is then

$$\bar{n}_i = \frac{g_i}{e^{\frac{E_i-\mu}{k_B T}} + 1} \tag{2.42}$$

where again E_i is the energy of the i'th state, μ the chemical potential, g_i the degeneracy of energy level i, k the Boltzmann constant and T the temperature. Compared to Bose-Einstein statistics, see eq. (2.5), the two distributions look very similar. However, the plus sign in the denominator of the Fermi-Dirac distribution keeps the denominator from ever becoming smaller than unity, and hence the whole expression bound below g_i. This reflects the aforementioned restriction, that two fermions cannot occupy the same state, unless they differ in another quantum number.

2.4.2 Those spinors

The Lorentz group is the set of rotations and boosts which preserve the Minkowski metric; $\Lambda^T g \Lambda$, where Λ are the generators of the group, so spacetime transforms as

$$X_\mu \to \Lambda_{\mu\nu} X_\nu \,. \tag{2.43}$$

The Lorentz group consists of three-dimensional rotations, generated by J_i, and spatial boosts, generated by K_i, so any group element can be uniquely expressed as

$$\Lambda = \exp(i\theta_i J_i + i\beta_i K_i) \,. \tag{2.44}$$

The algebra of the rotation generators is

$$\left[J_i, J_j\right] = i\varepsilon_{ijk} J_k \,, \tag{2.45}$$

and since the rotation group $SO(3)$ is a subgroup of the Lorentz group, every representation of the Lorentz group is also a representation of $SO(3)$. Why all this talk about the Lorentz group? Because there is depth to be found.

Imagining the spin of a fermion as an internal rotation suggests that there is some connection between the spin group and $SO(3)$. The imagined rotation can be projected onto a vector of length $\frac{1}{2}$, and the maximum and minimum values of the projection can be $\pm\frac{1}{2}$, i.e. the spin states along a particular direction can point up or down;

$$\psi \sim \begin{matrix} |+\rangle \\ |-\rangle \end{matrix} . \tag{2.46}$$

There are two representations with $J = \frac{1}{2}$ rotation generators: $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$, and a set of matrices which satisfy the algebra in eq. (2.45) are the familiar Pauli matrices, given in eq. (2.14), which form an irreducible basis for $SU(2)$ The elements in the vector space on which these spin $\frac{1}{2}$ representations act are called spinors, and in particle physics, $(\frac{1}{2}, 0)$ are called left-handed and $(0, \frac{1}{2})$ right-handed Weyl spinors. In terms of the expression in eq. (2.44), right- (left-) handed spinors get a plus (minus) sign in front of the β.

The Standard Model, describing both left- and right-handed fermions, hence contains $SU(2) \oplus SU(2)$, and this Lie algebra combined generates $SL(2, \mathbb{C})$; the group of complex 2×2 matrices with unit determinant. Spinors thus transform as representations of $SL(2, \mathbb{C})$, which is isomorphic[7] to the Lorentz group $SO(1,3)$.

To end this section with a slightly more intuitive picture: Two fermions together form combined states, and the possible combinations of the spins are, expressing now only whether the spin vector points "up" or "down" along a given direction,

$$\begin{aligned} &|+\rangle|+\rangle \,, \quad |-\rangle|-\rangle \,, \quad \frac{1}{\sqrt{2}}\left(|+\rangle|-\rangle + |-\rangle|+\rangle\right) \,, \quad \text{Symmetric} \\ &\qquad\qquad \frac{1}{\sqrt{2}}\left(|+\rangle|-\rangle - |-\rangle|+\rangle\right) \,, \quad \text{Antisymmetric} \end{aligned} \tag{2.47}$$

[7] *Isomorphism* stems from the Greeks words *iso*, meaning "equal", and *morphosis*, meaning "to shape". For groups, this is a one-to-one correspondence between the elements of the group which preserves the group operation. Put very bluntly, isomorphic groups contain the same elements, which just look differently.

i.e. three different symmetric spin-1 states and one antisymmetric spin-0 state, or, in group theory language,

$$(2,1) \oplus (2,1) = (1,1) \otimes (3,1). \tag{2.48}$$

2.5 A very brief history of the early universe

The prevailing cosmological model for the universe is Big Bang cosmology, which describes the evolution of the universe from approximately $t_P = 10^{-43}$ seconds after the Big Bang, where t_P is the Planck time. It is not known what happened before this time, or even certain that the concept of time made sense. If gravity can indeed be unified with the other fundamental forces, it is likely that it was during this period. Immediately after, the Planck era may have been followed by a Grand Unified Theory, short GUT, era, in which the three fundamental forces were unified into a single force.

These potential eras of unification will have ended as the universe expanded and cooled, and crossed temperatures at which phase transitions took place. As the universe had cooled to around 10^{28} Kelvin, the electroweak and strong forces probably, if they were ever united, separated.

Before approximately $10^{-32}\,\mathrm{s} \approx 10^{10} t_P$ after the Big Bang, the universe presumably entered the inflationary epoch, during which every spatial dimension rapidly expanded by a factor of around 10^{26}. What triggered this violent expansion known as inflation is not known. One potential explanation is that a scalar field, often nicknamed the inflaton, may have settled into its ground state and thereby generated an enormous repulsive force.

Neither is it known why the inflationary epoch came to an end, but whichever mechanism triggered it may have released additional energy — e.g. the decay of the inflaton — triggering an era of reheating. The universe, which due to inflation had been significantly cooled, was during this period again heated up to very high temperatures.

After reheating, the universe continued to expand and cool down until it passed the threshold for its assumed last symmetry breaking; the electroweak symmetry breaking described in section 2.3. The elementary particles became massive, and as the electroweak gauge bosons acquired large masses, the weak force became short-ranged.

After this, the physics of the early universe is better understood. At this point, the universe was filled with a hot quark-gluon plasma, and it continued to cool down and expand, forming after a while the chilly, galaxy-inhabited universe we find ourselves living in.

The events in the early universe have been summarised in a timeline in fig. 2.3, where the earliest events are written in italic to indicate that these are speculative.

Epoch	Event	Time[s]
Big Bang	*Fiat lux.*	0
Planck Era — $T \gtrsim 10^{19}\,\mathrm{GeV}$	*Quantum effects of gravity*	$< 10^{-43}$
GUT Era — $T \gtrsim 10^{16}\,\mathrm{GeV}$	*The strong force separates*	$10^{-43} - 10^{-36}$
Inflation starts — $T \sim 10^{15} - 10^{9}\,\mathrm{GeV}$	*Rapid expansion followed by*	$10^{-36} - 10^{-32}$
T_R	*reheating*	$10^{-32} - 10^{-12}$
$T \sim 100\,\mathrm{GeV}$	*Electroweak symmetry breaking*	10^{-12}
Quark Epoch — $T \sim 100$ MeV	Quark-gluon plasma (LHC reachable)	$10^{-12} - 10^{-6}$
Hadron Epoch — $T \sim 10$ MeV	Quarks form hadrons	$10^{-12} - 10^{-6}$
$T \sim 1$ MeV	Neutrinos decouple	1
BBN — $T \sim 0.1$MeV	Light elements are formed	2 – 20 minutes
Recombination — $T \sim 1$ eV	Universe becomes transparent	$3.8 \cdot 10^{5}$ years
Today — $T \sim 10^{-4}$ eV	Accelerated expansion	$13.8 \cdot 10^{9}$ years

Figure 2.3: Timeline of the universe — in temperature.

2.6 All the things we do not know

2.6.1 Theoretical incommodities

The Standard Model does not predict any of the three charges or gauge couplings, three lepton or six quark masses, three weak mixing angles, the Higgs VEV or the $\mathcal{CP}$-violation phase. All these 19 parameters — and nine more, if neutrino oscillations are to be accommodated — thus have to be determined experimentally. Furthermore, neither the different flavours nor their origins are predicted or explained, and the Standard Model contains no prescription for unifying gravity with the other fundamental forces.

2.6.2 Hierarchy problems

One hierarchy problem in particle physics concerns the surprisingly small mass of the Higgs. Theory predicts, and measurements confirm, that the strength of the interaction between the Higgs field and other other particles is proportional to the mass of the particle in question. After electroweak symmetry breaking, the Higgs field acquires a VEV, and its interaction terms with other particles yield mass terms in the Lagrangian. The puzzle is that since the strength of the interaction with the Higgs field scales with the particle mass, the heavier a particle is, the stronger the interaction becomes. Given that all particle masses, including that of the Higgs, are subject to quantum corrections from loops, heavy virtual particles could yield corrections to the Higgs mass larger than the Higgs mass itself. This is clearly not the case, and the simple yet daunting question is *Why not?*.

In general, a hierarchy problem in physics arises when the fundamental value of a physical observable in the Lagrangian differs strongly from the measured value. When this happens, it appears as if there has been a precise cancellation between quantum corrections to the fundamental quantity, often referred to as fine tuning. The problem with fine tuning touches upon the notion of naturalness in physics, where, as formulated by 't Hooft [19], "a small parameter is natural only if a symmetry is gained as it is set to zero". In quantum field theories, if a bare parameter is set to zero, radiative corrections lead to a renormalized non-zero value. Small renormalized values without a symmetry protecting them can thus require fine-tuning to a potentially much larger scale than the values themselves.

This insight demonstrates the severity of the hierarchy problem: If the Higgs mass has not been fine tuned, then what protects it?

The above is only one incarnation of particle physics hierarchy concerns. In general, the large difference in strength between the fundamental forces, or between the masses of the fundamental particles, all seem to owe an explanation. To illustrate: Scaling the electron to a mass of 1 kilogram corresponds to scaling the top quark to a mass of 340.308 kilograms. *Why do the building blocks of the universe have so immensely different scales?*

2.6.3 Dark topics

Observations of the rotation curves and velocities of galaxies both within and outside clusters have led to the conclusion that they must necessarily consist of more matter than the currently visible, since they would otherwise be torn apart. Without going into modified theories of gravity, one immediate solution could be to imagine more than the visible matter. Such invisible matter has later been calculated to constitute approximately 85% of the matter in the universe, and been dubbed dark matter, as it does not radiate. The Standard Model does not have a dark matter particle candidate.

Moving on to the next dark topic, the energy content of the universe has been mapped [20], and it is today distributed as shown in the chart in fig. 2.4. Even taking dark matter into account, a substantial part of the universe's energy content is not made up of matter, but another form of energy, about whose nature little is know. Its detection is limited to large-scale observations of a gravitational repulsion effect and the relation between redshift and distance to cosmological objects. It appears this unknown form of energy, called dark energy, is associated with empty space itself, and evenly distributed throughout the dimensions. Effectively, it contributes as a cosmological constant would, representing an inherent vacuum energy in spacetime. The Standard Model has no candidate for describing dark energy either.

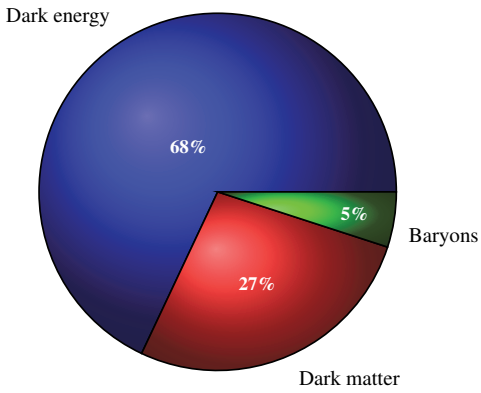


Figure 2.4: Present distribution of the energy contents of the universe.

2.6.4 The matter-antimatter asymmetry

This problem refers to the observed imbalance between baryons and anti-baryons in the universe. Under the assumption that matter and antimatter were produced in equal amounts during the Big Bang, some mechanism must have caused these two amounts to diverge at a later time. The cosmological matter-antimatter asymmetry expressed in terms of the ratio of the number density of baryons to photons, is[8]

$$\eta_B = \frac{n_B - n_{\bar{B}}}{n_\gamma} \simeq \frac{n_B}{n_\gamma} = 6.19 \pm 0.15 \times 10^{-10}, \tag{2.49}$$

again, an extremely small and seemingly fine-tuned number for which the Standard Model has no explanation.

2.6.5 Neutrino masses

According to the Standard Model, the three known neutrinos are all massless. However, oscillations between different neutrino flavours have been observed [22], which means that their interaction and mass eigenstates differ, which can in turn only be the case if neutrinos do indeed have a rest mass, implying also that right-handed neutrinos exist, and escape detection by virtue of being complete gauge singlets. Neutrino masses could be accommodated within the Standard Model, so this issue appears perhaps a bit less overwhelming than the shortcomings listed previously. However, the question regarding a possible Majorana nature [23], for which neutrinos are the only candidates in the Standard Model, and would be confirmed if neutrinoless double beta decay was observed, remains open. So does the question regarding a possible seesaw [24, 25, 26, 27, 28] mechanism which would explain the incredibly small neutrino masses, but in turn also introduce hitherto unobserved, extremely heavy additional neutrinos.

Whatever ends up resolving this issue could thus potentially bring about much more than merely a minor adaption of the theory. In short, the Standard Model does not yet know.

[8] This number has been most precisely determined via measurements of the angular distribution temperature fluctuations in the CMB [21].

Figure 2.5: Image from https://www.jpl.nasa.gov, used in NASA's article "Machines Teach Astronomers About Stars" [29].

2.7 . . . and very good reasons to be overwhelmed

There are around $250 \cdot 10^9$ stars in the Milky Way and another $\sim 10^{12}$ in our neighboring Andromeday galaxy. Together with another fifty galaxies, we form the local group, which is again part of the local supercluster, called the Virgo Supercluster. This is one of about $10 \cdot 10^6$ superclusters in the observable universe.

During Run 2 of the LHC, there were around $600 \cdot 10^6$ events per second, and the dataflow from all four experiments — Alice, ATLAS, CMS, LHCb — around 25 GB every second.

The sheer amount of data collected at particle accelerators, by space observatories and telescopes is overwhelming at the least, and an increasing challenge for physicists is to deal with the large amounts of data, and extract the information therein. It is not even unthinkable that data containing undiscovered physics have already been collected. The field of data science has become a natural and important part of many areas of physics, particle physics being among the forerunners. In many ways, data science could be considered a "fourth paradigm" of science, alongside empirical, theoretical and computational.

In the following work, different beyond Standard Model physics theories are considered. Parameter scans are used to identify valid regions in the parameter space of superymmetric theories, possible signals from dark matter annihilation are simulated and its possible detection discussed, and statistical methods including machine learning are used to analyse collider signatures and extract discriminatory information. All the projects in this work rely on programming and statistical methods, which are explained throughout. The code produced is available as indicated in the list of publications.

Chapter 3

Statistics and scans

In particle physics, observations and experimental outcomes are determined by a theory which only Nature knows. This constitutes a probability distribution — which may or may not be possible to analytically describe — and a collection of data is a random sample drawn from this underlying distribution.

Particle physics consists of mathematical models which describe physical processes, and the task of a particle physicists is to build and then adapt or discard these models after evaluating them in light of observed data. An important aspect of such an evaluation is to make quantitative statements about the parameters of a model. However, in order to make such statements, the connection between the observations and the model parameters must either be understood or assumed. The term model can in this context mean a theory described by a Lagrangian, or a specific point in parameter space. Finding out whether the predictions of a model are consistent with data is called hypothesis testing. On the other hand, determining the free parameters of a theory assumed to be correct is called parameter estimation.

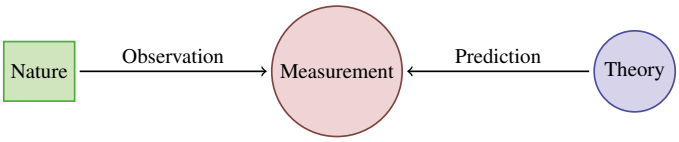


Figure 3.1: Nature follows some hidden rules which give rise to outcomes that can be observed. The aim is to find a theory which predicts the same outcomes.

Given a theory and its parameters, event generators and spectrum calculators can be used to simulate experimental outcomes or calculate mass spectra, i.e. predict which measurements it would give rise to. This is called deductive reasoning, and amounts to making predictions from a model. A model with fixed parameters constitutes a premise P, which yields a set of possible observations. Given a model, the possible outcomes O are specified.

Given a set of observations, on the other hand, it is not given which model produced those data. Several models, or even the same model with different parameters, can produce similar observations. Trying to 'go backwards' from outcome to premise is called inductive reasoning. An outcome O does not necessarily say much about which P it stems from, since there could have been $P_1 \to O$, $P_2 \to O$, $P_3 \to O$, and so forth. Were it possible to know all possible models, and only a certain P could produce the observed O, then P would for certain be true. In any other case, P can not be verified. The only thing which *is* possible, is to falsify a given P.

The following chapter contains definitions, rules, a selection of theorems and some hopefully helpful explanations. The parameter scans used in chapter 7 and *Trilinear-augmented gaugino mediation* [1] are explained, and the statistical concepts and techniques needed for chapter 5 and *Signal mixture estimation for degenerate heavy Higgses using a deep neural network* [2] are reviewed.

3.1 Probability and interpretation

Before beginning the statistics discussion, interpretations of probability from two different schools are reviewed. It will later become apparent, e.g. in likelihood guided parameter scans or when interpreting classifier output, that these different interpretations can potentially lead to different conclusions.

3.1.1 The frequentist

The frequentist, and perhaps the most intuitive, way of interpreting probability is by defining the probability of an event as the limit of its relative frequency in a large number N of trials, i.e.

$$P(event) = \lim_{N\to\infty} \frac{n_{\text{event}}}{N}. \tag{3.1}$$

In short, frequentists draw samples from a distribution, and try to make conclusions about its shape. Applying this interpretation to experimental outcomes amounts to assuming that each experiment produces statistically independent results, and relies on the possibility of repeating said experiment a large number of times.

Confidence

As a concrete example, assume a model with a single free parameter θ. The probability density function, or pdf for short, of the continuous observable x is usually denoted

$$p(x;\theta), \tag{3.2}$$

and its integral over an interval expresses the probability of x being contained in that interval. In the case of a discrete observable, the proper nomer for the above function is probability mass function, but the abbreviation pdf is often used in both cases.

Assuming that the model under consideration is correct, observations of x can help determine the true value of the unknown parameter θ. Repeated experimental outcomes are used to create confidence intervals $(\theta_{\text{min}}, \theta_{\text{max}})$ as estimates for the true value of θ. Since the observations of x are random samples from an underlying distribution, the confidence interval is also random, and the true value of θ may be contained in the interval, but it might also not.

Constructing many confidence intervals, the proportion containing the true value of θ is known as the confidence level. Constructing an interval $(\theta_{\text{min}}, \theta_{\text{max}})$ at e.g. 95% confidence level does *not* mean that the probability for finding the true value of θ in said interval is 95%. Rather, it means that had the experiment been repeated a large number of times, and the confidence interval been constructed in the same way each time, the true value of θ would be contained in 95% of the intervals $(\theta_{\text{min}}, \theta_{\text{max}})$. Under a hypothesis test, the level of significance is the complement of the confidence level; a 95% confidence interval corresponds to a significance level of 0.05. Note that the desired level of confidence is set by the researcher, not determined by data.

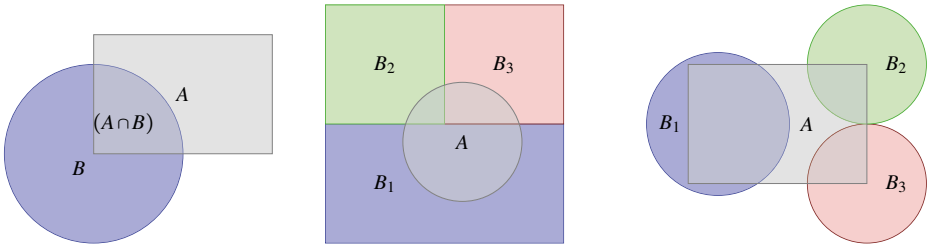
3.1.2 The Bayesian

Abandoning the requirement of basing the concept of probability on repeatable experiments, opens up the possibility of interpreting probability as a *degree of belief*, rather than just the limit of a number.

Bayes' theorem

Given two sets of outcomes A and B which are partially overlapping, see fig. 3.2a, define

- $P(A \cap B)$, the joint probability for both A and B
- $P(A|B)$, the conditional probability of A, given B.



(a) (b) (c)

Figure 3.2: Completely and partially overlapping probabilities.

The probability of drawing from the joint set $(A \cap B)$ is

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)\,. \tag{3.3}$$

Using only algebra, this is readily rearranged into

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{3.4}$$

which is Bayes' theorem, and in this form beyond mathematical dispute. Applied to point probabilities, the theorem is a result of the total law of probabilities, and is of course valid regardless of the interpretation of probability. However, a slight change in notation and interpretation of the above probabilities, yields a statement which is only valid in the Bayesian interpretation of probability:

$$p(\Theta|\mathcal{D},\mathcal{M}) = \frac{p(\mathcal{D}|\Theta,\mathcal{M})p(\Theta|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})} \equiv \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}}\,. \tag{3.5}$$

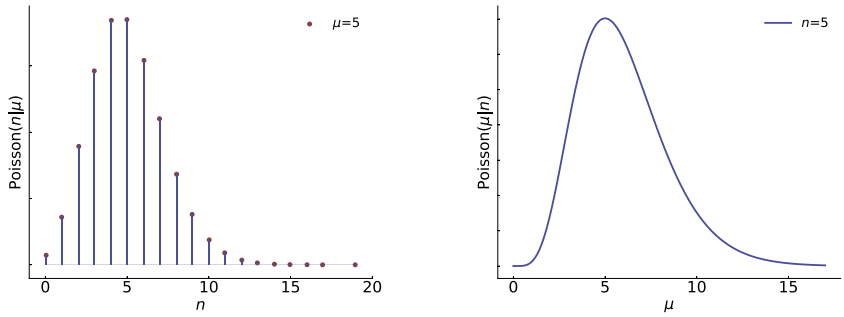
Here, Θ represents a set of free parameters, $\mathcal{D}$ a collected data sample, and $\mathcal{M}$ a model assumed to be correct, i.e. a sort of background information. In the Bayesian interpretation,

- $\pi(\Theta) \equiv p(\Theta|\mathcal{M})$ is known as the prior probability of Θ and expresses the degree of belief in Θ before taking data into account.
- $\mathcal{L}(\Theta) \equiv p(\mathcal{D}|\Theta,\mathcal{M})$ represents the probability of obtaining the given data, assuming that Θ is true. Interpreted as a function of Θ, it is the likelihood $\mathcal{L}(\Theta)$, see section 3.2.1.
- $\mathcal{Z} \equiv p(\mathcal{D}|\mathcal{M})$ normalizes the posterior over the domain of the model parameters,

  $$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta\,. \tag{3.6}$$

  It becomes large only if the likelihood is relatively large in a region of parameter space where the prior probability is also large. Specifically, given sets of parameters with comparable agreement with data, i.e. similar values for the likelihood, those whose priors extend over a large proportion of parameter space, implying smaller values for $\pi(\Theta)d\Theta$, have weaker evidence than those with narrow priors — a formalisation of Occam's razor.
- $p(\Theta|\mathcal{D},\mathcal{M})$ is the posterior probability of Θ, and represents the updated degree of belief in Θ after comparison to data.

Bayes' theorem thus provides an updated belief in light of data, based partially on the prior belief in a model. This perspective implies a subjective view of probability: Probability represents belief, which contrasts with the frequentist perspective. From the Bayesian perspective, any quantity for which the true value is uncertain, including model parameters, can be represented via probability distributions. This does not make sense from a frequentist perspective, where parameters are fixed quantities and only the data random.



(a) (b)

Figure 3.3: Two sides of the Poisson distribution: (a) The distribution of n, resulting from drawing ten thousand i.i.d. samples from the Poisson distribution with $\mu = 5$, and (b) the function of μ obtained from the Poisson distribution with $n = 5$.

3.2 Likelihood

Consider for illustration a distribution highly relevant for event counting experiments, the Poisson distribution

$$\text{Poisson}(n;\mu) = \frac{\mu^n}{n!}e^{-\mu}\,. \tag{3.7}$$

This describes the probability of measuring n outcomes when the mean expected number of outcomes is μ; the number n is discrete while μ is continuous. This is thus a two-dimensional function from which several probability density functions can be obtained. For instance, assuming $\mu = 5$ outcomes yields a pdf[1] for the outcome of n, shown in fig. 3.3a. On the other hand, assuming an outcome of for example $n = 5$, the distribution of the expected number μ looks as shown in fig. 3.3b. These two are independent and furthermore completely different functions; the last one is not even a probability density function — as can be easily verified since the area under the curve does not integrate to unity. It is however a very useful function, and often referred to as the likelihood $\mathcal{L}(\mu|n)$. It reaches its maximum at the value for μ which would have given the largest probability of the observed, or in this case fixed, n.

To generalise, drawing a dataset $\mathcal{D}$ of N samples from a Poisson distribution amounts to drawing the first N terms of an independent and identically distributed — i.i.d — sequence $\{x_1,\ldots,x_i,\ldots,x_N\}$ of Poisson variables. Independent measurements have the property that their combined probability is the product of each probability, so the total likelihood is

$$\begin{aligned}\mathcal{L}(\mu|\mathcal{D}) &= \prod_{i=1}^{N}\text{Poisson}(\mu\,|\,x_i)\\ &= \prod_{i=1}^{N}\frac{\mu^{x_i}}{x_i!}e^{-\mu}\,.\end{aligned} \tag{3.8}$$

Technical side note: Logarithms for computations

Equation (3.8) is a product of probabilities. A probability is a number between zero and one, so multiplying several of them inevitably results in a number close to zero. This is a challenge for computers, which use floating point arithmetics

[1] This is, as mentioned earlier, strictly speaking a probability mass function, since n is discrete.

to represent real numbers, and are thus sensitive to numbers very close to zero. The consequence of this is for the present case a potentially fragile computation, depending on how many probabilities are included in the product.

Furthermore, and as will become clear in the following, the aim is to maximise (minimise) the (negative) likelihood. This is an optimisation task, which involves calculating the value of a function along with its derivative. Optimising a product requires all the data to be loaded into memory at once, followed by calculation of the product and partial derivatives, before the next optimisation step can be performed. Conversely, the derivatives of a sum is the sum of the derivatives. Optimising a sum therefore requires only one datum to be loaded into memory and its partial derivative can be computed independently. The gradients can then be accumulated and the optimisation step applied, with significantly smaller memory consumption.

To circumvent these two challenges, it is common to work with the logarithm of the likelihood. This is safe since the logarithm is a monotonic transformation and preserves the locations of the extrema[2]. The convenience of this can be illustrated by trying to numerically find the extrema of two functions, where one is the logarithm of the other:

$$\begin{aligned} p(x) &= e^{-x^2} \\ q(x) &= \log p(x) = -x^2 \,. \end{aligned} \tag{3.9}$$

These both have global optima at the origin, and their gradients are

$$\begin{aligned} p'(x) &= -2xe^{-x^2} \\ q'(x) &= -2x \,. \end{aligned} \tag{3.10}$$

Consider first $q'(x)$: Multiplying this by $\frac{1}{2}$ yields the step size needed to reach the global optimum at $x = 0$, regardless of the value for x. Moving on to $p'(x)$, this contains the multiplication of e^{-x^2}, which decays faster than exponentially with x. At for example $x = 5$, the gradient is $p'(5) = 1.38879 \cdot 10^{-10}$. This would have to be multiplied by 10^{10} to reach a reasonable step size towards the optimum, making this option worse than useless; it would be better to just take a unit step in the positive direction.

This example might seem contrived at first, but it holds for most of the commonly encountered distributions. For instance, given a Gaussian distribution, see eq. (3.15), maximizing the likelihood reduces to a problem of least-squares.

3.2.1 The Log-Likelihood

The logarithm of the likelihood for the Poisson distribution is

$$\log \mathcal{L}(\mu) = \log \mu \sum_{i=1}^{N} x_i - N\mu - \sum_{i=1}^{N} \log(x_i!) \,. \tag{3.11}$$

Since the logarithm is a monotonic transformation which preserves the locations of the extrema, the value of μ which maximises the likelihood given a set of data, denoted $\hat{\mu}$, is also the one which maximises the loglikelihood. This value is therefore referred to as the estimator for the true value, and defined as

$$\hat{\mu} = \arg\max_{\mu} \, \log \mathcal{L}(\mu | x_1, \ldots, x_N) \,. \tag{3.12}$$

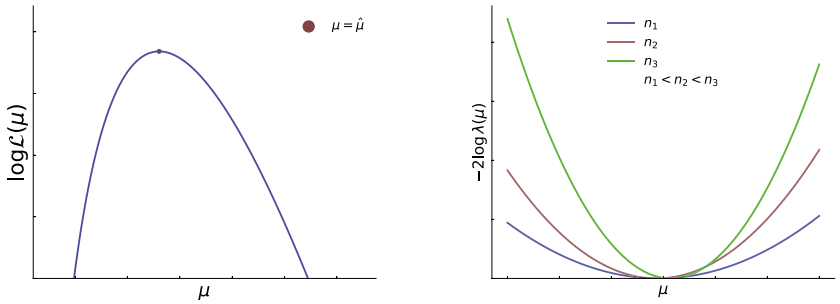
For the Poisson distribution, this is

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N} \,, \tag{3.13}$$

i.e. the sample mean of the N observations in the sequence. This makes intuitive sense because the sample mean is an unbiased estimator of the expected value of μ; the expectation value of a Poisson random variable.

The loglikelihood for a hundred draws from a Poisson distribution with parameter μ_{true} is shown in fig. 3.4a, where the maximum of the curve is located at $\hat{\mu} \approx \mu_{\text{true}}$. The location of the maximum in the x-direction does not change if the

[2] In particular, estimated best-fit parameters corresponding to the maximum of the likelihood are the same for the likelihood and the logarithm.



(a) (b)

Figure 3.4: (a) The loglikelihood curve resulting from drawing one hundred i.i.d. samples from a Poisson distribution. (b) Loglikelihood ratio curves for 500, 1000 and 2000 i.i.d. draws from a Gaussian distribution. The curves become narrower, i.e. the estimator $\hat{\mu}$ more certain, with increasing data.

loglikelihood of the maximum estimator $\hat{\mu}$ is subtracted, but the whole curve is shifted in the y-direction, such that the maximum is located at 0.

The loglikelihood with its maximum subtracted corresponds of course to the ratio $\log \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$, often referred to as the loglikelihood ratio. In the following, the notation

$$\lambda(\mu) \equiv \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})} \qquad 0 \leq \lambda(\mu) \leq 1 \tag{3.14}$$

is used. Larger (smaller) $\lambda(\mu)$ means better (worse) agreement between data and the hypothesised μ, and this quantity is easier to work with because its optimum is always located at 0. It is, again for computational reasons, more common to minimise the negative loglikelihood, in which case the minimum is still located at 0.

Before leaving the likelihood discussion, consider another distribution, this time the Gaussian, or normal, distribution

$$G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3.15}$$

where μ is the expectation value and σ the standard deviation. After drawing N i.i.d. samples $\{x_1, \ldots, x_N\}$ from such a distribution, the loglikelihood is

$$\log \mathcal{L}(\mu) = \log \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma_i^2}, \tag{3.16}$$

and all terms not dependent on μ will again cancel in the loglikelihood ratio. The above function is a parabola with extremum at some value $\mu = \hat{\mu}$.

The second derivative of a parabola is a constant, define here

$$h \equiv -\frac{\partial^2}{\partial \mu^2} \log \mathcal{L} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}, \tag{3.17}$$

to rewrite the entire expression as

$$\log \mathcal{L}(\mu) = \log \mathcal{L}(\hat{\mu}) - \frac{h}{2}(\mu - \hat{\mu})^2. \tag{3.18}$$

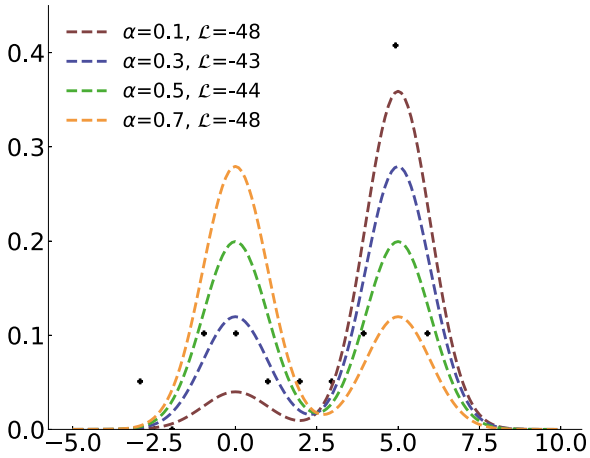


Figure 3.5: Different trial values of the mixture parameter α along with the resulting likelihood for 20 independent draws from the distribution in eq. (3.23) with mixture parameter $\alpha_{\text{true}} = 0.3$.

This corresponds to

$$\mathcal{L}(\mu) = \mathcal{L}(\hat{\mu}) e^{-\frac{h}{2}(\mu-\hat{\mu})^2}, \tag{3.19}$$

which represents a Gaussian distribution with mean $\hat{\mu}$ and standard deviation $h^{-1/2}$. Since h increases with the total number of data points, a larger data sample causes the width of the loglikelihood to decrease, see fig. 3.4b.

Expressing the loglikelihood in this way reveals one of its very handy properties: The uncertainty $\sigma_{\hat{\mu}}$ in the estimated $\hat{\mu}$ can be readily read off a loglikelihood plot: Evaluating eq. (3.18) at $\mu = \hat{\mu} \pm \sigma_{\hat{\mu}}$ yields

$$\log \mathcal{L}(\hat{\mu} \pm \sigma_{\hat{\mu}}) = \log \mathcal{L}(\hat{\mu}) - \frac{1}{2}, \tag{3.20}$$

i.e., the $1\sigma_{\hat{\mu}}$ interval corresponds to the point where the loglikelihood drops by $\frac{1}{2}$ from its extremum value,

$$-2 \log \lambda(\hat{\mu} \pm \sigma_{\hat{\mu}}) = 1. \tag{3.21}$$

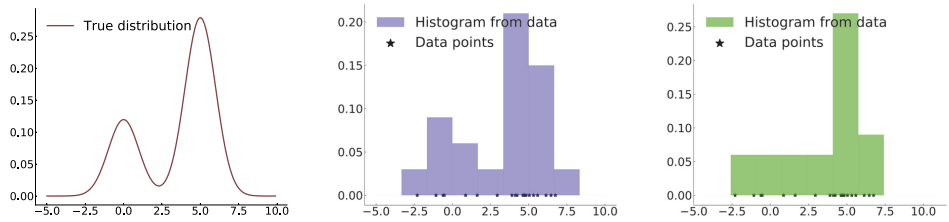
This equation can be generalised for $s \cdot \sigma_{\hat{\mu}}$ confidence intervals as [30]

$$-2 \log \lambda(\hat{\mu} \pm s\sigma_{\hat{\mu}}) = s^2. \tag{3.22}$$

3.3 Likelihood and data

3.3.1 Maximum likelihood fits

Provided a collection of data, these can as announced earlier be used to fit the parameters of a model. Consider for illustration a mixture model consisting of two Gaussians, with known means of zero and five respectively, standard



(a) (b) (c)

Figure 3.6: The mixture model in eq. (3.23) shown as (a) the true distribution, and (b) and (c) histograms resulting from 20 independent draws from the distribution, but with different bin locations.

deviation of one, but with unknown mixture parameter. The total pdf for this model is

$$p(x|\alpha) = \alpha G(x,0,1) + (1-\alpha)G(x,5,1)\,, \tag{3.23}$$

where α can take on a value in the interval $(0,1)$.[3]

As a concrete case, let the true but unknown value for the mixture parameter be $\alpha_{\text{true}} = 0.3$. While this value is unknown, random draws generated by the true model can be used to estimate the mixture parameter $\hat{\alpha}$. To demonstrate the procedure of maximum likelihood estimation, twenty random draws are used to calculate the likelihoods assuming four different trial values for α. These four trial mixture models, the draws from the true model and the resulting likelihood values are shown in fig. 3.5. Since the trial model with $\alpha = 0.3$ achieves the highest likelihood for generating the twenty random draws, the best-fit estimate is $\hat{\alpha} = 0.3$.

In practice, the underlying pdfs of the mixture model may not be so easily separable, the data more scattered, and one would use minimization algorithms to find the best estimate $\hat{\alpha}$. The results in *Signal mixture estimation for degenerate heavy Higgses using a deep neural network* [2] are obtained performing maximum likelihood fits. See also section 5.2.1 for further details.

3.3.2 Density estimation

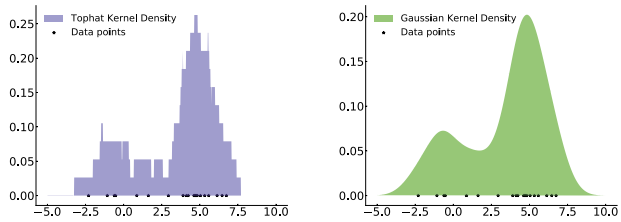
Using draws to represent and construct an estimate of a distribution is called density estimation, and a common choice for graphically representing a distribution of data is via histograms. The term 'histogram' was first used by Karl Pearson during his lectures on statistics, some time before 1894 [31], as

> "a term for a common form of graphical representation, i.e. by columns marking as areas the frequency corresponding to the range of their base.",

and everyone who has spent a little time playing around with bin widths and locations, knows that a histogram is not merely a histogram; it can be anything from an excellent representation of data to something completely confusing and misleading.

Consider again the mixture model from eq. (3.23): This pdf with $\alpha = 0.3$ is shown in fig. 3.6a, along with the two histograms in figs. 3.6b and 3.6c created from the same 20 independently drawn samples from the pdf. The difference between the two is merely the bin edges; the number and sizes of the bins are the same. An unlucky scientist who happened to choose the bin locations corresponding to fig. 3.6c would probably end up believing that the true model is asymmetric with only one peak and a large tail in one direction. Apparently, the arbitrary choice of binning can lead to fundamentally different interpretations of the data. This example serves to show how histograms are not smooth and

[3] α can of course also take on the values 0 and 1, but then the model seizes to be a mixture model.



(a) (b)

Figure 3.7: Two kernel density estimators: (a) with a tophat kernel and (b) with a gaussian kernel, both with $h = 0.95$.

depend strongly on the chosen bin widths and edges. Intuitively, histograms consist of blocks which each represent a data point, stacked in a grid, and it is this gridding which causes the problem.

Attempting a different form of density estimation, define the estimate $\hat{f}$ of the unknown underlying function f from which the data are drawn, at one single point x as

$$\hat{f}(x) = \sum_i K\left(\frac{x - x_i}{h}\right). \tag{3.24}$$

Here, K represents the kernel function, and h the bandwidth. The bandwidth determines the width around a point where the kernels overlap. A low bandwidth implies only points very close to a position are given weight, while a high bandwidth implies contribution from distant points.

This method is called Kernel Density Estimation, and the difference from histograms is that each block is now centered at the point it represents and the shape of the block is determined by the kernel function. Two examples are shown in fig. 3.7, one where the kernel is still just a block, often called a 'tophat' kernel, and one with a Gaussian kernel.

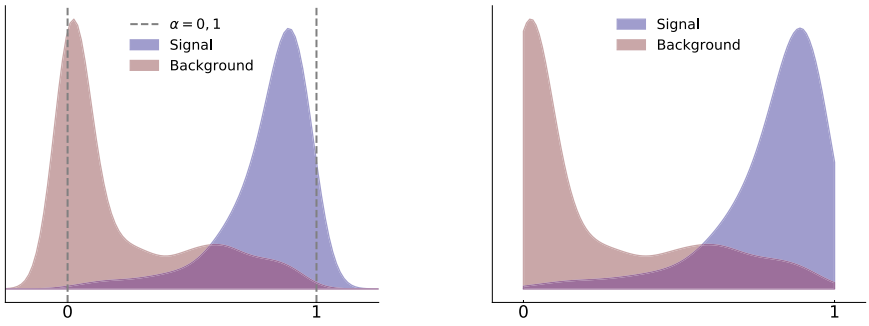
The Gaussian kernel function is commonly used in physics, and since the kernel function is continuous, so is the entire density estimate. This is also the kernel used in section 5.3 and [2].

Just as data representation via histograms depend on the choice of bins, kernel density estimation depends on the bandwidth: A too small value of h can lead to a very spikey estimate, while a too large h can cause the estimate to be over-smoothened — part of the histogram problem translated. Ideally, one wants to choose the smallest h allowed by the data.

Several methods for estimating the best bandwidth are of course available, but as expressed by Silverman [32]:

> The optimal width is the width that would minimise the mean integrated squared error if the data were Gaussian and a Gaussian kernel were used, so it is not optimal in any global sense. In fact, for multimodal and highly skewed densities, this width is usually too wide and over-smooths the density.

Since the standard mean squared error approach is problematic for distributions with several modes, and in the present work, bimodal distributions are the interesting ones, a more fruitful approach is cross-validation, which will also be important in the discussion on training machine learning models. This amounts to dividing the dataset at hand in two, performing the density estimation on one half, and via a maximum likelihood fit determining how well this density estimate fits the second half of the dataset. Maximising the likelihood for different bandwidths then yields the best-fit value for h.



(a) (b)

Figure 3.8: Kernel density estimation of signal and background templates, using a Gaussian kernel. The templates consist of points restricted to the interval $[0, 1]$. The KDEs in (a) are shown on the interval $[-0.25, 1.25]$ to illustrate how the Gaussian curves representing the points close to the edges at 0 and 1 extend beyond the interval spanned by the templates they represent. The KDEs in (b), have been cut off at the border and renormalised within the range $[0, 1]$.

3.3.3 Kernel density estimation on the edge

One potentially challenging scenario for kernel density estimation with a Gaussian kernel, is when the data is constrained to an interval, especially if the density of the data points is high around the edges.

When perfoming a kernel density estimate with a Gaussian kernel function, one Gaussian is created for each data point, before all the Gaussians are combined into the total estimate. The direct consequence of this is that the density estimation function is defined on the entire interval $(-\infty, \infty)$, as all Gaussian distributions are. The bulk of the area of each Gaussian is of course centered closely around its data point, but some of it is comprised by the tail, which only asymptotically vanishes.

This is a problem in cases such as the one described in section 5.3, where one of the tasks is to create a density estimate for the output from a classifier. The classifier outputs numbers in the range $(0, 1)$, so classifying many instances results of course in many numbers between 0 and 1. Also, the better the performance of the classifier, the more of its output is located close to 0 and 1. Points close to the edge on either side are represented by Gaussians whose area covers a non-negligible interval outside the defined range. This again causes the total area over the range to be smaller than unity, *and* the density estimation around the edges to be too small to accurately represent the density of the data there: The most important points are under-represented in the estimate!

There are of course ways around this. For instance, the workaround implemented in the particle physics package `ROOT` [33], where this problem is referred to as "spill out boundary effect", is to mirror the data at the edges, and project it back into the allowed interval. This strategy is, however, only useful for Gaussians with peak, i.e. zero derivative, exactly on the edge.

Another solution is to simply cut off the Gaussian at the edge of the chosen range, and renormalise[4] it so that its integral remains the same, but now bounded to the defined range, as illustrated in fig. 3.8. The curve is strictly speaking not a Gaussian anymore, but only because it has been spliced — it can still be described using an analytical function. This method has been implemented for the project on *Enabling sneutrino detection in weak signal scenarios using machine learning methods* [6], described in section 5.3, and the code is available on [34].

[4] This time in the literal sense of the word.

3.4 Hypothesis testing

The following section is strictly speaking mostly a series of nice occasions to discuss useful theorems and lemmas, and to explain terminology. This is done by introducing the necessary ingredients for hypothesis testing, a procedure which can be divided into the following steps

1. The null hypothesis $\mathcal{H}_0$ and alternative hypothesis $\mathcal{H}_1$ are formulated
2. A test statistic which can can be used to assess the truth of $\mathcal{H}_0$ is defined. The test statistic is a function that maps observed data into a single real rumber.
3. The p-value is computed. A smaller p-value means stronger evidence against $\mathcal{H}_0$.
4. The p-value is compared to a predetermined significance value α to decide whether the observation is statistically significant. If $p \leq \alpha$, $\mathcal{H}_0$ is falsified.

The details in these steps are explained in the following.

3.4.1 Wilks' theorem

Assume a set of probability density functions $p(x;\theta)$, depending on a d-dimensional parameter $\theta = (\theta_1, \ldots, \theta_d)$ contained in a parameter space $\mathcal{H}_1$. Define also an r-dimensional subset $\mathcal{H}_0$ of $\mathcal{H}_1$, so that $r < d$. Again, a dataset $\mathcal{D}$ containing N i.i.d. samples $\{x_1, \ldots, x_N\}$ is drawn from a distribution with $\theta \in \mathcal{H}_1$, so the likelihood function is

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} p(\theta|\mathcal{D}) \,. \tag{3.25}$$

The maximum of the likelihood for θ in $\mathcal{H}_1$ is $\mathcal{L}_d^{\max} \equiv \mathcal{L}(\hat{\theta}_d)$. In order to test the hypothesis $\theta \in \mathcal{H}_0$, calculate the maximum of the likelihood for θ in $\mathcal{H}_0$, $\mathcal{L}_r^{\max} \equiv \mathcal{L}(\hat{\theta}_r)$. Since $\mathcal{H}_0 \subset \mathcal{H}_1$, the relation $\mathcal{L}_r^{\max} \leq \mathcal{L}_d^{\max}$ holds, and the likelihood ratio

$$\lambda = \frac{\mathcal{L}_r^{\max}}{\mathcal{L}_d^{\max}} \tag{3.26}$$

is within the range $0 < \lambda \leq 1$. Interpreting the $\mathcal{H}$ as hypotheses, the null hypothesis $\mathcal{H}_0$ should be rejected if λ is small, and accepted if λ is sufficiently close to 1.

A test proposed by S.S.Wilks in 1938 [35] defines a test statistic

$$t_\theta = -2\log\lambda \qquad 0 \leq t_\theta < \infty \,, \tag{3.27}$$

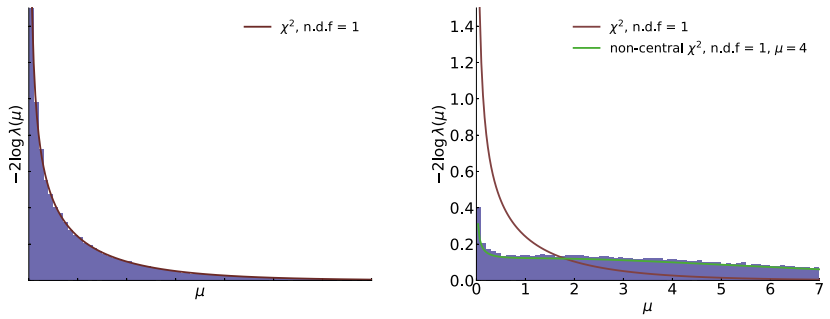
such that if the hypothesis $\mathcal{H}_0$ is true, the distribution of t_θ converges to a χ^2 distribution with $d - r$ degrees of freedom as $N \to \infty$, irrespective of the true θ. The null hypothesis can thus be rejected if t_θ is too large in the χ^2 distribution. Conversely, if the alternative hypothesis holds true, t_θ asympotically follows a non-central χ^2 distribution.

This can be illustrated using for instance a Gaussian distribution. Take as $\mathcal{H}_0$: $\mu = 0$ and $\mathcal{H}_1 : \mu = 4$, and construct the test statistic

$$t_\mu = -2\log\lambda(\mu) = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma_i^2} \,. \tag{3.28}$$

First, this quantity is calculated for ten thousand draws from a Gaussian distribution with $\mu = 0$, shown in fig. 3.9a, where the overlaid χ^2 probability density function with n degrees of freedom is defined as

$$\chi^2(x;n) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0; \\ 0, & \text{otherwise}. \end{cases} \tag{3.29}$$



(a) (b)

Figure 3.9: The distribution of $-2\log\lambda(\mu)$, under $\mathcal{H}_0$: $\mu = 0$ and $\mathcal{H}_1 : \mu = 4$, after drawing ten thousand samples from Gaussian distributions with (a) $\mu = 0$ and (b) $\mu = 4$. Both have a χ^2 distribution with one degree of freedom overlaid, and (b) shows also a non-central χ^2 distribution with sample mean at $\mu = 4$.

Next, it is calculated for another ten thousand draws, this time from a Gaussian distribution with $\mu = 4$, shown in fig. 3.9b. This plot has also the overlaid non-central χ^2 function, defined as

$$\chi^2_{\text{n-c}}(x;n,\mu) = \sum_{i=0}^{\infty} \frac{e^{-\frac{\mu}{2}}\left(\frac{\mu}{2}\right)^i}{i!}\chi^2(x;n+2i)\,. \tag{3.30}$$

3.4.2 The Neyman-Pearson lemma

Consider again a set of hypotheses and define a simple hypothesis test as one where the parameter of the hypothesis uniquely specifies the distribution of the population from which samples are drawn. For instance, given a Gaussian distribution with known variance and unknown mean, $\mathcal{H}_0$: $\mu = 1$ and $\mathcal{H}_1$: $\mu = 2$ is simple since μ is the only free parameter and uniquely defines the distribution[5]. By contrast, a hypothesis on the form $\mathcal{H}_0$: $\sigma < 5$ is not simple, since it does not uniquely define a distribution. Any hypothesis which is not simple is called a composite hypothesis.

Given a data sample $\mathcal{D}$ from a distribution with parameter θ, and a simple hypothesis test where $\mathcal{H}_0$: $\theta = \theta_0$ is tested against the alternative hypothesis $\mathcal{H}_1$: $\theta = \theta_1$, the Neyman-Pearson lemma [36] states that the most powerful test statistic is the likelihood ratio

$$t_\theta = \lambda(\theta_0, \theta_1 \,|\, \mathcal{D}) = \prod_{\mathcal{D}} \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)}\,. \tag{3.31}$$

This test statistic has an acceptance region of size α, such that

$$t_\theta \begin{array}{ll} < k_\alpha & \text{inside acceptance region} \\ \geq k_\alpha & \text{inside critical region} \end{array} \tag{3.32}$$

where k_α is a constant. The contour $\lambda(\theta_0, \theta_1|\mathcal{D}) = k_\alpha$ is thus the boundary of the region inside which the null hypothesis is not rejected. If a measurement x falls within the critical region, $\mathcal{H}_0$ is rejected, and if it falls within the acceptance region, $\mathcal{H}_0$ is not rejected[6]. The quantity α is called the size of the test or the expected significance level, is related to k_α (hence the subscript) and *independent of observations*.

[5] A binomial test is another, simpler example.
[6] Note that "not rejecting" is not the same as accepting.

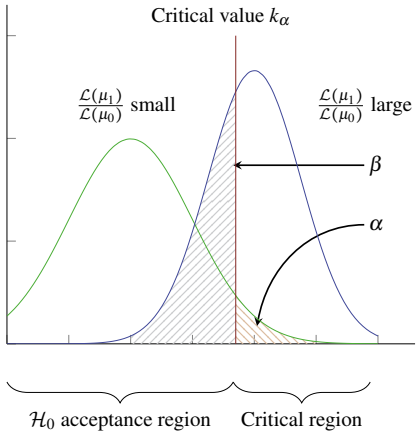


Figure 3.10: Two Gaussian distribution curves with mean values μ_0 and μ_1 respectively, serving to illustrate the acceptance and critical regions as delineated by the critical value k_α.

Denoting the pdf for the test statistic $p(t_\theta)$, α equals the probability for rejecting the null hypothesis when it is in fact true,

$$\int_{k_\alpha}^{\infty} p(t_\theta|\mathcal{H}_0)dx = \alpha\,, \tag{3.33}$$

while the probability for accepting the null hypothesis when it is not true, is

$$\int_{-\infty}^{k_\alpha} p(t_\theta|\mathcal{H}_1)dx = \beta\,, \tag{3.34}$$

and the quantity $1-\beta$ is called the power of the test.

The Neyman-Pearson lemma will not be proven here, but a simple example might provide some intuitive understanding: Assume drawing a single sample from a Gaussian distribution. Applying the Neyman-Pearson lemma to test the simple null hypothesis $\mathcal{H}_0$: $\mu = \mu_0$ against the simple alternative hypothesis $\mathcal{H}_1$: $\mu = \mu_1$, the test statistic

$$\lambda(\mu_0, \mu_1|x) = \frac{\mathcal{L}(\mu_1|x)}{\mathcal{L}(\mu_0|x)} = \frac{G(x;\mu_1,\sigma_1)}{G(x;\mu_0,\sigma_0)} \tag{3.35}$$

takes on different values depending on the value of x,[7]. The two curves and a critical value k_α are indicated in fig. 3.10, where it is obvious that drawing a sample x inside the acceptance region yields a small value for $\lambda(\mu_0, \mu_1|x)$, while the value is larger inside the critical region. The plot in fig. 3.10 also serves to illustrate that a biased test

$$\alpha \geq 1-\beta\,, \tag{3.36}$$

corresponds to a larger probability for not rejecting $\mathcal{H}_0$ when $\mathcal{H}_1$ is true than when $\mathcal{H}_0$ is true.

3.4.3 Significance and p-values

Consider an example from particle physics, looking for a New Physics scenario in an experiment where n events have been collected. Each event can belong to one out of two classes: signal (s) and background (b). The objective

[7]Please, ignore σ for now. The example could have been designed using two distributions of equal and known variance, but this would have rendered the figure less instructive.

is to determine whether one can claim discovery of a new signal, i.e. if the collected data are incompatible with the background-only hypothesis.

The mean expected number of events can be expressed as $\mu = \mu_s s + b$, where μ_s represents the strength of the signal process, and the aim is to test the null hypothesis $\mathcal{H}_0$: $\mu_s = 0$ against the alternative hypothesis $\mathcal{H}_1$: $\mu_s \neq 0$:

	Statistics term	Physics term	Probability function
$\mathcal{H}_0$	Null hypothesis	Background only	$Poisson(n;b)$
$\mathcal{H}_1$	Alternate hypothesis	Signal plus background	$Poisson(n;\mu_s s+b)$

(3.37)

The relevant likelihoods in this scenario are thus $\mathcal{L}_{s+b}$ and $\mathcal{L}_b$, and following the Neyman-Pearson lemma, the best test statistic is the loglikelihood ratio,

$$t = \log \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b} = n \log\left(1 + \frac{s}{b}\right) - s\,, \tag{3.38}$$

where the last equality holds for the case of the Poisson distributions stated in eq. (3.37).

In order to quantify the level of agreement between the data collected in the experiment and a hypothesis, a significance test can be performed. The pdf of the test statistic $p(t)$ is used to calculate the *p*-value of the observed data. This is defined as the probability, assuming a hypothesis $\mathcal{H}_0$, of finding data equally or more incompatible than the value x_{obs} observed in the experiment,

$$\begin{aligned} p_0 &= \int_{x_{obs}}^{\infty} p(t|\mathcal{H}_0)dx \\ p_1 &= \int_{x_{obs}}^{\infty} p(t|\mathcal{H}_1)dx\,. \end{aligned} \tag{3.39}$$

A small *p*-value thus implies incompatibility with the assumed hypothesis, and upon comparison to eq. (3.33), the hypothesis in question should be rejected if p is smaller than α, and the *p*-value is sometimes called the observed significance level.

The *p*-value is closely related to the observed significance, referred to as the *Z*-value, which is perhaps more intuitive, since a measurement has a high significance if it favors rejecting the null hypothesis. A large Z-value corresponds to a small *p*-value, and the two are related via [30]

$$Z = \Phi^{-1}(1-p)\,, \tag{3.40}$$

where Φ is the cumulative distribution function, short cdf, of the unit Gaussian,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x}{2}} dx\,. \tag{3.41}$$

In other words, the *p*-value corresponds to the tail-probability of the Gaussian cdf with $\mu = 0$ and $\sigma = 1$, at a distance Z standard deviations from the mean of the distribution — see fig. 3.11 for an illustration. The Gaussian cdf is closely related to the error function[8], which is more commonly used in statistics, via

$$\Phi(z) = \frac{1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)}{2}\,. \tag{3.42}$$

This can be used to express the *p*-value in terms of the error function as

$$p = \frac{1 - \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)}{2}\,. \tag{3.43}$$

In particle physics, a significance of $Z = 5$ is often required to claim discovery of a new signal. This corresponds to a *p*-value of $2.9 \cdot 10^{-7}$.

[8] See appendix B for a summary of the relevant special functions and their relations.

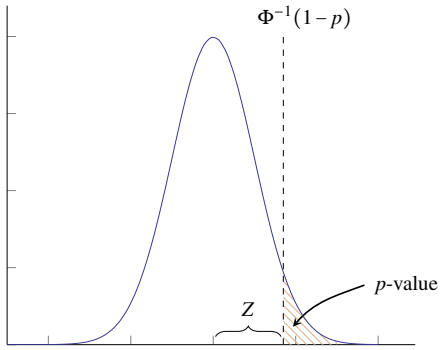


Figure 3.11: Illustration of the relation between the significance Z and the p-value.

3.5 Parameter scans

An important task within particle physics phenomenology is to identify interesting regions in the space of a theory's parameters, and since a physical theory may depend on many parameters, the parameter space can be high-dimensional. This is potentially very challenging, and the present chapter is concluded with a discussion of how to navigate through a high-dimensional parameter space, using the statistical concepts introduced.

3.5.1 Grid scans

A straightforward way of exploring parameter space, is by varying some or all of the parameters which specify a model, over a range with a predetermined step size. This is referred to as a grid scan, and was the strategy applied in [1], where a set of supersymmetric parameters at a high energy scale were varied, and physical quantities such as the identity and mass of the lightest supersymmetric particle, the supersymmetric spectrum and the Higgs mass, were calculated for each parameter point.

Grid scans are a nice and immediately intuitive way of mapping regions and identifying borders in parameter space, for instance where the lightest supersymmetric particle changes identity. Pseudo code for a three-dimensional grid scan over the parameters A, B and C is presented below.

```
for A in (start, stop, step size){
    for B in (start, stop, step size){
        for C in (start, stop, step size){
            do_calculation(A, B, C)
            store_information()
            }
        }
    }
```

The static nature of the predetermined step sizes implies that interesting and uninteresting regions in parameter space are granted the same amount of attention. This is a deterministic way of gathering information, but not an efficient way of exploring a parameter space. Indeed, if the parameter space is very large, it may even be infeasible to entirely cover it this way.

3.5.2 Random scans, or: the curse of dimensionality

If the dimensionality of the parameter space is d, the number of steps along each direction is n and the number of calculations per step, or parameter point, is m, then the total number of calculations in a scan over this space is

$$N = mn^d\,, \tag{3.44}$$

an exponentially increasing number.

In the hope of covering the whole parameter space, one could try sampling points at random. In two-dimensional space, random sampling over an area yields a selection of points which covers the whole area — the intuitive picture. However, as the dimensionality increases, the intuitive picture becomes increasingly useless. This can be illustrated by considering the difference in volumes between a d-dimensional cube and sphere, to illustrate that the majority of the volume is located around the outer edges, and consequently almost none of it around the center.

The volume of a d-dimensional cube with sides a is

$$V_D = a^d\,, \tag{3.45}$$

and volume of a d-dimensional sphere with radius r

$$V_D = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)} r^d\,. \tag{3.46}$$

Scaling the two d-dimensional objects so that the sphere fits exactly inside the cube, $a = 2r$, as illustrated for the two-dimensional case in fig. 3.12a, the difference between the volumes of the two is

$$\Delta V_D = \left(2^d - \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d\,. \tag{3.47}$$

This expression is not immediately enlightening, but plotting the prefactor — the equivalent of considering a unit cube and sphere — for different values of d, as is done in fig. 3.12b, shows that already for quite small numbers of d, the vast majority of the volume is contained outside the sphere. A consequence of this is that a random scan would never sample the region where most of the input parameters are small.

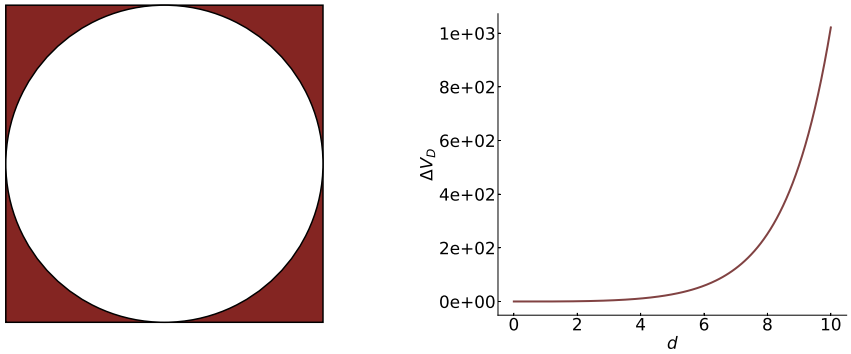
3.5.3 Guided scans

Turning at last to guided scans, the idea is that a scan should be given a range to consider, and then be able to efficiently identify and focus on interesting regions. This can be done by evaluating the likelihood in each point, and choose new points based on this. In this way, unlikely regions in parameter space receive little attention, while the interesting regions are explored with a greater resolution, increasing the chance of identifying the best regions in parameter space, or in statistical terms, identifying the optima of the likelihood function. This means that the likelihood function must be evaluated for each point in the scan.

The likelihood function usually consist of several components, e.g. representing agreement with experimental data or cosmological observations. Some components of the likelihood can also serve to maximise a particular value, such as the finding the highest allowed reheating temperature in [4], see also section 7.3.

Being a function of the scanned parameters, the total likelihood increases in complexity with the dimensionality of the parameter space. Hence, there is a numerical trade-off; the more expensive the computation of the likelihood, the less resources available for sampling the parameter space. Concurrently, if the likelihood is a complicated function with many minima or degenerate directions, a dense sampling may be required in order to adequately describe its topology.

There are several tools available for doing this in practice, and the tool used in [4] is the Bayesian inference algorithm `MultiNEST`, which will be presented in the following.



(a) (b)

Figure 3.12: (a) A two-dimensional depiction of a sphere within a cube, with the difference in volume ΔV_D shaded in dark red. (b) The difference in volume between a d-dimensional cube and sphere, as a function of d.

3.5.4 Nested sampling

The Bayesian evidence, introduced in eq. (3.6) and repeated here for convenience,

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta, \tag{3.48}$$

is an integral over the entire space spanned by Θ and thus potentially computationally expensive or even infeasible to evaluate. As a means of numerically approximating the integral, Skilling [37] developed an algorithm called Nested sampling, and as the `MultiNEST` tutorial states [38]:

> Nested sampling estimates the Bayesian evidence by transforming the multi-dimensional evidence integral over the prior density into a one-dimensional integral over an inverse survival function (with respect to prior mass) for the likelihood itself.

The reader for which everything was left crystal clear by this may proceed to the next section.

Before unnesting the inner workings of `MultiNEST`, the concept of a survival function, here conveniently denoted S, must be introduced. Let Λ be a random variable with probability density function f_Λ. Its cumulative distribution function

$$F_\Lambda(\lambda) = P(\Lambda \leq \lambda) = \int_{-\infty}^{\lambda} f_\Lambda(x)dx \tag{3.49}$$

where λ is a particular measurement, is the complement of the survival function,

$$S_\Lambda(\lambda) = P(\Lambda > \lambda) = 1 - F_\Lambda(\lambda) = \int_{\lambda}^{\infty} f_\Lambda(x)dx\,. \tag{3.50}$$

The latter represents the probability that the random variable Λ takes on a value larger than λ.[9]

[9] For a more intuitive picture, replace (λ, Λ) by (t, T) and interpret the probability as that for and event not having occurred by the time t.

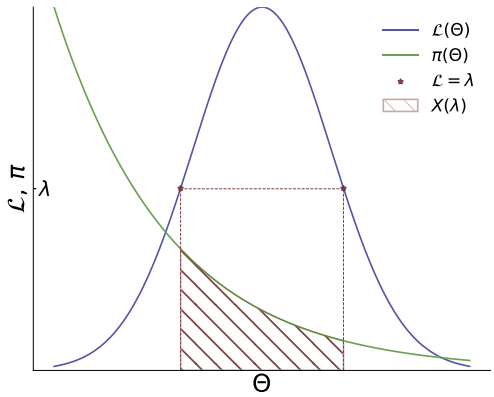


Figure 3.13: The integral over the prior $\pi(\Theta)$ in the region where the likelihood $\mathcal{L}(\Theta)$ is larger than some value λ. This corresponds to eq. (3.52), but since the illustration is one-dimensional, the shaded region comprises the prior *mass*, not the prior volume.

As will shortly become useful, the survival function is related to the expectation value of Λ via

$$\begin{aligned} E[\Lambda] &= \int_0^\infty x f_\Lambda(x)\, dx \\ &= \int_{x=0}^{x=\infty} \left(\int_{\lambda=0}^{\lambda=x} d\lambda \right) f_\Lambda(x)\, dx \\ &= \int_{\lambda=0}^{\lambda=\infty} \int_{x=\lambda}^{x=\infty} f_\Lambda(x)\, dx d\lambda \\ &= \int_0^\infty F_\Lambda(\infty) - F_\Lambda(\lambda)\, d\lambda \\ &= \int_0^\infty 1 - F_\Lambda(t) d\lambda \\ &= \int_0^\infty S_\Lambda(\lambda) d\lambda\,. \end{aligned} \tag{3.51}$$

Returning to `MultiNEST` business, the aforementioned prior mass is the one-dimensional version of the prior volume

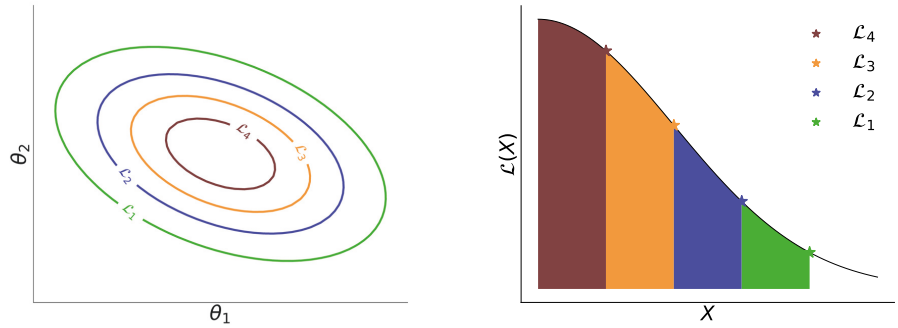
$$X(\lambda) = \int_{\{\Theta : \mathcal{L}(\Theta) > \lambda\}} \pi(\Theta)\, d\Theta = \int_{\mathcal{L}(\Theta)=\lambda}^{\mathcal{L}(\Theta)=\infty} \pi(\Theta)\, d\Theta\,, \tag{3.52}$$

which simply means the amount of prior probability contained in the region(s) where the likelihood $\mathcal{L}(\Theta)$ is larger than some value λ, see fig. 3.13 for a one-dimensional illustration.

The probability $P(\mathcal{L}(\Theta) > \lambda)$ is, treating the likelihood $\mathcal{L}(\Theta)$ like a random variable Λ, the survival function $S_\Lambda(\lambda)$. Using eq. (3.51), the evidence is thus

$$\mathcal{Z} = E_\pi[\mathcal{L}(\Theta)] = \int_0^\infty S_{\mathcal{L}(\Theta)}(\lambda) d\lambda = \int_0^\infty X(\lambda) d\lambda\,. \tag{3.53}$$

That is, the survival function is equal to the above defined $X(\lambda)$, which holds without any restrictions on $\mathcal{L}(\Theta)$ or $\pi(\Theta)$ [39].



(a) (b)

Figure 3.14: (a) The posterior of a two-dimensional problem with the iso-likelihood contours $\mathcal{L}_i$ indicated, and (b) the transformed $\mathcal{L}_i = \mathcal{L}(X_i)$ with the prior volumes X on the x-axis.

Looking again at fig. 3.13, the shaded region is obviously zero at $\lambda = \max_\Theta \mathcal{L}(\Theta)$. This means that the prior volume $X(\mathcal{L}_{\max}) = 0$, while it covers the whole range if evaluated from the origin, $X(0) = 1$. Hence, $X(\lambda)$ is a monotonically decreasing function of λ, and while λ can take values in the range $[0, \infty]$, $X(\lambda)$ is constrained to $[0, 1]$.

Taking a brief step back to the cumulative distribution function,

$$F_\Lambda(\lambda) = P(\Lambda \leq \lambda) = \alpha_\lambda \,, \tag{3.54}$$

represents the probability that Λ takes a value smaller than λ. Its inverse G, referred to as the percent point function, then represents the value of λ for which $P(\Lambda \leq \lambda) = \alpha$, i.e. $G(\alpha) = \lambda$.

Analogously, the inverse $\mathcal{L}$ of the survival function is also a percent point function, $\mathcal{L}(X(\lambda)) = \lambda$, and for the present case *strictly decreasing*, based on the earlier discussion about the behaviour of $X(\lambda)$. This can be used to rewrite the integral in eq. (3.53) one more time as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX \,. \tag{3.55}$$

This is now a one-dimensional integral which furthermore has a positive and decreasing integrand, and is thus well-behaved. Note that this function $\mathcal{L}(X)$ takes a scalar argument and must not be confused with $\mathcal{L}(\Theta)$. Comparing the new expression for $\mathcal{Z}$ with the one in eq. (3.48), the transformation from Θ to X involved dividing the prior volume into infinitesimal pieces and sorting them based on the likelihood $\mathcal{L}(X)$.

This ends the attempt at making the contents in that one sentence from the `MultiNEST` tutorial clearer, and the last step is to understand what actually happens during a scan.

A MultiNEST scan

If the recently introduced $\mathcal{L}(X)$ were known, it could be evaluated from $\mathcal{L}_i = \mathcal{L}(X_i)$ for a sequence of points X in increasing order as

$$0 < X_n < \cdots < X_2 < X_1 < 1 \,. \tag{3.56}$$

Then the numerical integral could be approximated as the weighted sum

$$\mathcal{Z} \approx \sum_{i=1}^{n} w_i \mathcal{L}_i \,, \tag{3.57}$$

with weights w_i equal to the widths ΔX_i, see the illustration in fig. 3.14b.

In practice, this is done by an iterative random mechanism, which is the reason behind the name "nested sampling".

1. Draw N_{live} independent live points Θ_i from the prior distribution $\pi(\Theta)$
2. Evaluate the likelihoods $\mathcal{L}_i = \mathcal{L}(\Theta_i)$
3. Discard the point Θ_l with the lowest likelihood $\mathcal{L}_l$
4. Replace the discarded point by a new draw Θ_{new} from $\pi(\Theta)$, conditional upon $\mathcal{L}_{\text{new}} \geq \mathcal{L}_l$
5. Repeat from step 3 until some stopping criterion is reached

Following this procedure results in samples from regions of increasing likelihood, i.e. the algorithm provides more likely points in parameter space. Concurrently, as the discarded points correspond to samples of increasing likelihood

$$0 < \mathcal{L}_1 < \mathcal{L}_2 < \dots , \tag{3.58}$$

the prior volume X contained within the contours of these likelihoods is steadily decreasing,

$$1 > X_1 > X_2 > \dots , \tag{3.59}$$

exactly what is needed to evaluate eq. (3.57) numerically, if the value of X_i can be found.

The relation $dX = \pi(\Theta)d\Theta$ implies that samples drawn from $\pi(\Theta)$ corresponds to a uniform sampling in X. Together with the relation in eq. (3.59), this means that at iteration i, the point to be discarded — the one with the lowest $\mathcal{L}$ and highest X — is

$$X_i = t_i X_{i-1} . \tag{3.60}$$

Here, t_i is the largest of N_{live} samples drawn uniformly from $[0, 1]$, i.e. the pdf for t is

$$f(t) = N_{\text{live}} t^{N_{\text{live}}} , \tag{3.61}$$

with mean and variance

$$\mu_{\log t} = -\frac{1}{N} \qquad \sigma^2_{\log t} = \frac{1}{N^2} . \tag{3.62}$$

Since the first sample in the scan is drawn from the entire prior volume, $X_0 = 1$, eq. (3.60) implies

$$X_i = t_i t_{i-1} \dots t_1 , \tag{3.63}$$

and correspondingly,

$$\log X_i = \log t_i + \log t_{i-1} + \dots + \log t_1 . \tag{3.64}$$

This can, using eq. (3.62) be approximated as

$$\log X_i \approx -\frac{i}{N} \pm \frac{\sqrt{i}}{N} \quad \Rightarrow \quad X_i \approx e^{-\frac{i}{N}} , \tag{3.65}$$

and the evidence integral can finally be approximated as

$$\mathcal{Z} \approx \sum_i \mathcal{L}_i \Delta X_i = \frac{1}{2} \sum_i \mathcal{L}_i (X_{i-1} - X_{i+1}) \tag{3.66}$$

where ΔX_i is exactly the prior mass (or volume) contained within the disk between two iso-likelihood contours, depicted in fig. 3.14a.

Once the evidence has been calculated, each of the sampled points can be assigned a posterior weight, in accordance with Bayes' law,

$$p_i = \frac{\mathcal{L}_i \Delta X_i}{\mathcal{Z}} . \tag{3.67}$$

Before ending this discussion, a couple of remarks: From a frequentist perspective, comparison of the likelihoods of different points is enough. Nested sampling is an efficient way of obtaining a likelihood-guided sampling of the parameter space, and does not require the Bayesian interpretation. It is in calculating the evidence and when making inferences based on the posterior that the Bayesian interpretation is inevitably chosen. On a more technical note, a requirement for this algorithm to work efficiently, is that it be able to efficiently sample the prior distribution under the constraint $\mathcal{L}_{\text{new}} > \mathcal{L}_{\text{discard}}$. If new samples are drawn from the entire prior distribution, this inevitably becomes more inefficient as the maximum of the likelihood is approached. This problem is evaded in `MultiNEST` due to improvements in [40], from which followed the implementation that new samples drawn from the prior distribution are done so from an ellipsoid containing the current set of live points. This of course comes with the side effect that any possible sharp features in the "true" posterior distribution are evened out to an ellipsoidal shape. Multimodal distributions, i.e. more than one maximum of the likelihood, are tackled by using clustering [41], which assigns one ellipsoid to each cluster of live points.

Chapter 4

Supersymmetry

The reason why supersymmetry, short SUSY, became so popular and has kept its place as one of the strongest candidates for a theory beyond the Standard Model, despite discouraging lack of experimental evidence, is certainly its wide range of appealing features. On the one hand, it could solve many of the challenges the Standard Model currently faces, e.g. by providing candidates for dark matter, and perhaps even relate the cosmological constant to theory [42] — in this case, that of SUSY breaking — and solving the Higgs hierarchy problem by cancelling the loop correction terms to each order exactly. On the other hand, it also has a purely theoretical appeal, independent of unanswered phenomenological questions, and can be motivated by more fundamental theories, such as string theories or supersymmetric grand unified theories. For instance, gauge coupling unification at a high energy scale does not occur in the Standard Model, while it does in supersymmetry, see fig. 4.1.

By providing a relation between bosons and fermions, i.e. force and matter, in the form of a symmetry, supersymmetry unifies all the constituents of particle physics. If supersymmetry is discovered, all mathematically consistent spacetime symmetries will have been realized in nature.

Supersymmetry is formulated using superfields, which correspond to the field operators known from quantum field theory, but formulated in superspace. The superspace considered in the present work is 4-dimensional, $N = 1$ super Minkowski space, meaning ordinary $\mathbb{R}^4$ spacetime, represented usually by the bosonic coordinates x^μ ($\mu = 0,1,2,3$), extended by the fermionic coordinates $\theta^\alpha, \bar{\theta}^\alpha$ ($\alpha = 1,2$), which transform as Weyl spinors.

The following contains a brief introduction to supersymmetry formalism, the minimal supersymmetric extension of the Standard Model, supersymmetry breaking and the mass spectrum, followed by an introduction to the topics discussed in *Trilinear-augmented gaugino mediation* [1].

4.1 Supersymmetry algebra

The supersymmetry algebra is a Lie superalgebra, of which the bosonic part is the Poincaré algebra. The supersymmetric extension of the Poincaré algebra consists of a fermionic part, constructed using Grassmann numbers. The supersymmetry operators, which act on superfields, are [44]

$$\hat{Q}_\alpha = i\frac{\partial}{\partial\theta^\alpha} - (\sigma^\mu\theta^\dagger)_\alpha\partial_\mu \qquad \hat{Q}^\alpha = -i\frac{\partial}{\partial\theta_\alpha} + (\theta^\dagger\overline{\sigma}^\mu)^\alpha\partial_\mu \tag{4.1}$$
$$\hat{Q}^{\dagger\dot{\alpha}} = i\frac{\partial}{\partial\theta^\dagger_{\dot{\alpha}}} - (\overline{\sigma}^\mu\theta)^{\dot{\alpha}}\partial_\mu \qquad \hat{Q}^\dagger_{\dot{\alpha}} = -i\frac{\partial}{\partial\theta^{\dagger\dot{\alpha}}} + (\theta\sigma^\mu)_{\dot{\alpha}}\partial_\mu\,,$$

where the indices and σ^μ matrices are explained in section 4.2. The $\hat{Q}$'s are sometimes referred to as the SUSY charges, and they satisfy the algebra

$$\begin{aligned}\{Q_\alpha, Q^\dagger_{\dot\beta}\} &= (\sigma^\mu)_{\alpha\dot\beta} P_\mu \\ \{Q_\alpha, Q_\beta\} &= \{Q^\dagger_{\dot\alpha}, Q^\dagger_{\dot\beta}\} = 0\,,\end{aligned} \tag{4.2}$$

where $P_\mu = i\partial_\mu$. It is convenient to introduce the SUSY covariant derivatives,

$$\mathcal{D}_\alpha = \frac{\partial}{\partial\theta^\alpha} - i(\sigma^\mu\theta^\dagger)_\alpha\partial_\mu \qquad \mathcal{D}^\alpha = -\frac{\partial}{\partial\theta_\alpha} + i(\theta^\dagger\overline{\sigma}^\mu)^\alpha\partial_\mu \tag{4.3}$$

$$\overline{\mathcal{D}}^\alpha = \frac{\partial}{\partial\theta^\dagger_{\dot\alpha}} - i(\overline{\sigma}^\mu\theta)^{\dot\alpha}\partial_\mu \qquad \overline{\mathcal{D}}_\alpha = -\frac{\partial}{\partial\theta^{\dagger\dot\alpha}} + i(\theta\sigma^\mu)_\alpha\partial_\mu\,, \tag{4.4}$$

and comparing these to the SUSY charges, it is obvious that the covariant derivative of a superfield is another superfield. In fact,

$$\{\hat{Q}_\alpha, \mathcal{D}_\beta\} = \{\hat{Q}^\dagger_{\dot\alpha}, \mathcal{D}_\beta\} = \{\hat{Q}_\alpha, \overline{\mathcal{D}}_{\dot\beta}\} = \{\hat{Q}^\dagger_{\dot\alpha}, \overline{\mathcal{D}}_{\dot\beta}\} = 0\,. \tag{4.5}$$

The Poincaré algebra, see eq. (2.11), is thus extended to a super-Poincaré algebra.

4.2 Formalism and notation

4.2.1 Spinors and indices

In this section, two-component notation for fermions is introduced, and will be used in the following. Building upon the short introduction to spinors in section 2.1.2, the Dirac spinors are re-introduced, now with explicit spinor indices,

$$\Psi_D = \begin{pmatrix}\xi_\alpha \\ \chi^{\dagger\dot\alpha}\end{pmatrix} \qquad \alpha,\dot\alpha = 1,2 \tag{4.6}$$

and again, using the projection operators defined in eq. (2.38), the different components are projected out as

$$P_L\Psi_D = \begin{pmatrix}\xi_\alpha \\ 0\end{pmatrix}, \qquad P_R\Psi_D = \begin{pmatrix}0 \\ \chi^{\dagger\dot\alpha}\end{pmatrix}, \tag{4.7}$$

so in this notation, left-handed, undaggered fields carry undotted indices, while right-handed, daggered fields carry dotted indices. In two-component formalism, the Clifford algebra is generated by the three Pauli matrices, eq. (2.14). Defining

$$\sigma^\mu \equiv (\mathcal{I}_2, \vec{\sigma})\,, \qquad \bar{\sigma}^\mu = (\mathcal{I}_2, -\vec{\sigma})\,, \tag{4.8}$$

the index convention corresponds to the one for the spinors; $(\sigma^\mu)_{\alpha\dot\alpha}$ and $(\bar{\sigma}^\mu)^{\dot\alpha\alpha}$. Dotted and undotted indices can be manipulated using the relations

$$\begin{aligned}\xi^\alpha &= \varepsilon^{\alpha\beta}\xi_\beta\,, \quad \xi_\alpha = \varepsilon_{\alpha\beta}\xi^\beta \\ \chi^{\dagger\dot\alpha} &= \varepsilon^{\dot\alpha\dot\beta}\chi^\dagger_{\dot\beta}\,, \quad \chi^\dagger_{\dot\alpha} = \varepsilon_{\dot\alpha\dot\beta}\chi^{\dagger\dot\beta} \\ (\bar\sigma)^{\dot\alpha\alpha} &= \varepsilon^{\alpha\beta}\varepsilon^{\dot\alpha\dot\beta}(\sigma^\mu)_{\beta\dot\beta}\,,\end{aligned} \tag{4.9}$$

and are usually suppressed, e.g. [44]

$$\xi^\dagger_{\dot\alpha}(\bar\sigma^\mu)^{\dot\alpha\alpha}\chi_\alpha = \xi^\dagger\bar\sigma^\mu\chi = -\chi\sigma^\mu\xi^\dagger = -\chi^\alpha(\sigma^\mu)_{\alpha\dot\alpha}\xi^{\dagger\dot\alpha}\,. \tag{4.10}$$

4.2.2 Supermultiplets

A supermultiplet is a representation of a supersymmetry algebra, and consists of particles which are superpartners of each other, meaning that they transform into each other under the operators introduced in eq. (4.1). Two kinds of supermultiplets are relevant for the present discussion.

Chiral supermultiplets

These contain the Standard Model matter particles. Here denoted Φ, chiral supermultiples contain a scalar ϕ, a Weyl-fermion ψ and an auxiliary field F. The latter is not a physical field, i.e. does not have kinetic nor mass terms in the Lagrangian, and is merely included to ensure the supersymmetry algebra close off-shell[1]. To find the exact form of a chiral supermultiplet, the only constraint needed is

$$\overline{\mathcal{D}}_{\dot{\alpha}}\Phi = 0\,, \quad \text{or} \quad \mathcal{D}_{\alpha}\Phi^{\dagger} = 0\,, \tag{4.11}$$

where a field satisfying the first one is a left-chiral, or just chiral for short, supermultiplet, while a field satisfying the second one is a right-chiral, or anti-chiral supermultiplet. Changing coordinates to $y^{\mu} \equiv x^{\mu} + i\theta\overline{\sigma}^{\mu}\theta$ and rewriting eq. (4.3) and eq. (4.4) yields

$$\mathcal{D}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} - 2i(\sigma^{\mu}\theta^{\dagger})_{\alpha}\frac{\partial}{\partial y^{\mu}} \qquad \mathcal{D}^{\alpha} = -\frac{\partial}{\partial\theta_{\alpha}} + 2i(\theta^{\dagger}\overline{\sigma}^{\mu})^{\alpha}\frac{\partial}{\partial y^{\mu}} \tag{4.12}$$

$$\overline{\mathcal{D}}^{\dot{\alpha}} = \frac{\partial}{\partial\theta^{\dagger}_{\dot{\alpha}}} \qquad \overline{\mathcal{D}}_{\dot{\alpha}} = -\frac{\partial}{\partial\theta^{\dagger\dot{\alpha}}}\,. \tag{4.13}$$

Combining the first condition in eq. (4.11) with eq. (4.13) reveals that a chiral supermultiplet is of the form

$$\Phi(y,\theta) = \phi(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y)\,, \tag{4.14}$$

where the factor $\sqrt{2}$ is conventional. It is obvious from eq. (4.14) that the gauge quantum numbers and mass dimension of the supermultiplet Φ are those of its scalar component ϕ. Integrating Φ over superspace, see eq. (A.1), yields

$$[\Phi(y,\theta)]_F \equiv \int d^2\theta\,\Phi(y,\theta) = \int d^2\theta\left(\phi(y,\theta) + \sqrt{2}\theta\psi(y) + \theta\theta F(y)\right) = F\,, \tag{4.15}$$

which is why this is referred to as 'taking the F-term'.

Vector supermultiplets

These contain the Standard Model gauge bosons. Denoted V, they contain a gauge boson field A_{μ}, a Weyl fermion λ, referred to as gaugino, and an auxiliary field D. This auxiliary field plays exactly the same role as F in the chiral supermultiplet[2]. A general form of V can be obtained by imposing the reality condition

$$V^{*} = V\,. \tag{4.16}$$

In general, V can have several auxiliary fields, but all except one can be transformed away via a gauge transformation in superspace. The simplest supermultiplet satisfying eq. (4.16) is in Wess-Zumino gauge, see e.g. [44] for details,

$$V_{\text{WZ}}(x,\theta,\theta^{\dagger}) = \theta^{\dagger}\overline{\sigma}^{\mu}\theta A_{\mu} + \theta^{\dagger}\theta^{\dagger}\theta\lambda + \theta\theta\theta^{\dagger}\lambda^{\dagger} + \frac{1}{2}\theta\theta\theta^{\dagger}\theta^{\dagger}D\,, \tag{4.17}$$

where the arguments of the fields have been, and will henceforth be, omitted. Integrating this over superspace, see eq. (A.1), yields

$$[V]_D \equiv \int d^2\theta d^2\theta^{\dagger} V = \int d^2\theta d^2\theta^{\dagger}\left(\theta^{\dagger}\overline{\sigma}^{\mu}\theta A_{\mu} + \theta^{\dagger}\theta^{\dagger}\theta\lambda + \theta\theta\theta^{\dagger}\lambda^{\dagger} + \frac{1}{2}\theta\theta\theta^{\dagger}\theta^{\dagger}D\right) = D\,, \tag{4.18}$$

which is why this is referred to as 'taking the D-term'.

[1] Meaning that $\delta\mathscr{L}_F = 0$ on-shell, while it exactly cancels the excess (fermionic) degrees of freedom off-shell.

[2] The auxiliary field D cancels both fermionic and bosonic degrees of freedom off-shell.

4.2.3 Superspace Lagrangians

The following contains a brief discussion of what a Lagrangian representing a supersymmetric theory should comprise and obey. More details can be found in the literature which forms the basis for the present discussion [44, 45].

As demonstrated in eq. (4.18), the integral of a function over superspace is independent of θ and $\theta^\dagger$, owing to the Grassmannian nature of the superspace coordinates. This means that for any superfield $S(x,\theta,\theta^\dagger)$, the variation of the following quantity under supersymmetric transformations vanishes

$$A = \int d^4x \int d^2\theta d^2\theta^\dagger S(x,\theta,\theta^\dagger)\,, \tag{4.19}$$

since the SUSY generators in eq. (4.1) are made up of derivatives of x, θ, $\theta^\dagger$. As indicated by the notation, A represents an action. This means that the above can be used to build a superspace Lagrangian. Based on eq. (4.15), the F-term of any holomorphic function $W(\Phi)$ of chiral superfields is also a constant, and can be included in the action. In that case, the hermitian conjugate must also be added, since the action must be real. Finally, the way to include vector supermultiplets in the F-term part is by realising that a chiral supermultiplet can be constructed from a vector supermultiplet as

$$\mathcal{W}_\alpha = \overline{\mathcal{D}\mathcal{D}}\mathcal{D}_\alpha V\,, \qquad \mathcal{W}^\dagger_{\dot\alpha} = \mathcal{D}\mathcal{D}\overline{\mathcal{D}}_{\dot\alpha} V\,. \tag{4.20}$$

Note that this chiral supermultiplet carries an index, which means that it is a spinor chiral superfield and transforms like a Weyl spinor, as opposed to the ones introduced earlier, which are scalars. To summarise, a general superspace Lagrangian is of the form

$$\mathscr{L}(x,\theta,\theta^\dagger) = \left[K(\Phi,\Phi^\dagger,V)\right]_D + \left(\left[f_{ab}(\Phi)\mathcal{W}^{a\alpha}\mathcal{W}^b_\alpha + W(\Phi)\right]_F + \text{h.c.}\right), \tag{4.21}$$

where K is a real function of superfields, and can in principle also be a function of spacetime derivatives of superfields.[3]

Additional requirements to the above Lagrangian are, as already mentioned, that it be real and invariant under gauge transformations, that each term have the correct mass dimension and, in the present case, that it reproduce the Standard Model.

4.3 The MSSM

The Minimal Supersymmetric Standard Model (MSSM) is minimal in the sense that it is the smallest extension, in terms of particle content and interactions, of the Standard Model necessary to accomodate supersymmetry and be consistent with observations. An overview of the MSSM supermultiplets' field content and gauge quantum numbers is presented in table 4.1, which is the supersymmetric extension of table 2.1. In the following, the three functions in eq. (4.21) for the MSSM are introduced and briefly explained.

4.3.1 The MSSM Lagrangian

The superpotential

All interactions not involving gauge fields or couplings, i.e. between chiral superfields alone, go into the superpotential W. This is where e.g. the Yukawa couplings are found and thus Standard Model fermion masses are generated. Demanding that the Lagrangian have mass dimension $[\mathscr{L}] = 4$, the mass dimension of the superpotential must be $[W] = 3$, meaning that only terms containing at most three chiral superfields are allowed, see also table 4.2. In the MSSM, the superpotential takes the form

$$W_{\text{MSSM}} = \bar u y_u Q H_u - \bar d y_d Q H_d - \bar e y_e L H_d + \mu H_u H_d\,, \tag{4.22}$$

where the $SU(2)$, generation and colour indices have been, and will henceforth be, suppressed, but behave as explained in section 4.2.1, i.e. $\bar u y_u Q H_u = \bar u^{ia}(y_u)^j_i Q_{ja\alpha}(H_u)_\beta \epsilon^{\alpha\beta}$ and $\mu H_u H_d = \mu (H_u)_\alpha (H_d)_\beta \epsilon^{\alpha\beta}$. The terms in the superpotential

[3] Note that $f_{ab} = \delta_{ab}$ in the MSSM.

Chiral supermultiplet	Scalar	Fermion	$SU(3)_c, SU(2)_L, U(1)_Y$
Q_i — Left-handed *s*quark isospin doublet	$(\widetilde{u}_{iL} \quad \widetilde{d}_{iL})$	$(u_{iL} \quad d_{iL})$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
$\bar{u}_i$ — Right-handed up-*s*quark	$\widetilde{u}^*_{iR}$	$u^\dagger_{iR}$	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
$\bar{d}_i$ — Right-handed down-*s*quark	$\widetilde{d}^*_{iR}$	$d^\dagger_{iR}$	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
L_i — Left-handed *s*lepton isospin doublet	$(\widetilde{\nu}_{iL} \quad \widetilde{e}_{iL})$	$(\nu_{iL} \quad e_{iL})$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
$\bar{e}_i$ — Right-handed charged *s*lepton	$\widetilde{e}^*_{iR}$	$e^\dagger_{iR}$	$(\mathbf{1}, \mathbf{1}, 1)$
H_u — Up-type Higgs*ino*	$(H_u^+ \quad H_u^0)$	$(\widetilde{H}_u^+ \quad \widetilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, \frac{1}{2})$
H_d — Down-type Higgs*ino*	$(H_d^0 \quad H_d^-)$	$(\widetilde{H}_d^0 \quad \widetilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
Vector supermultiplet	Vector boson	Fermion	$SU(3)_c, SU(2)_L, U(1)_Y$
g — Gluon *gluino*	g	$\widetilde{g}$	$(\mathbf{8}, \mathbf{1}, 0)$
W — Weak isospin bosons *winos*	$W^+ \; W^- \; W^0$	$\widetilde{W}^+ \; \widetilde{W}^- \; \widetilde{W}^0$	$(\mathbf{1}, \mathbf{3}, 0)$
B — Weak hypercharge boson *bino*	B^0	$\widetilde{B}^0$	$(\mathbf{1}, \mathbf{1}, 0)$

Table 4.1: The chiral and vector supermultiplets of the MSSM. The generation index $i = 1, 2, 3$ runs over (up, strange, top) and (down, charm, bottom) quarks, and (electron, muon, tau) leptons. The numbers in parentheses represent the transformation properties under the Standard Model gauge groups. Superpartner names are indicated in *italic*.

make it obvious why the MSSM requires two separate Higgs doublets[4] with different quantum numbers; given that W must be a holomorphic function of the supermultiplets, there is no way to couple only one Higgs supermultiplet to all quark and lepton supermultiplets in a gauge invariant way. In addition to the interaction terms which also end up giving masses to the quark and leptons after electroweak symmetry breaking, the superpotential contains a term coupling the two Higgses, commonly referred to as 'the μ-term'.

The field strength superfield

Moving on to the next F-term in the Lagrangian, there are the field strength superfields $\mathcal{W}^\alpha$, which are chiral superfields, accompanied by the gauge kinetic function $f_{ab}(\Phi)$, which is itself a chiral superfield. The field strength superfields are constructed from vector superfields, in line with eq. (4.20). Evaluating this in Wess-Zumino gauge yields

$$\mathcal{W}_\alpha = \lambda_\alpha + \theta_\alpha D + \frac{i}{2}(\sigma^\mu \bar{\sigma}^\nu \theta)_\alpha F_{\mu\nu} + i\theta\theta(\sigma^\mu \partial_\mu \theta^\dagger)_\alpha \,. \tag{4.23}$$

The Kähler potential

Now for the D-term in the Lagrangian. The Kähler potential is not a holomorphic function of superfields, and it can contain both chiral, anti-chiral and vector superfields. Having the (anti-)chiral supermultiplets invariant under super-gauge transformations as

$$\begin{aligned} \Phi &\to \mathrm{e}^{2ig\Omega^a T^a}\Phi \\ \Phi^\dagger &\to \Phi^\dagger \mathrm{e}^{-2ig\Omega^a T^a}\,, \end{aligned} \tag{4.24}$$

where the super-gauge transformation parameters Ω^a are also chiral superfields, a general Kähler potential can be written as

$$K(\Phi, \Phi^\dagger, V) = \Phi^\dagger \mathrm{e}^{2gT^a V^a}\Phi, \tag{4.25}$$

with g the gauge coupling of a single the gauge group, and the vector supermultiplets V^a containing the corresponding gauge bosons and gauginos. The generalisation to several groups is trivial. In the MSSM, using the chiral and vector

[4] Specifically, the MSSM contains a two-Higgs doublet model of Type II.

supermultiplets from eqs. (4.14) and (4.17), the explicit form is

$$\left[K(\Phi,\Phi^\dagger,V)\right]_D = F^\dagger F + \nabla_\mu\phi^\dagger\nabla^\mu\phi + i\psi^\dagger\bar{\sigma}^\mu\nabla_\mu\psi - \sqrt{2}g_a(\phi^\dagger T^a\psi)\lambda^a - \sqrt{2}\lambda^\dagger(\psi^\dagger T^a\phi) + g_a(\phi^\dagger T^a\phi)D^a\,. \tag{4.26}$$

Here, ∇_μ is the gauge-covariant spacetime derivative:

$$\begin{aligned} \nabla_\mu\phi_i &= \partial_\mu\phi_i - igq_iA_\mu\phi_i, & \nabla_\mu\phi^{*i} &= \partial_\mu\phi^{*i} + igq_iA_\mu\phi^{*i}, \\ \nabla_\mu\psi_i &= \partial_\mu\psi_i - igq_iA_\mu\psi_i\,, \end{aligned} \tag{4.27}$$

see [44] section 4.8 for details.

4.4 Broken supersymmetry

Supersymmetry must be broken at the energy scales covered by the Standard Model at present, since the Standard Model particles would otherwise be mass degenerate with their superpartners, which would already have been discovered. But just like the electroweak gauge bosons are massive due to the spontaneous breaking of the electroweak symmetry, it is possible that supersymmetry is spontaneously broken at low energies. Supersymmetry could thus still be a symmetry of the Lagrangian, while the vacuum state of the universe is not invariant under SUSY, i.e.

$$\hat{Q}_\alpha|0\rangle \neq 0 \quad \text{and} \quad \hat{Q}^\dagger_{\dot{\alpha}}|0\rangle \neq 0\,. \tag{4.28}$$

From the SUSY algebra, eq. (4.1), the above amounts to positive vacuum energy, $\hat{H}|0\rangle > 0$. For this to be true regardless of kinetic contributions to $\hat{H}$, the SUSY scalar potential $V(\phi,\phi^\dagger)$ must satisfy $\langle 0|V(\phi,\phi^\dagger)|0\rangle > 0$. The terms contributing to $V(\phi,\phi^\dagger)$ are contained in the Kähler potential part of the Lagrangian, eq. (4.26), and can be rewritten in terms of the auxiliary fields of the chiral and vector supermultiplets as

$$V(\phi,\phi^\dagger) = F^\dagger F + g_a(\phi^\dagger T^a\phi)D^a = F^\dagger F + \frac{1}{2}\sum_a D^aD^a\,. \tag{4.29}$$

The form of this equation reveals that spontaneous symmetry breaking can be classified as either F-term or D-term breaking, depending on which part of the potential acquires a non-zero VEV.

For reasons that will not be elaborated further in the present work, it is clear that within the MSSM, there is no suitable candidate supermultiplet whose F- or D-term could develop a VEV and thus give rise to the spontaneous breaking of SUSY. Because of this, SUSY is likely to be broken in a hidden sector scenario, while the field contents of the MSSM resides in a visible sector. One or several hidden sector fields would have an auxiliary field responsible for the breaking of SUSY, and the effects of this somehow mediated to the visible sector.

There are many different hidden sector models, but the phenomenological feature shared by all these models is that it is not the mechanism of the spontaneous SUSY breaking itself that prescribes how broken SUSY is manifested in the visible sector, but rather its mediation.

Planck scale mediated SUSY breaking (PMSB), or gravity mediated SUSY breaking, assumes that the SUSY breaking mediating interactions are gravitational, and as such connected to New Physics, perhaps including quantized gravity, which enters near the Planck scale. Gauge mediated SUSY Breaking (GMSB) on the other hand, postulates messenger fields which are chiral supermultiplets, charged under at least one MSSM gauge group. These also couple to the hidden sector auxiliary field which aquires a VEV, and thus contribute to mass terms for the visible sector fields through loop interactions.

These models for SUSY breaking are not discussed further in the present work, but a nice and comprehensive overview is given in [46]. The SUSY breaking mechanism considered in the following is gaugino mediated SUSY breaking.

4.4.1 Soft supersymmetry breaking

Though SUSY breaking happens at some scale not accessible to us and through some mechanism not assessable by us, it is possible to set aside this ignorance and simply consider all possible terms in the Lagrangian that would break

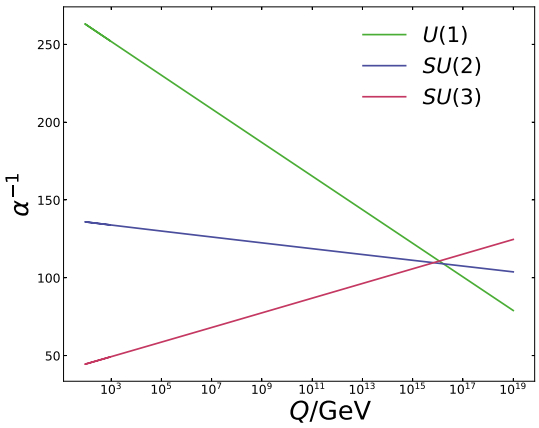


Figure 4.1: The gauge couplings evolve differently with energy and align at a high scale. Plot produced using `SPheno` [8].

SUSY. In order not to spoil the appeal of the theory by introducing quadratically divergent radiative corrections, only terms of positive mass dimension are included. These are termed soft, and the possible soft SUSY breaking terms are

$$\mathscr{L}_{\text{soft}} = -\left(\frac{1}{2}M_a\lambda^a\lambda^a + \frac{1}{6}a^{ijk}\phi_i\phi_j\phi_k + \frac{1}{2}b^{ij}\phi_i\phi_j + t^i\phi_i\right) + \text{h.c.} - \left(m^2\right)^i_j\phi^{j*}\phi_i\,, \tag{4.30}$$

where a runs over the gauge groups and thus enumerates the different gauginos, and $\left(m^2\right)^i_j$ are scalar squared-mass terms. Whenever a common gaugino mass is chosen, this is denoted $m_{1/2}$, while a common scalar mass is denoted m_0. The SUSY breaking interactions are a^{ijk}, combining three scalars and in short referred to as trilinears — a common trilinear coupling is denoted A_0 —, and b^{ij} coupling two scalars, which in the MSSM could only be the so-called $B\mu$ term for the Higgses, i.e. a SUSY breaking term with the same structure as the μ term. The tadpole coupling t^i is only allowed for a gauge singlet, which does not exist in the MSSM.

4.4.2 Energy scales

Any candidate theory for describing the mechanism of SUSY breaking must make predictions regarding the parameters in the SUSY breaking part of the Lagrangian. However, these are imposed at some high scale, much higher than the electroweak scale, and must be connected with the phenomenology at the electroweak scale in order to evaluate the theory at hand. In order to do this, the renormalisation group equations, short RGE, are solved, using as boundary conditions the measured values of the gauge couplings, the Standard Model fermion and electroweak boson masses at the weak scale. The one-loop RG equations for the gauge couplings are

$$\frac{d}{dt}g_a = \frac{1}{16\pi^2}b_a g_a^3, \quad a = 1,2,3\,, \tag{4.31}$$

where $t \equiv \ln(Q/Q_0)$, Q is the renormalisation scale and Q_0 the input scale. The corresponding equations for the MSSM gaugino mass parameters are [44]

$$\frac{d}{dt}M_a = \frac{1}{8\pi^2}b_a g_a^2 M_a, \quad a = 1,2,3\,, \tag{4.32}$$

from which three scale-independent ratios are extracted,

$$\frac{M_1}{g_1^2} = \frac{M_2}{g_2^2} = \frac{M_3}{g_3^2}\,. \tag{4.33}$$

The coefficients b_a are smaller in the Standard Model than the MSSM, since the MSSM has more particles and hence loops. The numerical values are

$$(b_1, b_2, b_3) = \begin{cases} (\frac{41}{10}, -\frac{19}{6}, -7) & \text{Standard Model} \\ (\frac{33}{5}, 1, -3) & \text{MSSM.} \end{cases} \tag{4.34}$$

The superpotential parameters of the third family evolve with the energy scale as [44]

$$\begin{aligned} \frac{d}{dt} y_t &= \frac{y_t}{16\pi^2}\left(6y_t^* y_t + y_b^* y_b - \frac{16}{3}g_3^2 - 3g_2^2 - \frac{13}{15}g_1^2\right), \\ \frac{d}{dt} y_b &= \frac{y_b}{16\pi^2}\left(6y_b^* y_b + y_t^* y_t + y_\tau^* y_\tau - \frac{16}{3}g_3^2 - 3g_2^2 - \frac{7}{15}g_1^2\right), \\ \frac{d}{dt} y_\tau &= \frac{y_\tau}{16\pi^2}\left(4y_\tau^* y_\tau + 3y_b^* y_b - 3g_2^2 - \frac{9}{5}g_1^2\right), \\ \frac{d}{dt} \mu &= \frac{\mu}{16\pi^2}\left(3y_t^* y_t + 3y_b^* y_b + y_\tau^* y_\tau - 3g_2^2 - \frac{3}{5}g_1^2\right), \end{aligned} \tag{4.35}$$

and the first and second family squark and slepton squared-mass parameters as

$$16\pi^2 \frac{d}{dt} m_{\phi_i}^2 = -\sum_{a=1,2,3} 8C_a(i) g_a^2 |M_a|^2 + \frac{6}{5} Y_i g_1^2 S, \tag{4.36}$$

where

$$S \equiv \mathrm{Tr}[Y_j m_{\phi_j}^2] = m_{H_u}^2 - m_{H_d}^2 + \mathrm{Tr}[m_Q^2 - m_L^2 - 2m_{\bar{u}}^2 + m_{\bar{d}}^2 + m_{\bar{e}}^2]. \tag{4.37}$$

The $C_a(i)$ are known as the quadratic Casimir group theory invariants for the superfield Φ_i. These are defined in terms of the Lie algebra generators T^a, akin to the discussion around eq. (2.9), as

$$(T^a T^a)_i{}^j = C_a(i)\,\delta_i^j\,. \tag{4.38}$$

For the MSSM supermultiplets, these are explicitly

$$\begin{aligned} C_1(i) &= 3Y_i^2/5 \quad \text{all } \Phi_i \text{ with weak hypercharge } Y_i \\ C_2(i) &= \begin{cases} 3/4 & \Phi_i = Q, L, H_u, H_d, \\ 0 & \Phi_i = \bar{u}, \bar{d}, \bar{e}, \end{cases} \\ C_3(i) &= \begin{cases} 4/3 & \Phi_i = Q, \bar{u}, \bar{d}, \\ 0 & \Phi_i = L, \bar{e}, H_u, H_d. \end{cases} \end{aligned} \tag{4.39}$$

The squared-mass parameters of the Higgs scalars and third family squarks and sleptons receive the same contributions as eq. (4.36), but there are additional contributions from large Yukawa couplings and the soft trilinears. These expressions are large and their derivation tedious, and they will not be stated explicitly here. Perhaps just as enlightening, the energy evolution of all the particle masses for a single parameter point is shown in fig. 4.2.

Similarly tedious is the job of evolving these RG equations between energy scales. Among the most popular programs for doing this numerically are at the time of writing `SPheno` [8], `Softsusy` [9] and `SuSpect` [10]. These programs are called spectrum generators, since their output includes the mass spectrum of the parameter point.

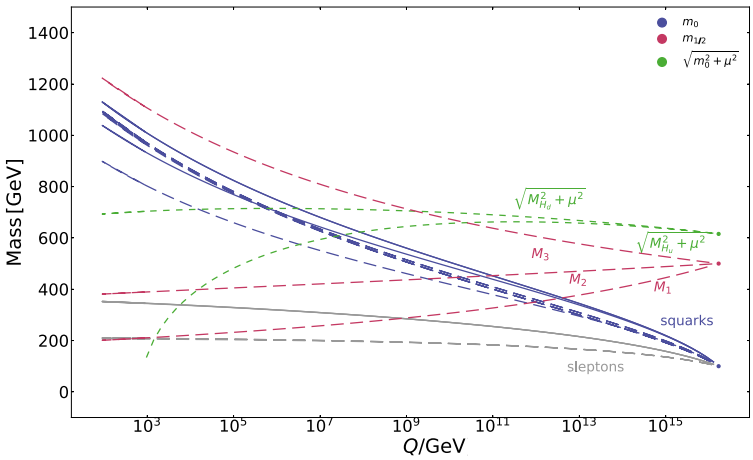


Figure 4.2: The evolution of the SUSY soft masses, as function of energy scale. The calculations are made by `SPheno`, and the input parameters are $m_0 = 100\,\text{GeV}$, $M_1 = M_2 = M_3 = 500\,\text{GeV}$, $\tan\beta = 10$, $A_0 = 0\,\text{GeV}$.

4.4.3 A note on electroweak symmetry breaking and naturalness

Electroweak symmetry breaking in SUSY is slightly complicated by the fact that the Higgs sector of the MSSM is larger than that of the Standard Model; it contains a two-Higgs doublet model, and the Higgs scalar potential is

$$\begin{aligned} V = &(|\mu|^2 + m_{H_u}^2)(|H_u^0|^2 + |H_u^+|^2) + (|\mu|^2 + m_{H_d}^2)(|H_d^0|^2 + |H_d^-|^2) \\ &+ \left(b\,(H_u^+ H_d^- - H_u^0 H_d^0) + \text{h.c.}\right) \\ &+ \frac{1}{8}(g^2 + g'^2)(|H_u^0|^2 + |H_u^+|^2 - |H_d^0|^2 - |H_d^-|^2)^2 + \frac{1}{2}g^2|H_u^+ H_d^{0*} + H_u^0 H_d^{-*}|^2 . \end{aligned} \tag{4.40}$$

The requirement for electroweak symmetry breaking is thus extended to $\partial V/\partial H_u^0 = 0$ and $\partial V/\partial H_d^0 = 0$,[5] and imposing additional requirements such as that the potential be bounded from below, yields a relation between the Higgs VEVs

$$\begin{gathered} \langle H_d^0 \rangle \equiv v_d \qquad \langle H_u^0 \rangle \equiv v_u \\ v_u^2 + v_d^2 = v^2 \end{gathered} \tag{4.41}$$

and the other MSSM parameters, after electroweak symmetry breaking.

The common scale for the sparticle masses after SUSY breaking is M_{SUSY}, necessarily larger than the electroweak scale. Although SUSY solves the hierarchy problem, it comes with a smaller hierarchy problem of its own, which is exactly this large gap between scales. At tree-level, the mass of the Z boson is

$$m_Z^2 = \frac{1}{2}(g^2 + g'^2)(v_u^2 + v_d^2) . \tag{4.42}$$

[5] And the minimum of the Higgs potential satisfying $\partial V/\partial H_u^+ = 0$ must also have $\partial V/\partial H_d^- = 0$, to avoid breaking the electromagnetic symmetry.

Expanding this in large $\tan\beta = \frac{v_u}{v_d}$ yields

$$\begin{aligned} m_Z^2 &= -2(m_{H_u}^2 + |\mu|^2) + \frac{2}{\tan^2\beta}(m_{H_d}^2 - m_{H_u}^2) + \mathcal{O}(1/\tan^4\beta) \\ &= -2(m_{H_u}^2 + |\mu|^2) + \mathcal{O}(1/\tan^2\beta)\,. \end{aligned} \tag{4.43}$$

The value of m_Z is determined by experiment, so the right-hand side of the above equation must combine to yield the measured value. The first term contains m_{H_u}, which is one of the soft SUSY-breaking parameters, eq. (4.30), and thus of the order M_{SUSY}. The second term, on the other hand, contains μ which is a supersymmetric parameter. There is no part of the theory that relates μ to M_{SUSY}, and it might have a completely different origin. This has become known as the "μ-problem", or the "little hierarchy problem" in SUSY. However, even provided some relation between the two, there would remain a fine-tuning problem, since two large numbers have to cancel to yield a very exact, small number. In short, the larger m_{H_u} and μ, the larger the fine-tuning required.

Various measures have been suggested to quantify the amount of fine-tuning required within a setup. For instance, the spectrum generator `Softsusy` uses the measure of absence of fine-tuning, or 'naturalness', c_a of a parameter a, as defined in [47], eq. (4)

$$c_a = \left| \frac{\partial \ln m_Z^2}{\partial \ln a^2} \right|. \tag{4.44}$$

4.5 Gaugino mediated supersymmetry breaking

4.5.1 The brane, the bulk and the breaking

One possibility for mediating supersymmetry breaking is via additional spatial dimensions. In the framework considered in [1], spacetime is in general D dimensional, referred to as the bulk. Out of the D spacetime dimensions, $D-4$ are compact with volume V_{D-4}, which determines the energy scale needed to resolve these dimensions, referred to as the compactification scale $M_c \equiv (1/V_{D-4})^{\frac{1}{D-4}}$. The bulk contains a number of 4-dimensional branes, and all quantum fields are either $D-$ or 4-dimensional, i.e. they are either bulk or brane fields. The D-dimensional Lagrangian is [48]

$$\mathscr{L}_D = \mathscr{L}_{\text{bulk}}\left(\hat{\Phi}(x,y)\right) + \sum_j \delta^{(D-4)}(y-y_j)\,\mathscr{L}_j\left(\hat{\Phi}(x,y_j), \phi_j(x)\right), \tag{4.45}$$

where j runs over the branes, x are coordinates on the branes, y are coordinates in the bulk, $\hat{\Phi}$ is a bulk field and ϕ_j is a field localised on the jth brane. The 4-dimensional field theory on the brane containing the MSSM is an effective field theory valid on energy scales that do not resolve the additional dimensions. This will be discussed more thoroughly in the following, but for now, note the hats on the bulk fields, which indicate that they are canonically normalised in D dimensions, see section 4.5.2 below.

The work in [1] considers two branes; the MSSM brane with all its matter fields, and a source brane at a spatial distance R away from the MSSM brane. The source brane contains at least one chiral superfield S, which is a singlet under the Standard Model gauge groups, and breaks supersymmetry via some mechanism that results in a non-zero expectation value $\langle F_S \rangle$ of its auxiliary field. The SUSY breaking mechanism itself remains unknown, since S is constrained to its brane and no direct interactions with it can be observed from the MSSM brane. However, the breaking of supersymmetry is necessarily communicated to the MSSM brane, and to this end, the mediators are introduced. These are D-dimensional fields that propagate in the bulk and also interact with the fields on each brane, provided that they carry the same charge. An immediate consequence is that these fields obtain soft masses proportional to $\langle F_S \rangle$.

In the original work on gaugino mediated supersymmetry breaking [49], only the gauge superfields were allowed to propagate in the bulk. Supersymmetry breaking in this model is a consequence of gauginos receiving a SUSY breaking mass upon interacting with S, and communicating this to the MSSM brane. This gave rise to the name of gaugino mediation which has stuck ever since, although Higgs superfields were promoted to bulk fields soon after [50]. The present work considers both gauge and Higgs superfields as bulk fields.

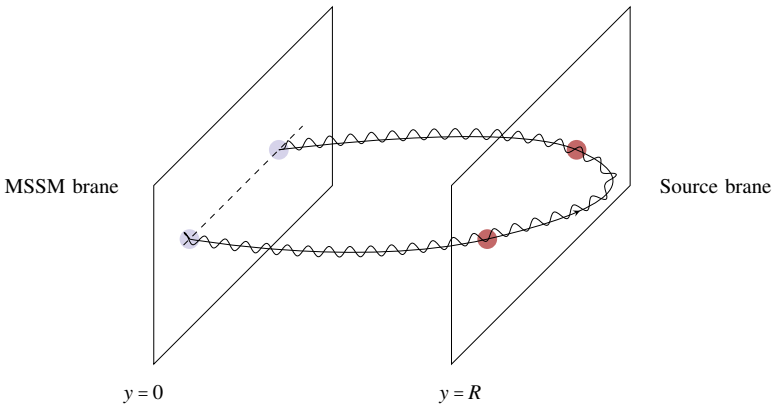


Figure 4.3: A gaugino propagating in the bulk, interacting twice with a scalar on the MSSM brane (blue points) and twice with some chiral field on the source brane (red points).

The process of a gaugino or higgsino propagating from the MSSM brane to the source brane and back is depicted in fig. 4.3. The propagator in this figure is drawn as a gaugino line, but in principle, any bulk superfield which is charged under an MSSM gauge group and interacts with a source brane field could undertake this journey. Also, the propagator on the MSSM brane itself is a scalar, so this particular process shows the correction to squark and slepton masses.

4.5.2 The fields and their dimensions

As mentioned in the introductory section 2.1.2, specifically eq. (2.36) and repeated here for convenience, canonically normalised kinetic terms in D dimensions have mass dimension of

$$[\partial_\mu \phi \partial^\mu \phi] = [\bar{\psi} \partial\!\!\!/ \psi] = [\mathscr{L}] = D\,, \tag{4.46}$$

and as mentioned, the hats on the bulk fields in eq. (4.45) indicate that these fields are have canonically normalised kinetic terms in D dimensions. See table 4.2 for an overview of mass dimensions of the relevant fields, operators and couplings. Following the notation of eq. (4.45), $j = 1$ for the MSSM brane and $j = 2$ for the source brane. Writing down the Lagrangians at hand, the MSSM, containing both bulk and brane fields, has

$$\mathscr{L}_{\text{MSSM}} = \mathscr{L}_{\text{bulk}} + \delta^{(D-4)}(y - y_1)\,\mathscr{L}_1 = \left[W(\hat{\Phi}, \phi_1) + \frac{1}{4}\hat{\mathcal{W}}_\alpha \hat{\mathcal{W}}^\alpha\right]_F + \text{h.c.} + \left[K\left(\hat{\Phi}, \hat{\Phi}^\dagger, \phi_1, \phi_1^\dagger, e^V\right)\right]_D\,, \tag{4.47}$$

where W is the MSSM superpotential, $\hat{\mathcal{W}}$ the field strength superfield and K the Kähler potential. For the source brane, the part of the Lagrangian that is interesting to us contains interactions between the gauge and Higgs superfields, and the chiral field S,

$$\begin{aligned}\mathscr{L}_2 = {} & \frac{1}{M^{D-3}}\left[\lambda S H_u H_d + \frac{h}{4} S \hat{\mathcal{W}}^\alpha \hat{\mathcal{W}}_\alpha\right]_F + \text{h.c.} \\ & + \frac{1}{M^{D-3}}\left[S\left(a\hat{H}_u^\dagger \hat{H}_d^\dagger + b_u \hat{H}_u^\dagger \hat{H}_u + b_d \hat{H}_d^\dagger \hat{H}_d\right) + \text{h.c.}\right]_D \\ & + \frac{1}{M^{D-2}}\left[S^\dagger S\left(c_u \hat{H}_u^\dagger \hat{H}_u + c_d \hat{H}_d^\dagger \hat{H}_d + \left(d\hat{H}_u \hat{H}_d + \text{h.c.}\right)\right)\right]_D + \ldots\,,\end{aligned} \tag{4.48}$$

where h, a, $b_{u,d}$, $c_{u,d}$ and d are dimensionless couplings. The dots refer to terms containing three or more source fields interacting with MSSM bulk fields, terms containing only source fields and any interactions unobservable on the MSSM

Operator	4 dim	5 dim	d dim	
$[\delta^D(y)]$	4	5	d	Delta function
$[\theta]$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	Superspace coordinate
$[d\theta] = [\frac{d}{d\theta}]$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	Superspace coordinate differential
$[\partial]$	1	1	1	Partial derivative
$[\phi]$	1	$\frac{3}{2}$	$\frac{d-2}{2}$	Chiral superfield, scalar field
$[V]$	0	$\frac{1}{2}$	$\frac{d-4}{2}$	Vector superfield
$[\psi] = [\lambda]$	$\frac{3}{2}$	2	$\frac{d-1}{2}$	Spin $\frac{1}{2}$ Fermion field
$[F] = [D]$	2	$\frac{5}{2}$	$\frac{d}{2}$	Auxiliary fields
$[\mathcal{W}]$	$\frac{3}{2}$	2	$\frac{d-1}{2}$	Superfield strength
$[g_D^2]$	0	-1	$4-d$	D-dimensional gauge coupling

Table 4.2: Mass dimensions of fields, couplings and operators in 4, 5 and d dimensions.

brane. The superpotential for this framework has at least one more term compared to that in eq. (4.22), and is now

$$W = \overbrace{\bar{u}y_u Q H_u - \bar{d}y_d Q H_d - \bar{e}y_e L H_d + \mu H_u H_d}^{W_{\text{MSSM}}} + \overbrace{\lambda S H_u H_d}^{W_S}, \tag{4.49}$$

to account for the possible holomorphic interaction between the Higgs and source fields. Again, the spatial separation between the MSSM and source branes prevents direct interactions between MSSM matter fields and S, so direct trilinear couplings from terms like $S\bar{u}\hat{H}_u Q$ and sfermion soft masses from terms like $SQ^\dagger Q$ are forbidden.

Since interactions between the bulk fields and S are non-renormalisable in 4 dimensions, $\mathscr{L}_D$ describes an effective theory valid up to some fundamental scale M. To obtain the 4-dimensional effective field theory valid below the compactification scale, the higher dimensions are integrated over and only the zero modes of the bulk fields, which are constant in the extra dimensions, are kept. Terms in the Lagrangian coupling bulk fields to brane fields come with a delta function, see eq. (4.45), so after integrating over the extra dimensions, the volume factors in front of these terms cancel. On the other hand, terms involving only bulk fields do not come with this delta function, thus the integration yields a volume factor V_{D-4} in the kinetic terms of the bulk fields. For the effective theory to contain only fields with canonical kinetic terms in 4 dimensions, bulk fields are scaled as $\Phi \equiv \sqrt{V_{D-4}}\hat{\Phi}$. Hence, the part of the effective 4-dimensional Lagrangian describing direct interactions between MSSM fields and S is

$$\begin{aligned}\mathscr{L}_{D=4} \supset \frac{1}{V_{D-4}} \Bigg\{ & \frac{1}{M^{D-3}} \left[W_S + \frac{h}{4} S \mathcal{W}^\alpha \mathcal{W}_\alpha \right]_F + \text{h.c.} \\ & + \frac{1}{M^{D-3}} \left[S \left(a H_u^\dagger H_d^\dagger + b_u H_u^\dagger H_u + b_d H_d^\dagger H_d \right) + \text{h.c.} \right]_D \\ & + \frac{1}{M^{D-2}} \left[S^\dagger S \left(c_u H_u^\dagger H_u + c_d H_d^\dagger H_d + \left(d H_u H_d + \text{h.c.} \right) \right) \right]_D \Bigg\}. \end{aligned} \tag{4.50}$$

4.5.3 Calculating the soft terms

Deriving the soft terms and thus seeing the direct SUSY breaking effect of the F-term of the source field S, is quite straightforward. Calculations are outlined in the following, and details can be found in appendix A. The first term

in eq. (4.50) does not contribute to the present discussion, but the interested reader can see what it does in appendix A. Moving on to the second term, that gives rise to the gaugino mass $m_{1/2}$ when F_S acquires a VEV. This follows from eq. (4.23), keeping only the terms with the gaugino fields λ_α. Taking the F-term yields

$$\begin{aligned}\frac{1}{V_{D-4}}\frac{1}{M^{D-3}}\frac{h}{4}\left[S\mathcal{W}^\alpha\mathcal{W}_\alpha\right]_F + \text{h.c.} &\supset \frac{M_c^{D-4}}{M^{D-3}}\frac{h}{4}\int d^2\theta\left(\langle F_S\rangle\lambda^\alpha\lambda_\alpha + \text{h.c.}\right)\theta\theta \\ &= \frac{M_c^{D-4}}{M^{D-3}}\frac{h}{4}\langle F_S\rangle\lambda^\alpha\lambda_\alpha + \text{h.c.} \\ &= \left(\frac{M_c}{M}\right)^{D-4}\frac{h\langle F_S\rangle}{2M}\lambda^\alpha\lambda_\alpha\,,\end{aligned} \tag{4.51}$$

so the gaugino mass is

$$m_{1/2} = \left(\frac{M_c}{M}\right)^{D-4}\frac{h\langle F_S\rangle}{2M}\,. \tag{4.52}$$

The next term in eq. (4.50) gives rise to

$$\begin{aligned}\frac{1}{V_{D-4}M^{D-3}}\left[aS^\dagger H_u H_d + \text{h.c.}\right]_D &= \frac{M_c^{D-4}}{M^{D-3}}a\int d^2\theta d^2\bar\theta S^\dagger H_u H_d + \text{h.c.} \\ &\supset \left(\frac{M_c}{M}\right)^{D-4}\frac{a\langle F_S\rangle^\dagger}{M}\int d^2\theta H_u H_d + \text{h.c.} \\ &= \left(\frac{M_c}{M}\right)^{D-4}\frac{a\langle F_S\rangle^\dagger}{M}\left[H_u H_d\right]_F + \text{h.c.}\,,\end{aligned} \tag{4.53}$$

which, upon comparison with eq. (4.22), is clearly a contribution to the μ term. The equation of motion for the Higgs auxiliary field is solved in eq. (A.9), and repeated here for convenience

$$F_{H_{u,d}} = -b_{u,d}\left(\frac{M_c}{M}\right)^{D-4}\frac{\langle F_S\rangle}{M}\phi_{H_{u,d}}\,. \tag{4.54}$$

Inserted into the third and fourth terms and combined with the fifth and sixth terms of eq. (4.50), this yields the Higgs soft masses

$$\begin{aligned}(m_{H_u}^2, m_{H_d}^2) &= \langle F_S\rangle^\dagger\langle F_S\rangle\left\{\frac{(b_u^2, b_d^2)}{(V_{D-4}M^{D-3})^2} + \frac{(c_u, c_d)}{V_{D-4}M^{D-2}}\right\} \\ &= \frac{\langle F_S\rangle^\dagger\langle F_S\rangle}{M^2}\left\{\left(\frac{M_c^{D-4}}{M^{D-4}}\right)^2(b_u^2, b_d^2) + \left(\frac{M_c^{D-4}}{M^{D-4}}\right)(c_u, c_d)\right\}.\end{aligned} \tag{4.55}$$

The last term yields

$$\frac{1}{V_{D-4}M^{D-2}}\left[S^\dagger S\left(dH_u H_d + \text{h.c.}\right)\right]_D = \frac{\langle F_S\rangle^\dagger\langle F_S\rangle}{M^2}\frac{M_c^{D-4}}{M^{D-4}}d\left(\phi_{H_u}\phi_{H_d} + \phi_{H_u}^\dagger\phi_{H_d}^\dagger\right), \tag{4.56}$$

and compared with eq. (4.30), this contributes to the $B\mu$ term.

Directly substituting F_{H_u} and F_{H_d} from eq. (4.54) in the Lagrangian gives rise to trilinear scalar couplings proportional to the VEV of the SUSY-breaking field and the Yukawa matrices,

$$\mathscr{L}_{\text{trilinear}} = \left(\frac{M_c}{M}\right)^{D-4}\frac{\langle F_S\rangle}{M}\left(-b_u\phi_{\bar u}y_u\phi_{H_u}\phi_Q + b_d\phi_{\bar d}y_d\phi_{H_d}\phi_Q + b_d\phi_{\bar e}y_e\phi_{H_d}\phi_L + \text{h.c.}\right). \tag{4.57}$$

Defining

$$a_u = A_{u0}\,y_u \quad , \quad a_d = A_{d0}\,y_d \quad , \quad a_e = A_{d0}\,y_e \tag{4.58}$$

with

$$A_{u0} = \left(\frac{M_c}{M}\right)^{D-4}\frac{\langle F_S\rangle}{M}b_u \quad , \quad A_{d0} = \left(\frac{M_c}{M}\right)^{D-4}\frac{\langle F_S\rangle}{M}b_d\,. \tag{4.59}$$

This is a very interesting result, and forms the basis for the main result in *Trilinear-augmented gaugino mediation* [1], as these trilinear terms were previously assumed to vanish in the gaugino mediation model.

Chapter 5

Machine learning for classification

In the field of computer science, there is a method of data analysis which allows computers to learn without being explicitly programmed: Algorithms which use data to adjust their parameters[1] are collectively referred to as machine learning algorithms and constitute a branch of artificial intelligence. These can be very powerful tools for data analysis, especially in cases where there is much data available, but its underlying structure is too complex or elusive to model or describe analytically — as is often the case in particle physics.

There are many types of machine learning algorithms, each with their drawbacks and advantages. There is no single algorithm which outperforms the others in all areas, so the choice of algorithm depends on the problem at hand, number of inputs, accuracy desired and resources available. One fundamental machine learning concept, and an algorithm most scientists are familiar with, is decision trees, which also have a long history at CERN. In their most basic format, these work by dividing datasets into regions with mostly the same properties — or, simply put, asking yes/no questions —, and hence make parallel, discrete decision boundaries, resulting in classification plateaus.

On the other end of the machine learning scale, is an algorithm with pretty much opposite properties from decision trees: A more advanced algorithm which requires much data to train, much tuning, is computationally expensive to train and whose decisions are not reducible to simple yes/no statements. These are the artificial neural networks — henceforth: neural networks — which have become increasingly popular over the past years, and, as shown in the following, can be useful for problems too elusive for the more basic machine learning algorithms, as is the case for the problem in *Signal mixture estimation for degenerate heavy Higgses using a deep neural network* [2].

5.1 Neural networks

Before delving into the topic of neural networks, let the following be said: A neural network does not present an easily understandable or transparent model in the way for instance a decision tree can be converted into a series of if-then-else-statements. A neural network is, to repeat the perhaps most common critique directed towards them from the particle physics community, a "black box", whose output depends on its internal state. How to interpet neural network output is discussed by use of examples later in the present chapter, after a brief introduction.

5.1.1 What they are

Neural network algorithms consist of nodes arranged in layers and connected in a systematic way, forming a network. Each node receives input and produces output, and the behaviour of a neural network is an emergent property, ultimately the result of the decisions made by its nodes. The nodes are connected via links, which serve to propagate the output of node i in layer $l-1$ to node j in layer l. Each link connected to layer l also has a weight W_{ij}^{l}, which is just a number determining the sign and strength of the connection, so the weights are the parameters of a neural network.

[1] So yes, linear regression is machine learning.

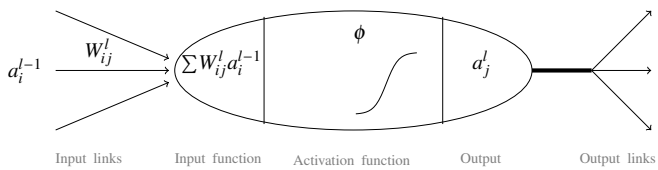


Figure 5.1: A simple sketch depicting the functioning of a node, i.e. a single neural network unit. The output a_j is based on the sum of all input the node receives, passed through the node's activation function ϕ. This node has a sigmoid shaped activation function as example.

Each node in a network calculates the sum of its input, and the output of node j in layer l, referred to as the activation and denoted a_j^l, is calculated using an activation function ϕ,

$$a_j^l = \phi\left(\sum_{i=0}^{n} W_{ij}^l a_i^{l-1}\right) \equiv \phi(\text{in}_j^l)\,. \tag{5.1}$$

A conceptual sketch of a node and its functionalities is presented in fig. 5.1.

Activation functions

Any nonlinearity of a neural network model comes from the activation function. This function must return ~ 1 for good and ~ 0 for bad inputs, where the notion of "good" and "bad" of course depends on the problem at hand. The simplest behaviour of a node would be binary, i.e. returning either 0 or 1. This can be achieved using a Heaviside step function as activation, see fig. 5.2a. The derivative of the Heaviside function is zero in the entire domain, except in one single point, where it diverges, i.e. the Dirac delta function. While intuitive, the Heaviside is not a good choice of activation function, since neural network training is an optimisation problem using gradient descent to update the weights W_{ij}^l — a method which can obviously not be used on a function which is either non-differentiable or has zero derivative everywhere.

One function which can fit the role is the sigmoid function, defined and differentiated as follows

$$\sigma(x) = \frac{1}{1+e^{-x}}\,, \qquad \frac{d\sigma(x)}{dx} = \sigma(x)\big(1-\sigma(x)\big)\,. \tag{5.2}$$

The sigmoid function stays approximately zero until the correct input is received, after which it increases smoothly and approaches an asymptote, see fig. 5.2b. However, while the sigmoid function is differentiable and serves its purpose, it comes with some challenges for the learning process: Its gradient is first zero, then it gradually increases until it reaches its maximum of $\frac{1}{4}$ at $x = 0$. After that it again decreases and slowly approaches zero. This small derivative in regions of good input causes what is commonly referred to as a "vanishing gradient problem"[2], and results in artificially low learning rates for the network.

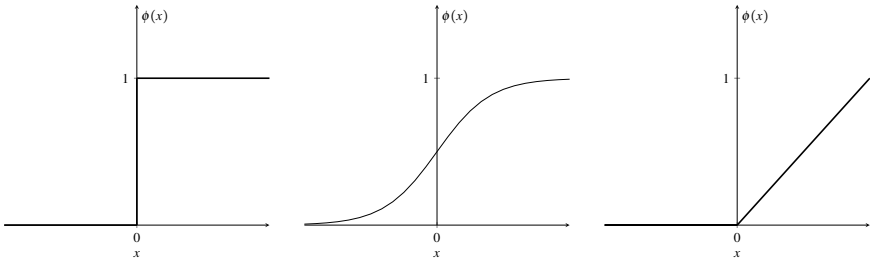
Similar to the sigmoid function is the softmax, which takes a vector of elements x_i and normalises it such that each element is mapped to the interval $[0,1]$, and $\sum_i x_i = 1$.The difference from the sigmoid function is that the softmax considers not only one input value, but also the remaining N elements inside the vector, as

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}\,, \tag{5.3}$$

which ensures the normalisation of the total output.[3]

[2] More precisely, when training a network with n stacked layers, each node contributes a factor σ', and thus the maximum gradient decreases exponentially with n.

[3] Note that the softmax is a generalisation of the sigmoid function, and the two have the same behaviour for $N = 2$, meaning binary classification.



(a) The Heaviside function. (b) The sigmoid function. (c) The Rectified Linear function.

This is useful in the context of neural networks, because the non-normalized values of the nodes in a layer are via this activation function mapped to a probability distribution. If this is done in the output layer of the network, see below, the output values from the softmax activation function can represent probabilistic predictions for class proportion of a data sample. Due to this, the softmax is a commonly used activation function in the output layer of classifier neural networks. However, the probabilistic interpretation comes with underlying assumptions, see section 5.2.2.

Introducing one last activation function, consider the rectified linear function, defined

$$R(x) = \max(0, x)\,. \tag{5.4}$$

Nodes using this activation function are also called Rectified Linear Units, abbreviated ReLU. The gradient of this function is either zero or one, avoiding the vanishing gradient problem. This is important especially since the trend for neural networks these days is to be increasingly deeper. Another benefit worth mentioning is sparsity: When $x < 0$, the sigmoid function returns a very small but non-zero value, meaning that the node remains active, though with little influence. The ReLU on the other hand, causes such nodes to "switch off", resulting in a sparse network — a network with fewer active nodes —, while the sigmoid activation function would create a denser network. Other variations of this activation function include the Leaky ReLU and softplus [51], which do not abruptly become zero, but die off slowly.

Network output

The two main types of neural networks are feed-forward networks and recurrent networks. In the former, signals travel in one direction only; from input to output. There are no loops, so the output of one layer cannot affect that same layer, the network represents a function of its input, and its entire internal state is represented by the weights W. The latter kind, on the other hand, allows signals to travel in both directions, so these networks can become quite complicated. They are dynamical and their internal state changes continuously, it may reach an equilibrium point, or it may oscillate or exhibit chaotic behaviour. Recurrent networks can display a short-term memory, which is not the case for feed-forward networks. The present work considers only feed-forward networks, but the interested reader can find a comprehensible explanation of recurrent networks in [52].

To demonstrate how the output of a feed-forward network is calculated, consider as an example a small network which is given an input vector $\mathbf{x} = (x_1, x_2)$. The number of nodes in the network's input layer corresponds to the size of the input vector. Subsequently, the network can in principle have any number hidden layers, of any size. In the present example, there is one hidden layer consisting of two nodes. The final layer is the output layer, which here has one node. This layout is depicted in fig. 5.3. Each node h in the hidden layer receives input from both nodes j in the input layer, and calculates its output as

$$a_h = \phi\left(W_{hj} x_j\right) = \phi\left(W_{h1} x_1 + W_{h2} x_2\right), \quad h = 1, 2\,. \tag{5.5}$$

The same holds for the output node, so the final output of this network is

$$a_5 = \phi\left(W_{35}\left(\phi\left(W_{13} x_1 + W_{23} x_2\right) + W_{45}\phi\left(W_{14} x_1 + W_{24} x_2\right)\right)\right), \tag{5.6}$$

which justifies the claim that the weights are the parameters of the network.

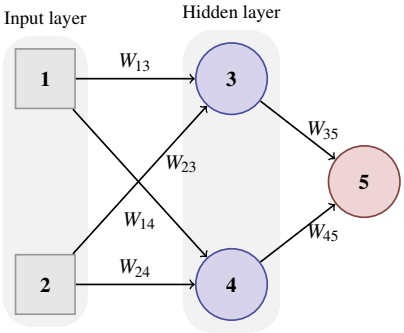


Figure 5.3: A small neural network with two input nodes, one hidden layer with two nodes, and one output node.

5.1.2 How they learn

The process of training a neural network involves adjusting the aforementioned weights W_{ij} to minimise some loss function[4]. Learning is thus an optimisation problem in weight space, spanned by all W_{ij}. Neural network training can be divided into two categories, namely supervised and unsupervised training. During supervised training, the target values of the data are known, which is not the case for unsupervised training. Only supervised training are considered in the present work.

During supervised training, data samples from a training set are passed through the network, and the network's output y is compared to the target value, denoted t. A loss function is used to numerically represent how well the network performed, and the weights of the network are updated accordingly.

A simple loss function is the mean squared error,

$$\mathrm{L}_{\mathrm{MSE}} = \frac{1}{2}(t-y)^2, \tag{5.7}$$

where the factor half is present only to cancel the factor two which appears upon differentiation. The term inside the parenthesis is often referred to as the residual vector. A popular, and slightly more advanced, loss function, which takes the uncertainty of the prediction into account, is cross entropy. It is defined for C classes as

$$\begin{aligned} \mathrm{L}_{\mathrm{CE}} &= -\sum_{i}^{C} t_i \log y_i \\ &= -\left(t_1 \log y_1 + (1-t_1)\log(1-y_1)\right), \quad C = 2, \end{aligned} \tag{5.8}$$

the second expression being valid for binary classification, where $t_2 = 1 - t_1$ and $y_2 = 1 - y_1$, in which case the loss function is also called log loss.

As the aim is to reduce the loss, the training process of a neural network should result in the weights being updated by moving in the direction of steepest descent along the loss surface — which has given the method its name, gradient descent[5].

This process can be demonstrated by calculating first the partial derivative of the loss with respect to the weights for

[4]Different authors use different terminology, but in all essence, loss, cost and error functions for neural networks express the same thing; a measure of deviance between the actual and desired output of the network.

[5]Stochastic gradient descent refers to gradient descent for one sample from the training set at a time, where the stochasticity arises from randomly shuffling the data samples between each epoch, in contrast to batch gradient descent, where the whole training dataset must be loaded into the memory.

each node,

$$\frac{\partial \mathrm{L}}{\partial W_{ij}} = \frac{\partial \mathrm{L}}{\partial a_j}\frac{\partial a_j}{\partial \mathrm{in}_j}\frac{\partial \mathrm{in}_j}{\partial W_{ij}}. \tag{5.9}$$

Expressing the input and output of each node as in eq. (5.1), the last term in the above equation is simply

$$\frac{\partial \mathrm{in}_j}{\partial W_{ij}} = \frac{\partial}{\partial W_{ji}} W_{kj} a_k = a_i, \tag{5.10}$$

which is just x_i if the node under consideration is an input node. The second term is

$$\frac{\partial a_j}{\partial \mathrm{in}_j} = \frac{\partial}{\partial \mathrm{in}_j}\phi(\mathrm{in}_j) = \phi'(\mathrm{in}_j), \tag{5.11}$$

which illustrates explicitly why the activation function ϕ must be differentiable.

The first term is simple if the node under consideration is the output node, $a_j = y$, e.g. for mean squared error loss $\frac{\partial}{\partial y}\mathrm{L}_{\mathrm{MSE}} = t - y$. If, on the other hand, a_j is a node somewhere inside the network,

$$\begin{aligned}\frac{\partial \mathrm{L}}{\partial a_j} &= \frac{\partial \mathrm{L}}{\partial \mathrm{in}_i}\frac{\partial \mathrm{in}_i}{\partial a_j} \\ &= \frac{\partial \mathrm{L}}{\partial a_i}\frac{\partial a_i}{\partial \mathrm{in}_i} W_{ji},\end{aligned} \tag{5.12}$$

a recursive expression which ends when an output node is reached. Finally, the impact a change in the weights W_{ij} has on the loss function is

$$\frac{\partial \mathrm{L}}{\partial W_{ij}} = \frac{\partial \mathrm{L}}{\partial a_j}\phi'(\mathrm{in}_j) a_i. \tag{5.13}$$

Since the aim is to reduce the error on target prediction, the weights should be updated as

$$W_{ij} \to W_{ij} - \alpha \frac{\partial \mathrm{L}}{\partial W_{ij}}. \tag{5.14}$$

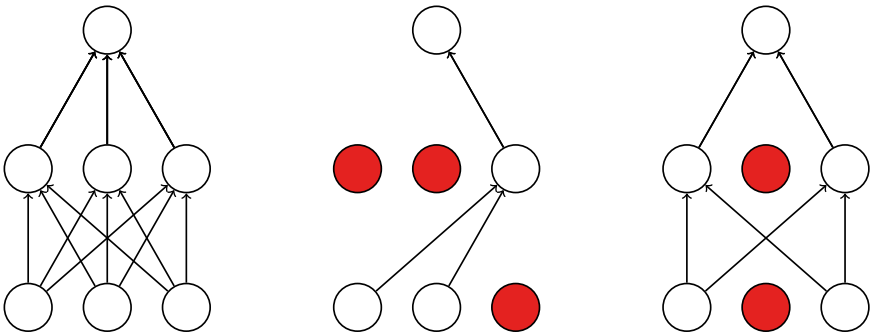
The parameter α is the learning rate, and represents the step size toward the minimum of the loss function. It is one of the many network hyperparameters. Each cycle through the entire training set is called an epoch, and the learning rate is maintained constant during each epoch when using stochastic gradient descent. How many epochs of training a network needs depends of course on the size of the network and the complexity of the problem, but a common stopping criterion for training is that the weights change very little. Still, there are several things to keep in mind when training neural networks:

Regularisation

Neural networks have a large number of free parameters, and with increasing network depth comes an increasing ability to make complex models. Complexity is nice, but as John von Neumann stated, [53]

> *With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.*

Increasing complexity in machine learning models can amount to simply internalising the training data into the model parameters, known as overtraining. An overtrained model performs well on training data, but fails to generalise on unseen data. There are several methods by which this can be avoided, and collectively, modifications to a learning algorithm which reduces the error on validation or test data but not that on the training data, are referred to as regularisation. Effectively, regularisation helps the algorithm to generalise beyond the training data.



(a) (b) (c)

Figure 5.4: A small neural network with (a) all nodes active and (b) and (c) some nodes dropped out, representing different network architectures.

One type of regularisation technique is based on weight norm penalties, and works by adding a penalty term to the loss function as

$$\mathrm{L} \to \mathrm{L} + \lambda f_p(W)\,, \tag{5.15}$$

where λ is the regularisation strength, and $f_p(W)$ grows larger with the norm of the weights W. If $\lambda = 0$ there is no regularisation, and if λ is very large, an increase in any of the weights is significantly penalised, resulting in the network trying to keep the weights as small as possible instead of finding weights which perform well. Finding a good value for λ is thus crucial, and may involve some trial and error. The intuitive interpretation of this is that large weights are penalised, while small weights are favoured; sharp peaks are disfavoured, while smooth behaviour is encouraged. The desire is that the network use all the input values, instead of assigning a high degree of importance to only some of them. The parameter λ is another example of a hyperparameter.

There are two common regularisation methods, and the difference between the two lies in the penalty term. One is referred to as L1 regularisation or Lasso regression, and adds the absolute sum of the weights to the loss function, i.e. $f_p(W) = \sum_i |W_i|$. The other type is L2 regularisation, or Ridge regression, which adds the absolute square of the weights to the loss function, i.e. $f_p(W) = \frac{1}{2}\sum_i W_i^2$. Note that the L2 regularisation scheme

$$\mathrm{L} \to \mathrm{L} + \frac{\lambda}{2}\sum_i W_i^2\,, \tag{5.16}$$

is equivalent to updating the weights as

$$W_k \to W_k - \alpha\frac{\partial \mathrm{L}}{\partial W_k} - \alpha\lambda W_k\,, \tag{5.17}$$

for gradient descent.

These two regularisation schemes have different advantages and drawbacks, and the correct one must be chosen for each problem. For instance, in the case of outliers, L2, which penalises large weights more sharply, adjusts the model to minimise the few outlier cases at the expense of the common data points. By contrast, L1 rather disregards outliers and is in this case more robust.

Another and more random regularisation technique assigns a probability p for activation to each node in the network, so that any node can at any point during training become inactive, see fig. 5.4 for an illustration. As such, the network is forced to ensure no information depends on one specific node, but rather spread it out over several nodes. Effectively, overfitting is prevented by combining many different network architectures. The dropout probability p is another hyperparameter.

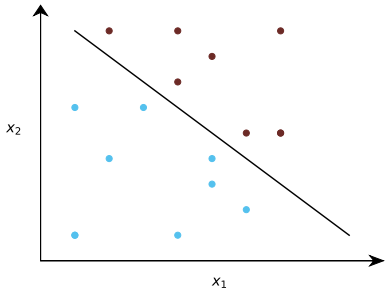


Figure 5.5: A simple model of a linear, two-dimensional classifier with input features x_1 and x_2. The coloured points represent instances from two different classes, and the decision surface is in this case a straight line.

Finally, a regularisation method which provides guidance for how many training epochs are sufficient for the learning algorithm to learn, but not overtrain: The dataset available for training is divided into into three subsets, and only the first subset used for the actual training, i.e. for adjusting the weights of the network. The second subset can then be used for validation during training, by calculating the loss on the validation set throughout. This is known as cross-validation, akin to the method described in section 3.3.2. Initially, the validation loss decreases along with the loss on the training set, but while the training loss keeps decreasing even when the network learns the training data, i.e. overtrains, the validation loss starts to increase as the network loses its ability to generalise. When the validation loss increases for some given number of iterations, the training stops. The weights corresponding to the minimum validation loss are treated as the best fit for the network for the current training set. The third subset can then be used as a blinded test to evaluate the performance.

5.1.3 What they learn

To summarise and generalise, neural network classifiers receive a dataset $\mathcal{D}$ containing input features $\mathbf{x}$, which can be mapped to an output variable y via some function f, i.e. $f(\mathbf{x}_i) = y_i$. The algorithm tries to learn, meaning numerically approximate, f so that y can be predicted for new $\mathbf{x}$.[6]

The space spanned by all the input features is called feature space, and the task of the algorithm is sometimes referred to as feature extraction. This is most easily pictured for the case of classifying data from two classes: The network finds a decision boundary, which is a hypersurface in feature space that separates the underlying vector space into one set for each class, see fig. 5.5 for a simple illustration. Mathematically speaking, each layer in a neural network creates a hyperplane serving as decision surface in the vector space spanned by the previous layer.

The classifier assigns all instances, represented by points in feature space, on one side of the decision boundary to one class, and all those on the other side to the other class. However, the decision boundary is not an inherent property of the classification problem itself, but of the learned classifier; the function f is numerically approximated, not found analytically. Given the nature of the learning process, the classifier is based entirely on data it has seen during training; a dataset henceforth denoted $\mathcal{D}_{\text{train}}$.

5.1.4 How well they learn

To start by introducing some useful terminology, the veracity of any classifier output can be divided into the four categories true or false and positive or negative, which reflect in the first term whether the classifier output matches

[6] In all essence, neural networks multiply matrices together and perform high dimensional curve-fitting, but don't tell anyone.

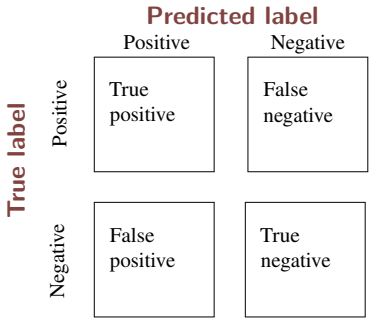


Table 5.1: The confusion matrix for a binary classifier, showing the four categories to which the network output can belong.

the target, and in the second term to which class the target belongs. This is summarised in table 5.1, which shows the so-called confusion matrix for a binary classifier. This can be used to define some terms to describe the quality of the classifier, meaning which mistakes it makes and the interpretation of these, see table 5.2.

A Bayesian interlude for intuition

In the Bayesian interpretation of probability, sensitivity and specificity are conditional probabilities. Denoting by $t_{P/N}$ a true positive or negative, and by $y_{P/N}$ a predicted positive or negative, and using the notation and rules given in section 3.1.2,

$$\begin{aligned} \text{Sensitivity} &= \frac{p(y_P, t_P)}{p(t_P)} = \frac{p(t_P|y_P)p(y_P)}{p(t_P)} = p(y_P|t_P) \\ \text{Specificity} &= \frac{p(y_N, t_N)}{p(t_N)} = \frac{p(t_N|y_N)p(y_N)}{p(t_N)} = p(y_N|t_N)\,. \end{aligned} \tag{5.18}$$

Expressing also the precision as a probability in the Bayesian sense yields

$$\text{Precision} = \frac{p(y_P, t_P)}{p(y_P)} = \frac{p(y_P|t_P)p(t_P)}{p(y_P)}\,, \tag{5.19}$$

which reveals something quite interesting: In the Bayesian interpretation, where the sensitivity $p(y_P|t_P)$ is a conditional probability, the class proportion $p(t_P)$, often called the prevalence, is the prior, and the resulting precision the posterior probability. The denominator is the overall probability for predicting the positive class, and as mentioned earlier, the classifier can only base its predictions on one set of information — its training data — rendering $p(y_P)$ an unsettling component of the statement.

The above expression can be expanded as

$$\begin{aligned} \frac{p(y_P|t_P)p(t_P)}{p(y_P)} &= \frac{p(y_P|t_P)p(t_P)}{p(y_P|t_P)p(t_P) + p(y_P|t_N)p(t_N)} \\ &= \frac{p(y_P|t_P)p(t_P)}{p(y_P|t_P)p(t_P) + p(y_P|t_N)\,(1 - p(t_P))}\,, \end{aligned} \tag{5.20}$$

which *only in the case of balanced classes*, i.e. $p(t_P) = p(t_N) = 0.5$, reduces to

$$\text{PPV}|_{t_P = t_N} = \frac{p(y_P|t_P)}{p(y_P|t_P) + p(y_P|t_N)}\,, \tag{5.21}$$

which is sometimes given as the positive predictive value, short PPV. However, balanced test sets are, as will become clear in the following, neither in general nor in particle physics a safe assumption.

Table 5.2: Terms describing different measures for the quality of a classifier. Here, total positive and negative refer to the total number of positive and negative samples inside the test set.

Term	Expression	Also known as	Explanation
True positive rate	$\equiv \frac{\sum \text{True positive}}{\sum \text{Total positive}}$	Sensitivity, Recall	Rate of positive samples correctly identified
False negative rate	$\equiv \frac{\sum \text{False negative}}{\sum \text{Total positive}}$		Miss rate, β in eq. (3.34)
False positive rate	$\equiv \frac{\sum \text{False positive}}{\sum \text{Total negative}}$		Rate of false alarm p(falsely reject null hypothesis)
True negative rate	$\equiv \frac{\sum \text{True negative}}{\sum \text{Total negative}}$	Specificity	Rate of negative samples correctly identified
Positive predictive value[1]	$\equiv \frac{\sum \text{True positive}}{\sum \text{Predicted positive}}$	Precision	Rate of positive class assignment correct
False discovery rate	$\equiv \frac{\sum \text{False positive}}{\sum \text{Predicted positive}}$		Rate of type I errors
False omission rate	$\equiv \frac{\sum \text{False negative}}{\sum \text{Predicted negative}}$		Rate of type II errors
Negative predictive value[1]	$\equiv \frac{\sum \text{True negative}}{\sum \text{Predicted negative}}$		Rate of negative class assignment correct

[1] This is the commonly used definition, although it is a bit ambiguous, see the discussion leading up to eq. (5.21).

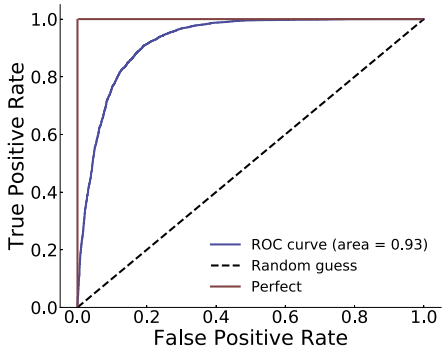


Figure 5.6: The receiver operating characteristic curve for a — very strong — classifier (blue), along with the theoretically best possible performance (red) and the result of random guessing (dashed black).

. . . and one on optimising

Evidently, precision and recall will often have an inverse relationship, since the process of maximising the number of true positives requires the number of predicted positives to increase. To illustrate the impact of maximising one at the expense of the other, consider a dataset containing many background events B and only a few signal events S. A classifier trained to identify signal events can in the two ends of the precision-recall spectrum behave in two ways: If it has high precision but low recall, it avoids misclassifying background as signal by returning only those signal events which have been classified with a very low uncertainty. The consequence is of course that many signal events are "lost", in the sense that they are over-carefully assigned to the background class. In the opposite case, with high recall but low precision, many events are classified as signal, but most of these will actually belong to the background class.

The ideal scenario is of course to maximise both recall and precision, but in the common case where one must be prioritised, the best course of action depends on the problem at hand. For instance, in searches for new physics, some false positives and ambulance-chasing papers on the arXiv is better than missing out on new physics.

The ROC curve and the area under it

There is a myriad of model evaluation tools available for different classifiers, but in the case of binary classification, the receiver operating characteristic, short ROC, curve is simple both conceptually and in use. Consider the false positive rate, FPR, versus the true positive rate, TPR. Since the classifier outputs a number between zero and one, a threshold is chosen such that samples above it are indicative of one class, and those below of the other. The FPR and TPR vary with this threshold: In the extrema, where the threshold is exactly zero or one, all samples are assigned to the negative or positive class, respectively. The ROC curve is created by evaluating the class predictions across a continuum of thresholds between zero and one. The curve must as argued start at $(0,0)$ and end up at $(1,1)$, but it can take almost any path between the two corners. When its gradient is large, the sensitivity increases at a larger rate than the specificity decreases. Such insight can be used to adjust the classification threshold in the interest defined by the problem at hand.

A perfect classifier which completely separates the two classes would have TPR= 1 and FPR= 0, i.e. both sensitivity and specificity equal to one. Graphically, this would correspond to a single step from $(0,0)$ to $(0,1)$ and then a constant value from $(0,1)$ to $(1,1)$. Conversely, a classifier which guesses the class of each sample at random would produce a diagonal ROC curve, see fig. 5.6 for an illustration.

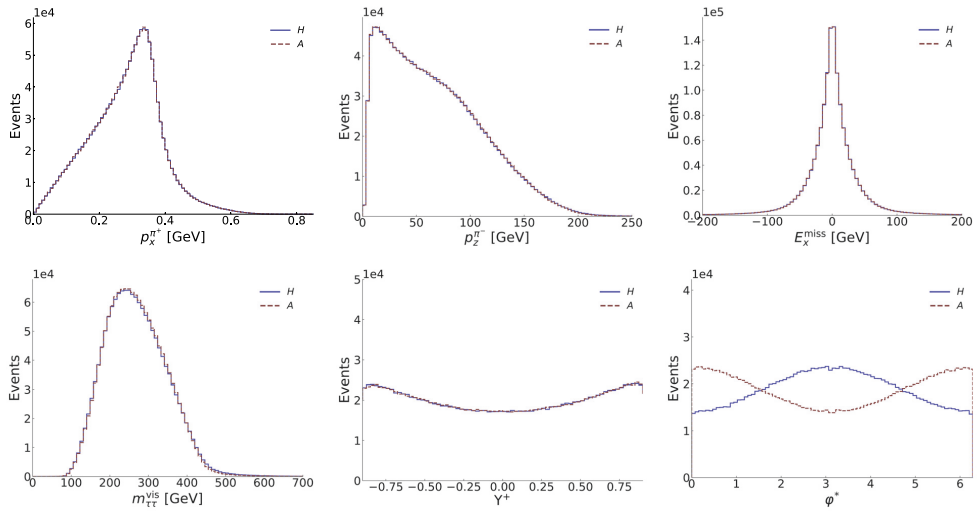


Figure 5.7: Some of the kinematic features from the decays of $H/A \to \tau^+\tau^- \to (\pi^+\pi^0\nu_\tau)(\pi^-\pi^0\nu_\tau)$. The top row shows low-level features, i.e. direct observables, while the bottom row shows high-level features, i.e. derived quantities from low-level features. The plots are produced using the same data as in [2].

5.2 Neural network predictions

To put the previous discussion into use, consider a concrete case, and let it be difficult. The scenario studied in [2] is high-dimensional and the feature distributions severely overlapping, making it perfect for illustrating some of the challenges related to interpreting the output of a machine learning algorithm, which is of course a crucial part of using it.

5.2.1 Mixture parameter estimation in a scenario with overlapping features

The results in this section are taken from [2], where the scenario is any generic two-Higgs doublet model with production of mass degenerate heavy Higgses H and A, which can consequently only be distinguished by means of their $\mathcal{CP}$ charges. A state's $\mathcal{CP}$ charge affects the angular distributions of its decay products, but $\mathcal{CP}$ cannot be sharply determined in single events.

Some of the low-level features, meaning direct observables such as visible decay products, for this problem are shown in the top row of fig. 5.7. The high degree of overlap between the H and A distributions is representative for all low-level features in this problem, indicating that the task of separating the classes is not an easy one. In fact, separation of the classes depends on high-level correlations between the features. Such correlations can in principle be found by a neural network, and sometimes also by physicists. These are called high-level features, and a prominent example in the current discussion is the $\mathcal{CP}$-sensitive parameter φ^*, which represents an angle between decay planes, defined in eq. (6) in [2].

Comparison between a one-dimensional likelihood fit to this φ^* variable and the result obtained by a neural network, is the main result in [2]. The reason why the neural network method outperforms the one-dimensional method, is of course that the network has access to all observables and derived quantities, and can use these to find correlations within the data, not contained in the φ^* variable, *and* that there are additional correlations to be found.

In theory, all the available information could be used by constructing an n-dimensional likelihood for all the available features, and then performing a multi-dimensional maximum likelihood fit to data. Unfortunately, this is not feasible in practice: Generating sufficient data for mapping the likelihood function in many dimensions is impossible with the limits set by the computing power available today— in addition to the fact that there is no known criterium by which to identify the global maximum of a non-convex function, which the likelihood could be. This is exactly the aforementioned curse of dimensionality, which would be circumvented if the information from all the features could be projected onto a fewer, or even a single variable.

Enter machine learning! A classifier can be trained on data containing low- and high-level features, and its output, or prediction, have dimensionality as low as one.

5.2.2 Prior and prejudice

A neural network with $(500, 1000, 100)$ nodes in its three hidden layers using a ReLU activation function, and whose output layer has a softmax activation function, mapping the output to the range $(0, 1)$, is trained on a dataset $\mathcal{D}_{\text{train}}$. This dataset contains H and A events with mixture parameter, defined as

$$\alpha \equiv \frac{n_A}{n_A + n_H}\,, \tag{5.22}$$

of $\alpha_{\text{train}} = 0.5$. The network is then used to predict on three different datasets $\mathcal{D}_{\text{test}}$ with mixture parameters $\alpha_{\text{test}} = 0.5, 0.7, 0.9$.

Since the network output y_i for each test sample is a number between zero and one, the interpretation of y_i as the *probability* of event i belonging to a class seems to suggest itself. If this is the case, the mixture parameter of the test set is simply

$$\hat{\alpha}_{\text{test}} = \frac{\sum_i y_i}{n_{\text{test}}}\,. \tag{5.23}$$

Using this interpretation, the distributions shown in fig. 5.8a are generated. For comparison, predictions from a one-dimensional maximum likelihood fit to φ^* are shown in fig. 5.8b. The expected result was of course that the neural network prediction would be as accurate yet more precise than the one-dimensional maximum likelihood fit. What instead happened, is that the neural network result is very biased towards the mixture parameter of the training set, which is a direct result of the interpretation of y_i as a probability. This interpretation implies the assumption that the test data is drawn from the same distribution as the training data, as will be shown in the following. Note that classification problems where labeled training data can only be obtained through Monte Carlo generation, as is often the case in particle physics, are particularly sensitive to this effect, here with the added complication that the class proportion in the test data is exactly the quantity of interest.

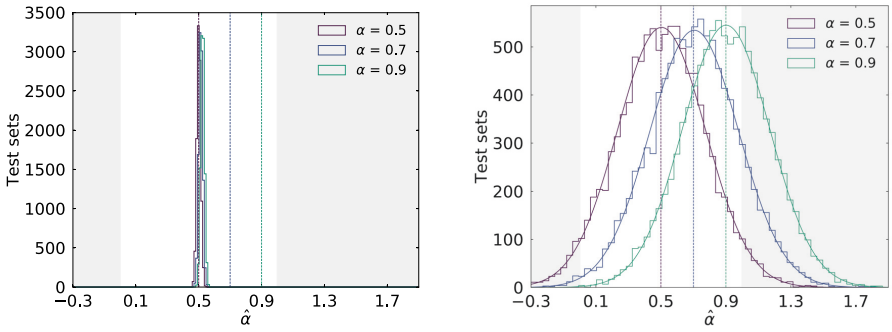
In machine learning communities, faulty results caused by test data drawn from a different distribution than the training data is sometimes referred to as data mismatch. It is merely the Bayesian effect of prior probability shift, which arises as the probabilistic interpretation amounts to assigning a probability to both the data *and* the model:

$$y_i \sim p(A_i|\mathbf{x}_i, \mathcal{D}_{\text{train}}) \tag{5.24}$$

is the probability that event i in the test set belongs to class A, given its features $\mathbf{x}_i$ and everything the network knows about the world, namely the training data $\mathcal{D}_{\text{train}}$. To see explicitly what this entails, use Bayes' law to expand the above probability,

$$p(A_i|\mathbf{x}_i, \mathcal{D}_{\text{train}}) = \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})}\; p(A_i|\mathcal{D}_{\text{train}})\,. \tag{5.25}$$

This can be marginalised over the in principle unknown mixture parameters used to generate the training set (α_{train}) and



(a) (b)

Figure 5.8: Estimating the mixture parameter of ten thousand test sets containing a hundred samples each, for $\alpha_{\text{test}} = 0.5, 0.7, 0.9$, using (a) directly the output of a neural network trained on a dataset with $\alpha_{\text{train}} = 0.5$, and (b) the one-dimensional maximum likelihood fit to φ^*. The plots were produced for [2].

test set (α_{test}):

$$\begin{aligned} p(A_i|\mathbf{x}_i, \mathcal{D}_{\text{train}}) &= \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})} \iint p(A_i, \alpha_{\text{test}}, \alpha_{\text{train}}|\mathcal{D}_{\text{train}})\, d\alpha_{\text{test}}\, d\alpha_{\text{train}} \\ &= \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})} \iint p(A_i|\alpha_{\text{test}})\, p(\alpha_{\text{test}}|\alpha_{\text{train}})\, p(\alpha_{\text{train}}|\mathcal{D}_{\text{train}})\, d\alpha_{\text{test}}\, d\alpha_{\text{train}} \\ &= \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})} \iint \alpha_{\text{test}}\, p(\alpha_{\text{test}}|\alpha_{\text{train}})\, p(\alpha_{\text{train}}|\mathcal{D}_{\text{train}})\, d\alpha_{\text{test}}\, d\alpha_{\text{train}}. \end{aligned} \tag{5.26}$$

Here,

- $p(A_i\,|\,\alpha_{\text{test}}, \alpha_{\text{train}}, \mathcal{D}_{\text{train}}) = p(A_i\,|\,\alpha_{\text{test}}) = \alpha_{\text{test}}$ because the probability of drawing an A_i event from the test set only depends on the distribution α_{test} in the test set; not the distribution nor the kinematics in the train set.
- $p(\alpha_{\text{test}}\,|\,\alpha_{\text{train}}, \mathcal{D}_{\text{train}}) = p(\alpha_{\text{test}}\,|\,\alpha_{\text{train}})$ because the (network) assumption for the test set distribution can be based on the train set distribution (or something completely different), but not on the kinematics in the train set. Once α_{train} is known, $\mathcal{D}_{\text{train}}$ contains no other additional information than the kinematics of the particular events contained therein. Under the assumption that the test and train sets have equal distributions, this collapses to $p(\alpha_{\text{test}}\,|\,\alpha_{\text{train}}) = \delta(\alpha_{\text{test}} - \alpha_{\text{train}})$.

These arguments, after the delta function collapses the double integral, yield

$$\begin{aligned} y_i \approx p(A_i|\mathbf{x}_i, \mathcal{D}_{\text{train}}) &= \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})} \int \alpha_{\text{train}}\, p(\alpha_{\text{train}}|\mathcal{D}_{\text{train}})\, d\alpha_{\text{train}} \\ &= \frac{p(\mathbf{x}_i|A_i, \mathcal{D}_{\text{train}})}{p(\mathbf{x}_i|\mathcal{D}_{\text{train}})} \frac{n_{A,\text{train}}}{n_{\text{train}}}, \end{aligned} \tag{5.27}$$

where the integral is simply the expectation value for α_{train} given the training dataset. The first factor represents the kinematic distributions for H and A, which according to the earlier feature analysis is close to unity for most $\mathbf{x}_i$.

This is a demonstration of how the prior in Bayesian analysis can strongly affect the posterior if the data is not sufficiently strong. Whenever the influence of the prior dominates that of the data, the results must be interpreted with care, regardless of the choice of interpretation. It is obvious from the above derivation that a different choice of prior could potentially completely change the conclusion.

5.2.3 The likelihood fits again

Recall that the most powerful test statistic for a simple hypothesis test is, according to the Neyman-Pearson lemma, see section 3.4.2, the likelihood ratio. In the current notation,

$$\lambda(\mathcal{D};\alpha,\alpha') = \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(\mathbf{x}|\alpha)}{p(\mathbf{x}|\alpha')}\,, \tag{5.28}$$

where $\mathcal{D}$ denotes the dataset, $p(\mathbf{x}|\alpha)$ the model and α any model parameter — not necessarily a mixture parameter. In terms of the input observables $\mathbf{x}$, the Higgs mixture model is

$$p(\mathbf{x}|\alpha) = \alpha p_A(\mathbf{x}) + (1-\alpha)p_H(\mathbf{x})\,. \tag{5.29}$$

It can be shown [54, 55], that eq. (5.28) is *equivalent* to the likelihood ratio

$$\lambda_s(\mathcal{D};\alpha,\alpha') = \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(s(\mathbf{x};\alpha,\alpha')|\alpha)}{p(s(\mathbf{x};\alpha,\alpha')|\alpha')}\,, \tag{5.30}$$

based on the univariate density $p(s(\mathbf{x};\alpha,\alpha')|\alpha)$, provided that the transformation $s(\mathbf{x};\alpha,\alpha')$ is a strictly monotonic function of the density ratio $p(\mathbf{x}|\alpha)/p(\mathbf{x}|\alpha')$.

This is not a very strict requirement; in many cases one need merely look at the pdfs to realise that their ratio is monotonic, but it has been formalized [56] which loss functions result in neural networks that satisfy the requirement. Cross-entropy, the one used in [2], is among these. Replacing thus the general univariate density s by the network output y, the maximum likelihood estimator for the total pdf is now

$$\begin{aligned}
\hat{\alpha} &= \arg\max_{\alpha} \prod_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x}|\alpha) \\
&= \arg\max_{\alpha} \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(\mathbf{x}|\alpha)}{p(\mathbf{x}|\alpha')} \\
&= \arg\max_{\alpha} \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(y(\mathbf{x})|\alpha)}{p(y(\mathbf{x})|\alpha')} \\
&= \arg\max_{\alpha} \prod_{\mathbf{x}\in\mathcal{D}} p(y(\mathbf{x})|\alpha) \\
&= \arg\max_{\alpha} \prod_{\mathbf{x}\in\mathcal{D}} \Big[\alpha p_A(y(\mathbf{x})) + (1-\alpha)p_H(y(\mathbf{x}))\Big],
\end{aligned} \tag{5.31}$$

having used in the fourth step that the network output is (unfortunately!) independent of the true, unknown mixture parameter α'.

This very handy result implies that the value of α which maximises a fit of the total n-dimensional likelihood to data, is the same as that which maximises the neural network output — provided of course that the network does a good job of approximating the univariate density, which depends on its architecture and the training process. An unbinned kernel density estimation is done to generate the pdfs p_A and p_H from the network output, which are then fitted to test data. This produces the main result in [2].

As a concluding remark, note that the two methods described above, of which one was severely prior dependent and hence produced useless results, while the other was prior independent and furthermore outperformed the conventional method, were both based on the same neural network. No further feature engineering or network output calibration was done, and the classifier itself performed just as well or as poorly in the two cases: The difference consisted in how the network output was interpreted and used.

5.3 Boosted learning

Each layer in a neural network introduces a new vector-space into which the previous layers' representation of the features is projected numerically. As discussed, this makes deep neural networks strong candidates for extracting

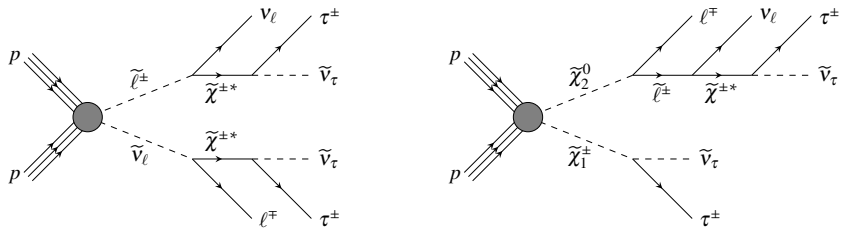


Figure 5.9: Examples of the processes referred to referred to in the text as (a) slepton and (b) electroweakino production.

high-level correlations from data, but it does not mean that "stack more layers" is the best solution to any data driven problem. When computational power and CPU time are not strictly limited, it can of course be tempting to use deep learning even on problems which do not strictly speaking call for it. Beware though, that the more free parameters a model has, the more data is required to tune it. Because of this, training neural networks inherently depends on large datasets, and the process of tuning and training them is on the more time-consuming side of the machine learning scale. Consider for illustration one last scenario, which forms the basis for the project on *Enabling sneutrino detection in weak signal scenarios using machine learning methods* [6]:

5.3.1 Sneutrino detection

Remaining in a collider setting, consider a supersymmetric model featuring the sneutrino as the lightest observable supersymmetric particle. The aim is to separate the supersymmetric signal from the Standard Model background, i.e. again a mixture parameter estimation task, this time defining

$$\alpha \equiv \frac{n_s}{n_s + n_b}, \tag{5.32}$$

where s abbreviates signal or supersymmetry, and b background. Since the task is now signal detection, the reader may already anticipate that this time, the mixture parameter could be very small.

Divide the interesting supersymmetric production processes within the model at hand into three categories, according to which sparticle pairs are produced in the hard collision:

- Anything involving squarks and/or gluinos
- A slepton pair, i.e. any $\widetilde{l}$ and/or $\widetilde{\nu}_l$
- A neutralino and/or chargino pair, $\chi^{0,\pm}\chi^{0,\pm}$

The first of the three stands out, as it is easier to separate from the Standard Model background than the other two, thanks to its accompanying jets. Hence, the two latter, which can give rise to very similar collider signatures, are most interesting in the present context. Examples of the slepton pair production and electroweakino production processes are shown in fig. 5.9, and the detector signature of interest is that containing two taus and a muon.

A selection of the feature distributions from a simulation with these two processes is shown in fig. 5.10. It is clear that the features are so overlapping that a cut-and-count analysis will have trouble separating the classes, but in comparison to fig. 5.7, the features are not *extremely* overlapping. The question begs itself: Is it necessary to go deep?

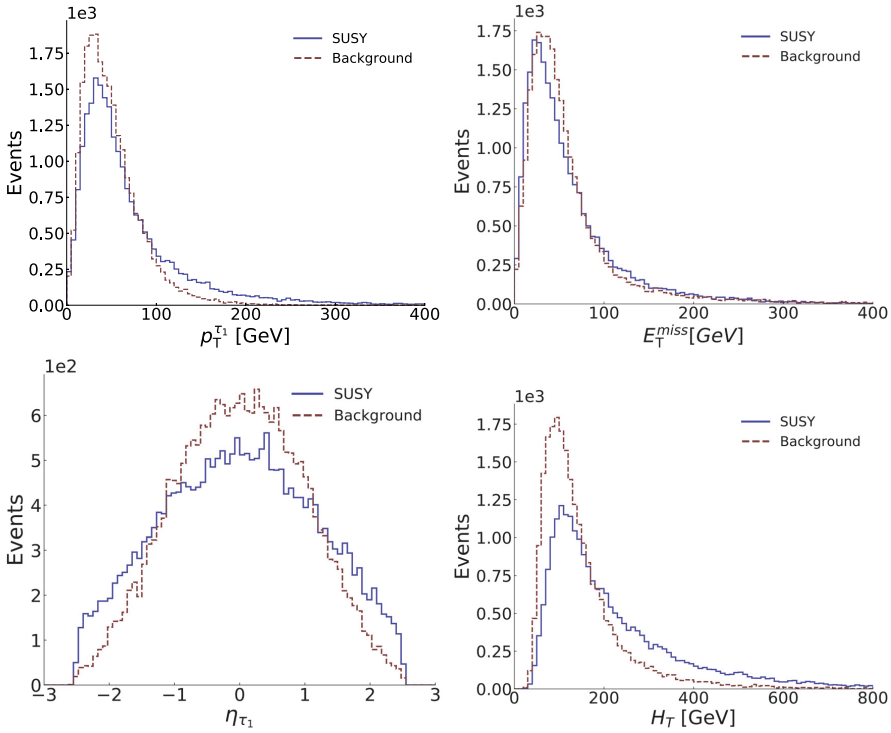


Figure 5.10: Some of the kinematic features from the supersymmetric signal coming from the slepton pair production and electroweakino production channels (solid blue) and background (dashed red).

5.3.2 Boosted decision trees

Going back to a classification algorithm mentioned earlier, the decision trees: As described, their internal yes/no structure makes it impossible for them to learn multicollinearity[7] in data.

Fortunately, these transparent machine learning algorithms are continually improved, and a machine learning discussion would not be complete without them. One important concept is boosting. This is a meta-model concept, which combines many weak prediction models — decision trees — in the form of an ensemble, into one strong. Boosted decision trees are frequently used in the ATLAS, CMS and LHCb collaborations.

Gradient boosting [57] is an extension of the previously discussed learning process — of moving in the direction of the steepest descending gradient —, where the parameters of a model were adjusted based on the change in the loss function. Gradient boosting models also use the gradient for optimization, hence the name, but the optimization occurs not on the parameters of the underlying models themselves, but on the prediction of the composite model. This means that gradient boosting models are trained on two levels; on the weaker, underlying models, and on the composite model. The overall training of the composite model performs gradient descent using the residual vector on the combined model output, and this gives rise to the strength of boosted algorithms, which is impressive in its simplicity: The individual trees merely have to do slightly better than random guessing, and still a large combination of them produces strong

[7] Variables are multicollinear when they combine to yield another, independent variable. Two variables are perfectly collinear if they are a linear combination of each other. Here, a decision tree will have to pick only one, and will mostly go for the one with the strongest correlation.

results.

Just like neural networks, boosted decision trees have hyperparameters. These range from the intuitive ones, like the number of underlying trees and the maximum depths of these, to how many data samples must be considered before splitting an internal node, or leaf. And just as for neural networks, the correct tuning of the hyperparameters is important for optimal performance of the model. Still, the list of boosted decision tree hyperparameters does not even come close to that of neural networks.

The boosted trees are also more efficiently trained, given that they have less free parameters to adjust, and focus, by construction, on the most important features to the extent that features with little or no discriminatory power are ignored completely.

While no boosted decision tree algorithms were successful in the extremely overlapping case of [2], this does not have to be the case for the signal vs. background problem described in the present section. To illustrate the performance of boosted decision trees, consider the classification problem at hand, with the added complication of using a scarce dataset. This can be the case e.g. when simulating certain processes takes a long time, or when data cannot be simulated and must be collected.

Using a balanced training dataset of fifty thousand events and a test dataset of ten thousand events, a simple neural network achieves a ROC AUC of ~ 0.9 in around fifteen minutes. By comparison, the `XGBoost` [58] — short for eXtreme Gradient Boost — algorithm achieves a ROC AUC of ~ 0.93 after less than two minutes of training, and `CatBoost` [59][8] the same ROC AUC in just under three minutes. Based on the previous discussion, this result is not surprising.

Note that these numbers can of course not be taken as a definite conclusion of how well either algorithm can perform on this particular problem, and they are certainly strongly dependent upon the architecture on which they run. However, they do serve for illustration: Other machine learning algorithms, such as boosted decision trees, can compete with — and for certain problems also outperform — neural networks, and are in general more transparent and computationally light-weight. Especially since neural networks are so opaque, a good practice is to always use another classification algorithm as benchmark.

5.3.3 Mixture parameter estimation for sneutrino detection

An `XGBoost` algorithm is trained on a dataset containing equal amounts of Standard Model background and signal from the slepton and electroweakino production processes described earlier.[9]

The algorithm is then used to create templates for the signal and background class respectively. This is done by feeding the classifier one dataset for each class, containing ten thousand previously unseen events, and using kernel density estimation to transform the classifier output for each class into an analytical function. A technical challenge encountered here, which was not the case in the heavy Higgs classification, is that this classifier has an easier job separating the signals, causing more of the predictions to fall closer to 0 or 1. The template fit to the network output must thus be robust on the edges, as described in section 3.3.3.

Next, the classifier is used to predict on test sets with different sizes and mixture parameters, and the classifier output used as data in a maximum likelihood fit for the aforementioned templates. Fits to mixture parameters of $\alpha_{\text{true}} = 0.1, 0.2$ with test set sizes of 200 and 100 events respectively are shown in fig. 5.11. The histograms indicate the classifier's predictions for the test datasets in each case, and the templates for each class have been scaled according to the best-fit value for the signal mixture parameter, $\hat{\alpha}$.

The above described method is shown more systematically in fig. 5.12, again for different test set sizes and mixture parameters, where the gray shades indicate physically sensible regions for the mixture parameter. As expected, larger test sets correspond to narrower parabolas, i.e. stronger estimates of the mixture parameter. This is in agreement with e.g. fig. 3.4b, and the potential discovery significance can now be accessed using the relation in eq. (3.22), indicated by the $n\sigma$ bands in the plots.

[8] The name presumably coming either from familiar examples of image classification into categories of cats and dogs, or the logo, which resembles a cat's paw.

[9] The internal distribution in the background dataset is in accordance with the Standard Model prediction.

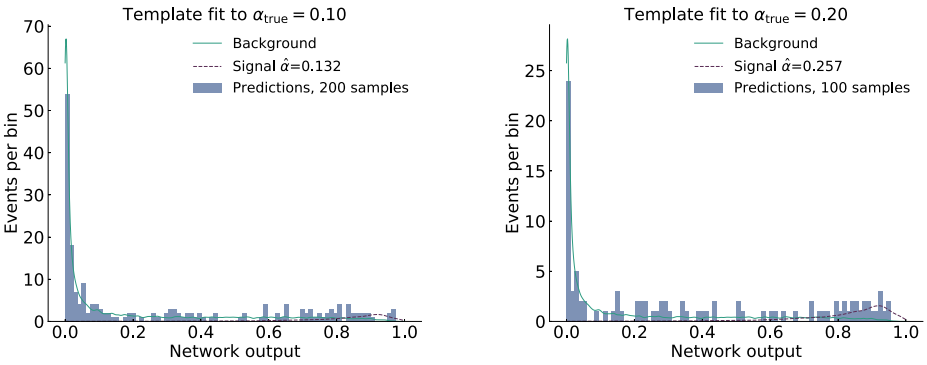
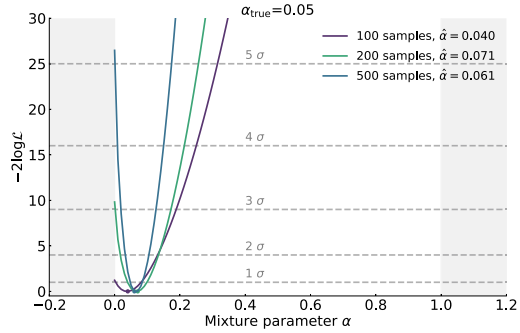
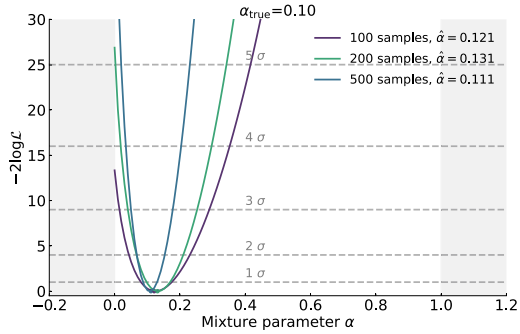


Figure 5.11: Template fits to the classifier output showing the best-fit mixture parameter $\hat{\alpha}$ and the true mixture parameter α of each test set.

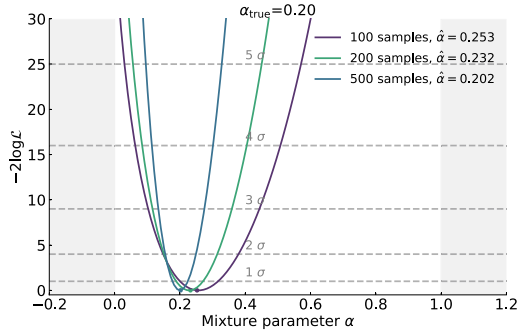
Also, the smaller the mixture parameter, the steeper the parabola needs to be in order to cross the lines indicating the confidence intervals before crossing over to $\alpha < 0$. This makes sense, as larger test datasets mean a larger number of events per class.



(a)



(b)



(c)

Figure 5.12: The negative loglikelihood curves for the template fits to the classifier output, for test set mixture parameters of (a) 0.05 (b) 0.1 and (c) 0.2. Each of the fits is shown for 100 (purple), 200 (green) and 500 (blue) test set events. The intersections with the dashed gray lines indicates the $n\sigma$ confidence intervals.

Future work

The project *Enabling sneutrino detection in weak signal scenarios using machine learning methods* [6] is still in a phase where it requires much work, although the main method is in place. On the technical side, there is room for improving the classifier by further tuning the hyperparameters. Also, a well-tuned neural network given a large amount of training data could still compete with the boosted decision tree. Especially in models where only few events are expected, even a small classifier improvement can make a difference, so it is important to thoroughly investigate this possibility.

On the physics side, the model's parameter space is enormous, and it is anything but trivial to map out the interesting regions. Interesting meaning having a signal dominated by production not from strong processes, but still with a large enough production cross section to expect a fair number of signal events at least at the HL-LHC.

To give a rough idea, three example points' mass spectra are presented in fig. 5.13. These points have expected mixture parameters of 0.073, 0.105 and 0.163 respectively, i.e. in the range of the previously presented sample scenarios, at $150\,\mathrm{fb}^{-1}$ integrated luminosity.[10] Again, this does not in any way constitute a thorough study of the parameter space or indicate upper or lower limits on expected mixture parameters, but merely test points for the method described.

[10] This, or strictly speaking $149\,\mathrm{fb}^{-1}$, is the total integrated luminosity from Run 2 of the LHC.

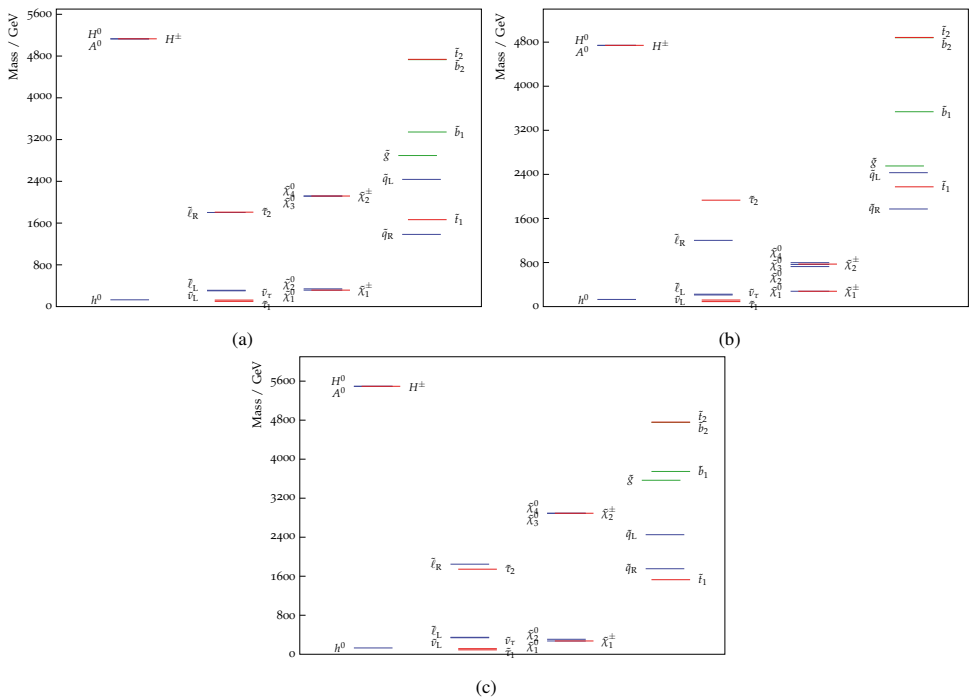


Figure 5.13: The mass spectra of three sample points in the sneutrino NLSP scenario discussed, with expected signal to background ratios of (a) 0.078, (b) 0.117 and (c) 0.195.

Chapter 6

Dark matter indirect detection

6.1 Dark matter

The earliest evidence of gravitationally interacting, non-luminous dark matter was presented in the early paper by Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln* [60], in which he "concluded from the results of observations that the amount of non-luminous matter in the Universe must be greater than that of luminous matter, and thus becoming an early precursor of the dark matter idea" [61].

Since then, stronger astrophysical evidence for the existence of dark matter has been collected, including studies of the CMB, structure formation, galactic rotation curves, galaxy clusters and intergalactic nebulae [62, 63, 64, 65, 66, 67, 68, 69]. According to the current understanding, dark matter makes up more than a fourth of the total energy density of the universe [20].

The few things that are as of today certain about dark matter, are that it interacts gravitationally[1], that its self-interactions are small — observations from colliding galaxy clusters put strong constraints on dark matter self-interaction cross sections, a prominent example being that of the Bullet Cluster $(1E0657-558)$, which recently, on cosmological time scales, passed through another cluster [71] — and that it neither emits nor absorbs photons, and is hence not "dark" in the light-emitting sense of the word, but rather of a, for the time being, obscure nature.

To quantify the energy content which dark matter constitutes in the universe, some quantities must first be introduced. The evolution of the expanding universe with time can be described by a single parameter, denoted $a(t)$. This dimensionless parameter is known as the scale factor of the universe. The 00-component of the Einstein equation [72] with the Friedmann-Robertson-Walker metric [73, 74, 75] yields the Friedmann equation, which defines the Hubble parameter as

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3M_{Pl}^2}\sum_i \rho_i \,. \tag{6.1}$$

Here, $\dot{a}$ is the time derivative of $a(t)$, M_{Pl} is the Planck mass and ρ_i are the contributions to the total energy density of the universe. The scale factor today a_0 thus yields the value H_0 of the Hubble parameter today. The latter enters the expression for the critical density of the universe

$$\rho_c = 3M_{Pl}^2 H_0^2 \,. \tag{6.2}$$

Now, the density parameter of the universe Ω is defined as the ratio of the observed energy density ρ of the universe to the critical density. Neglecting a possible curvature term, the energy density of the universe contains contributions from radiation Ω_R, dark energy Ω_Λ and matter Ω_m. The latter consists of a baryonic and a dark matter component. The

[1] Effort has been made to explain the observed gravitational anomalous phenomena commonly ascribed to dark matter, by modifying only the gravitational theory involved, and thus obviating the need for a dark matter model. A well-known group of such models is that of Modified Newtonian Dynamics [70].

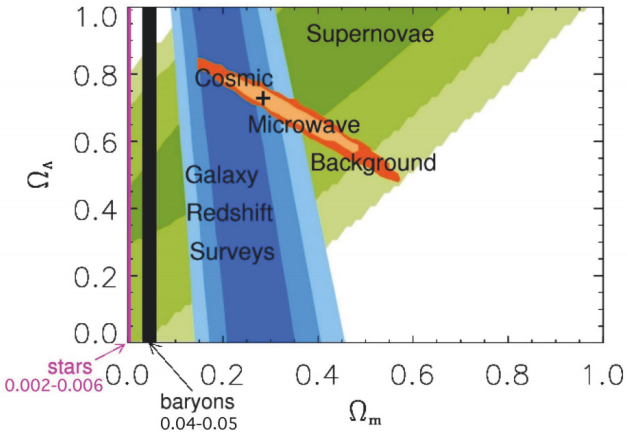


Figure 6.1: The matter density Ω_m versus the energy density Ω_Λ in the universe, the current value indicated by a cross. The matter content of the universe which stretches along the Ω_m-axis from the baryonic density Ω_b band to said cross, must necessarily be made up of non-baryonic matter. Note that this particular plot was made in 2004, and that the measured values have been updated since then. However, the qualitative features of the illustration still holds. *The plot has been taken from [78].*

density parameter determines the geometry of the universe, as

$$\begin{aligned} \Omega = 1: &\quad \text{a flat universe} \\ \Omega > 1: &\quad \text{a spherical, or closed, universe} \\ \Omega < 1: &\quad \text{a hyperbolic, or open, universe}. \end{aligned} \tag{6.3}$$

The current measurements of the matter density Ω_m and the energy density Ω_Λ are summarised in [76], and agree in the region where $\Omega_m \simeq 0.3$ and $\Omega_\Lambda \simeq 0.7$. Using also measurements of the cosmic microwave background and Big Bang Nucleosynthesis, the baryon density Ω_b in the universe has been constrained to less than ~ 0.05. These results combined are presented in fig. 6.1, which shows a plot first presented in [77] and later adapted by [78]. Evidently, dark matter must make up a significant portion of Ω_m.

The criteria a dark matter candidate should fulfill are being be non-baryonic, long-lived on cosmological time scales and electrically neutral. In fact, the Standard Model already has such a particle; the neutrinos. This possibility has been considered, but the neutrinos' relativistic nature ruined their claim to this particular fame [79]. It is in general not possible to have dark matter consist predominantly of relativistic particles, known as hot dark matter, because they could not in the early universe have formed structures as large as galaxies, given that the initial state was very smooth. This smoothness was evidenced by the map of the cosmic microwave background — based on measurements by the COBE (Cosmic Background Explorer) and the later WMAP (Wilkinson Microwave Anisotropy Probe) and Planck satellites, see, e.g., fig. 1.1.

Turning thus to cold dark matter [80], this is a major ingredient in the widely accepted Lambda-CDM cosmological model, in which the universe contains a cosmological constant representing dark energy, denoted Λ. If dark matter is a new kind of particle, charged only very weakly or not at all under the Standard Model gauge groups, it could be accommodated within an extension of the Standard Model. Examples of such particles are heavy sterile neutrinos [81], axions [82], light scalars, Kaluza-Klein excitations of Standard Model fields — which would require a framework with extra dimensions and a bulk — and the list goes on. A supersymmetric model with the gravitino as lightest

Angular momentum l	J^{PC}	Type
0	0^{-+}	Pseudoscalar
0	1^{--}	Vector
1	1^{+-}	Pseudo-vector
1	0^{++}	Scalar
1	1^{++}	Axial vector
2	2^{++}	Tensor

Table 6.1: Different meson states by quantum numbers, from ground states with $l = 0$ to states with orbital excitations.

supersymmetric particle and dark matter candidate is considered in chapter 7.

6.1.1 Indirect detection

Indirect dark matter searches look for visible products of dark matter interactions, focussing often on stable Standard Model particles produced by the decay or annihilation of dark matter, or secondary decay products of these. Astrophysical observations of gamma rays are one example of a potentially promising source for indirect dark matter detection.

As mentioned previously, light sterile neutrinos are candidates for non-baryonic dark matter, and their possible role in astrophysics [83] would yield observable phenomena in the form of monochromatic photons with low energy, i.e. astrophysically detectable in the keV range.

On the high energy side, observations of the Milky Way halo by the Fermi-LAT (Large Area Telescope) collaboration, have searched for gamma ray spectral lines in the energy range 200 MeV to 500 GeV [84], and possible gravitino dark matter decay signatures in the energy range 100 MeV to 10 GeV [85].

Evidently, these searches for photons originating from processes involving dark matter, have left an energy range with relatively little attention, namely that from around 100 keV up to just below 100 MeV. This range has been coined the MeV-gap, and if dark matter couples somehow to Standard Model particles, it is possible that the signatures reside within this gap. Furthermore, it is interesting to observe that this is the energy range for photons coming from the decays of heavy Standard Model mesons at rest.

This is the motivation behind a project on gamma rays from the annihilation of dark matter into heavy mesons, which will be explained in the following. The project formed a contribution to the e-ASTROGRAM white book [3], see also section 6.4, under the title *Smoking gun dark matter signatures in the MeV range*.

6.2 Heavy mesons

Mesons are bound states of a quark and an antiquark, as such have baryon number 0, and all the known ones are unstable. They are classified using the three quantum numbers J, P and C, and the notation is J^{PC}. Denoting the orbital angular momentum by l and the spin by s, the total spin J of the meson is constrained via the familiar relation

$$|l - s| \leq J \leq |l + s| \,. \tag{6.4}$$

The spins of the two quarks in the meson can align in antiparallel, forming a total $s = 0$ state, or in parallel, forming a total $s = 1$ state. The parity and charge conjugation of a meson state are

$$\begin{aligned} P &= (-1)^{l+1} \\ C &= (-1)^{l+s} \end{aligned} \tag{6.5}$$

where the latter relation is defined *only* for mesons consisting of quarks with their own antiquarks. The allowed combinations are shown in table 6.1.

The combinations of quarks can be light-light, light-heavy or heavy-heavy. The lightest mesons, and also the lightest hadrons, are the pions $\pi^0, \pi^\pm$, consisting of combinations of the up and down quarks. The non-perturbative computation of many meson properties in QCD are available only using techniques from lattice QCD. [2] These are computationally and conceptually much harder for light than for heavy quarks, due to e.g. relativistic and vacuum effects [86].

Bound states of two heavy quarks are sometimes called quarkonia, and examples of these are

- The J/ψ meson: the charmonium $c\bar{c}$ ground state
- The Υ meson: bottomonium $b\bar{b}$

One might at this point ask why there is no toponium, a $t\bar{t}$ bound state, and the answer is that the very heavy top quark decays electroweakly before it can form a bound state — a rare example of a weak process happening faster than a strong process.

6.2.1 Meson decays

Mesons decay via lighter mesons and leptons, in chains which end up in pions, electrons, muons, neutrinos and photons. The pions and muons themselves also decay into leptons and photons, meaning that these chains all have electrons, neutrinos and photons with different energies as stable particles at the bottom of the chain.

Measuring the photon energies from such a decay chain would yield a relatively featureless spectrum looking more like a broad bump than anything else, and as such resemble the astrophysical background. Said background consists primarily of interstellar gamma rays [87] from inverse Compton scattering, decaying free-flying pions and Bremsstrahlung from particles flying off from some process they were lucky enough to be involved in, perhaps something as spectacular as star formation or a neutron star collision.

Since mesons are composite systems, they can have different energy levels. If the two quarks have a higher relative angular momentum than the lowest $l = 0$, they form an excited meson state, usually denoted by an asterisk in the name of the meson. Such excited mesons can decay into their $l = 0$ state by radiating off a photon or a pion. For instance, the excited B-meson state decays dominantly as $B^* \to B\gamma$, and the excited D-meson state as $D^* \to D\pi^0$. Since the photon or pion energy corresponds to the difference in mass between the meson states, a signal consisting of these photon energies would in the meson rest frame stand out from the astrophysical background as either a sharp line or box feature, as realised by [88].

The above discussion is interesting in the context of dark matter. Suppose that dark matter has some coupling to either a Standard Model particle, or to some non-Standard Model particle which in turn does carry $SU(3)_C$ charge. This would open up the possibility of dark matter annihilations resulting in production of Standard Model quarks.

The authors of [88] demonstrated the possible existence of spectral features from dark matter annihilation via Standard Model quarks through excited mesons using `Pythia`. Before taking the same approach here, a brief introduction to event generation, and how `Pythia` can help simulate astrophysical spectra.

6.3 Event generation

Event generators, meaning here general-purpose Monte Carlo[3] generators, simulate high-energy collisions and are a central tool in high-energy physics. For instance at the LHC, where particles are produced in proton-proton collisions, analytically calculating detailed distributions on a particle level would rely on an analytical model for strong interactions at the non-perturbative level.

[2] Lattice QCD is a non-perturbative, mathematically well-defined framework where gauge theories in spacetime have been discretised onto a lattice, using the path integral formalism of quantum field theory.

[3] Monte Carlo algorithms are a wide range of algorithms using random sampling, and are used in particle physics to generate draws from a probability distribution, as a subsistute for modelling the underlying quantum field theory.

For the projects in this thesis, the role of event generators is in some sense twofold: In the present chapter, the use is to simulate data from decay processes and record final state particle energies. In chapter 5 and chapter 7, the use is to, together with detector simulation, provide collider signals for statistical analysis of beyond Standard Model theories.

An event generation process can be divided into the following steps

1. **Hard process**: These are in general processes at energy scales above a few GeV, that can be treated perturbatively. For instance, the initial parton-parton collision is a hard process.
2. **Parton showering**: This generates the QCD ratiation from the partons coming from the hard event.
3. **Early decay**: Some of the produced heavy resonances decay rapidly, e.g. the top quark, Higgs, vector bosons and heavy sparticles.
4. **Hadronisation, or soft processes**: The partons left after showering combine and form hadrons
5. **Unstable hadrons decay**: These decays happen in accordance with experimentally measured, or manually entered, branching ratios.

6.3.1 Heavy meson spectra

As mentioned, the event generator `Pythia` can be used to simulate possible photon spectra from dark matter annihilation, and the central functionality resides in item 5 in the above list.

If the dark matter mass and energy is within the same range as the Standard Model mesons, dark matter could annihilate as

$$\chi\chi \to M_1 M_2 M_3 \ldots,$$

where the M_i are mesons produced with little kinetic energy. To simulate the spectrum from the resulting decays of such a process using `Pythia`, this is initialised with centre-of-mass energy corresponding to the combined dark matter mass $2m_\chi$. The event contains one particle representing the state formed by a dark matter annihilation, and the open decay channels specify which Standard Model mesons are produced in the annihilation. For each event, the energies of the final photons are collected and stored. Figure 6.2 show example results of such a simulation, for two different dark matter energies. Here, `Pythia` decayed the initial particle into the excited states B_s^* and $\bar{B}_s^*$, accompanied by a pion. Evidently, the line feature is sharpest when the dark matter mass is closest to the production threshold of the three mesons.

6.3.2 Quarkonium resonances

Another kind of process which could take place and yield detectable signatures, is

$$\chi\chi \to (Q\bar{Q})\gamma, \tag{6.6}$$

where $(Q\bar{Q})$ is a quarkonium state. This can result in monochromatic photons in the same way as discussed before, but in addition, the photon produced back-to-back with the quarkonium would have a sharp energy spectrum, given by the difference in mass between the dark matter particle and the quarkonium. The annihilation of dark matter into quarkonia and a photon was considered in early work [89, 90, 91], and interesting for the present discussion is heavy meson production through a quarkonium resonance where the photons from the meson decays land in the MeV gap.

As an example, consider the process where two dark matter particles annihilate into $\Upsilon(10860)$ and a photon, where the former decays primarily into B and B_s mesons and their excited states. This can be again be simulated using `Pythia`, but the event must be initialised in a different way. This time, it contains two initial particles; the $\Upsilon(10860)$ and a photon with opposite momentum. Furthermore, the decay table of the quarkonium given to `Pythia` corresponds to experimental values [76]. Figure 6.3 shows the resulting final photon energies from such a simulation.

In agreement with the results in [88], the spectral features from processes such as those described above would be astrophysically visible in the lower part of the GeV range of the energy spectrum. With the simulations in place, this chapter ends with a brief discussion about an astrophysical mission with the possibility of detecting these signals.

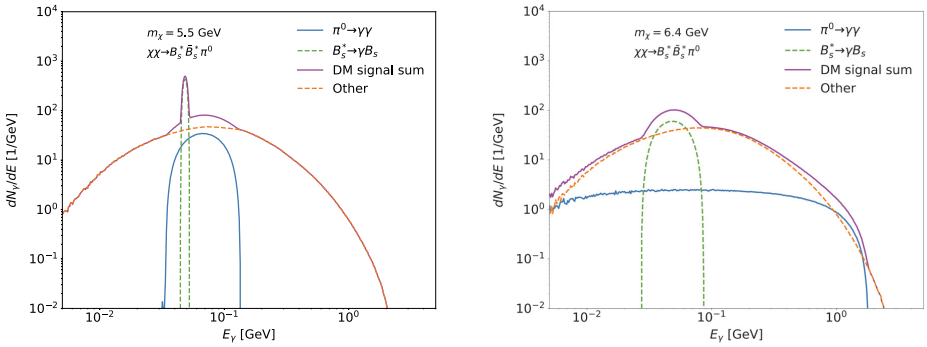


Figure 6.2: The photon spectra from a Pythia simulation of a dark matter annihilation process into Standard Model mesons, in this case $B_s^* \bar{B}_s^* \pi^0$, and subsequent decays.

6.4 e-ASTROGRAM

The enhanced ASTROGRAM [3] is an observatory space mission, designed for observations in the 0.3 MeV to 3 GeV photon energy range. As mentioned previously, this domain remains astrophysically largely unexplored, after NASA's COMPTEL[92] stopped taking data in 2000, and no successors to that mission are planned.

Potential e-ASTROGRAM observations would enable studies of a range of astrophysical processes, reaching from relativistic jets via low-energy cosmic rays to the recent history of supernova explosions in the Milky Way and detecting supernovae in nearby galaxies, to mention a few. Still, of relevance for the present discussion is the potential dark matter annihilation signatures which could be found in the range which e-ASTROGRAM is, at the time of writing, the only candidate to observe.

As a part of the presentation of the e-ASTROGRAM mission, a white book was written, containing a presentation of a wide range of physics questions which could be answered given observations by e-ASTROGRAM. Among the topics belonging to fundamental physics, dark matter is strongly represented. Under the title *Smoking gun dark matter signatures in the MeV range*, the annihilation of dark matter into Standard Model mesons and the resulting detectable features, was discussed. A plot similar to that in fig. 6.3, was contributed[4], alongside similar results from Bringmann et al. [88], see [3], chapter 4, section 4.

After the e-ASTROGRAM mission was unfortunately not selected as three of ESA's M5 mission concept candidates, the e-ASTROGRAM collaboration joined with NASA's AMEGO [93], the All-sky Medium Energy Gamma-ray Observatory, an astrophysical mission with the same scientific objectives and similar detector design. Efforts to launch the e-ASTROGRAM mission, probably together with AMEGO, continue, and the potential this would have for observations of dark matter processes in the MeV range is unprecedented.

The results contributed to the e-ASTROGRAM white book also form the basis of an ongoing project, under the name *Gamma rays from the annihilation of dark matter into heavy mesons* [5], which investigates the potential role played by quarkonium resonances in enhancing the production of excited mesons. In the above simulations for fig. 6.2, as well as those in [88], underlies the assumption that the dark matter mass is exactly on the production threshold for the excited mesons. The farther away from the combined mass of the meson states, the broader the photon spectrum from the decays becomes, and the harder it becomes to distinguish from the astrophysical background. Realising the scenario described above hence has a delicate dependence on the dark matter mass.

However, as long as the theory for the dark matter interaction has the vector operator $\bar{Q}\gamma^\mu Q$, the annihilation can

[4]The contribution to the e-ASTROGRAM white book differs from fig. 6.3 in that the dark matter mass is instead $m_\chi = 6.4$ GeV in the simulation.

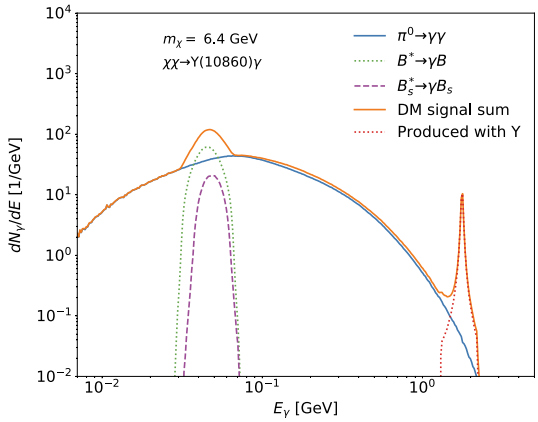


Figure 6.3: The photon energies from a Pythia simulation of a dark matter annihilation process into a Standard Model quarkonium state and a photon, and subsequent decays. The feature in the lower end of the spectrum stems from the subsequent meson decays, and that in the higher end from the photon produced along with the quarkonium state. This plot is similar to the one in [3], discussed further in section 6.4.

happen via a quarkonium resonance, such as the one illustrated in fig. 6.3. Since such a resonance can decay into excited meson states, the sharp feature in the photon spectrum remains, but less of the delicate dependence on being in exactly the right mass range. Additionally, annihilation through a resonance may lead to enhancement or suppression in $\langle \sigma v \rangle$, depending on temperature through the velocity distribution of the dark matter. This can lead to potential enhancements in the annihilation cross sections for indirect detection today, compared to the cross section that sets the relic density.[5] Quarkonium resonances may thus play a particularly important and hitherto unnoticed role in enhancing signals from dark matter annihilation.

[5]The thermally averaged cross section and the dark matter relic density are discussed in chapter 7.

Chapter 7

The early universe and supersymmetric dark matter

One of the many features of supersymmetry is that it offers several candidates for particle dark matter. The nature of a supersymmetric dark matter particle has strong implications for the physics of the early universe. The present chapter starts by giving a short introduction to the model-independent mechanisms in the early universe relevant for particle physics phenomenology, before moving on to discuss supersymmetric dark matter scenarios.

This forms the basis for a project which was unfortunately not finished within the time limitation of the doctoral period. The working title of the project is *An open window for high reheating temperatures in gravitino dark matter scenarios*, and effort is made for it to be finished in the near future. This chapter builds a summary of what has been done in the project at the time of writing, and discusses the motivation behind it.

7.1 The particle contents of the early universe

7.1.1 Abundances

Think of the early universe as a hot thermal bath in which particles are constantly colliding and interacting. The temperature of the thermal bath corresponds to the average energy per particle, and as the universe expands, the temperature decreases. Consider next a stable particle χ which forms part of the thermal bath. As long as $T >> m_\chi$, the particle is assumed to be in thermal equilibrium. The equilibrium abundance is maintained by the balance between two processes

- $\chi\bar{\chi} \to X\bar{X}$ — annihilation of the particle into lighter particles X.
- $X\bar{X} \to \chi\bar{\chi}$ — production of the particle via thermal scattering of other particles. This is sometimes referred to as thermal production of the particle.

This process is described by the Boltzmann equation, which contains the temporal development of the χ number density,

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\langle\sigma_{\chi\bar{\chi}\to X\bar{X}}v\rangle\left(n_\chi^2 - (n_\chi^{eq})^2\right). \tag{7.1}$$

The term containing the Hubble parameter H accounts for dilution due to cosmic expansion, while the right-hand side represents interactions between particles, and n^{eq} denotes the equilibrium number density. The thermally averaged cross section $\langle\sigma v\rangle$ contains the cross section σ for the relevant process and the relative velocity v between the particles colliding. It is an average over all velocities at a given temperature, so eq. (7.1) is not a function of v, but rather of the temperature T. Simply put, the interaction rate is determined partly by the cross section, as calculated using a particle physics model, and partly by the temperature of the particles.

As the universe expands and the temperature decreases below $T_f^\chi < m_\chi$, the particle χ freezes out of the thermal bath. Look again at the distribution functions for bosons eq. (2.5) and fermions eq. (2.42): as the temperature drops below the mass of the particle, this becomes non-relativistic and its distribution function becomes exponentially suppressed as $f \to e^{-m/T}$. The number density n_χ approaches the relic abundance when the annihilation rate is on the order of the Hubble, or expansion, rate

$$\Gamma = n_\chi \langle \sigma_{\chi\bar{\chi}\to X\bar{X}} v \rangle \sim H \,, \tag{7.2}$$

i.e. when the density of the particle is so small that its annihilation is too inefficient to keep it in thermal equilibrium. Note that $\sigma_{\chi\bar{\chi}\to X\bar{X}}$ denotes the total annihilation cross section for the particle into any species.

7.1.2 Inflation and reheating

Assuming an epoch of inflation [94], during which the scale factor of the universe grew exponentially and the particle density in the universe was consequently strongly diluted, can explain the overall homogeneity, spatial flatness and large size of the current universe.

The first of these issues, often called the horizon problem, emerges from the realisation that the cosmic microwave background radiation has the same temperature in all directions. In fact, it seems evident that now disconnected regions in the universe must previously have been connected, allowed to equilibrate to a common state. If these regions were thermally connected prior to inflation, and only driven apart during inflation, this explains the observation of homogeneity throughout the universe.

The flatness issue is a fine-tuning problem concerning the density parameter of the universe. Had this had an only slightly different value in the past, its present-day value would have differed vastly from the observed value. As mentioned in eq. (6.3) , the density parameter of the universe Ω is defined as the ratio of the observed density ρ to the critical density. Now, the current density of the universe is observed to be very close to this critical value, and since the density parameter evolves in cosmic time, the deviation of the Planck-time value of the density parameter cannot have been more than one part in 10^{60}. This seeming need for fine-tuning is alleviated by inflation, as a small, but not necessarily critically small, value for the initial curvature of the universe before inflation, would be enough to render the universe as flat as we observe it today.

Exactly which process drove inflation is not yet known, but if particle physics can supply the cause, this would likely be by introducing one or several scalar fields which transition from a flat region to a minimum of the scalar potential. The fields' associated particles, commonly called inflatons, would then have decayed as the inflationary period ended[1], and the energy thereby released transferred into a hot thermal bath of elementary particles. This process would then have caused the temperature of the universe to increase again, known as reheating. The maximum temperature reached during the period of reheating is denoted T_R.

7.1.3 Thermal leptogenesis

As mentioned in the introduction section 2.6.5, the baryon asymmetry in the universe is not predicted by the Standard Model, and it is natural to inquire about the mechanisms which may have produced it. Sakharov formulated three conditions for the generation of a baryon asymmetry [95]

- Baryon number violation
- $\mathcal{CP}$- and $\mathcal{C}$-violation
- Departure from thermal equilibrium

The three of these can be satisfied by heavy Majorana neutrinos — the seesaw-partners of the light neutrinos of the Standard Model.

[1] Actually, the inflaton field would likely have oscillated about the minimum of its potential, before rapidly decaying and needing some time to reach thermal equilibrium. This process is sometimes referred to as preheating.

Once the temperature of the universe drops below the mass of any heavy Majorana neutrinos, these do not follow the changes in equilibrium distribution as quickly as the interacting particles. This deviation from thermal equilibrium results in a large number density of heavy Majorana neutrinos, compared to the equilibrium density of other particles. The heavy neutrinos decay through $\mathcal{CP}$- and $\mathcal{C}$-violating processes, generating a lepton asymmetry, which is converted into a baryon asymmetry through so-called sphaleron processes[2].

This mechanism both yields baryogenesis and makes a statement about neutrino properties, as it constrains the masses of both the light and heavy neutrinos. The lightest neutrino mass is for instance bound to $m_\nu \leq 0.1$ eV [97], to be compared with cosmological probes, until laboratory-based experiments can provide more stringent limits. At the time of writing, limits are set at $m_\nu \leq 0.120\,\text{eV}$[3]. The observation of neutrino oscillations, and hence a definitely non-vanishing neutrino mass, supports thermal leptogenesis. In order for the mechanism to produce the observed baryon asymmetry, a high reheating temperature of $T_R \gtrsim 10^9\,\text{GeV}$ is required [96, 98].

7.2 The supersymmetric early universe

In the spectrum of supersymmetry, there are several massive particles which carry neither colour nor electric charge. If the lightest supersymmetric particle has these properties and is in addition stable, either by virtue of an R-parity or due to extremely weak couplings, it is an excellent candidate for particle dark matter.

7.2.1 The gravitino and its problem

A local supersymmetric model implies the existence of the gravitino $\widetilde{G}$, the superpartner of the graviton. As supersymmetry is spontaneously broken, the associated spin-$\frac{1}{2}$ goldstino is absorbed by the gravitino, which thus inherits both its mass and spin-$\frac{1}{2}$ modes. Depending on the scheme of SUSY breaking, the gravitino mass $m_{3/2}$ can range from the eV to beyond the TeV region [44]. However, a too light gravitino, if identified as the dark matter particle, is not favoured by structure formation in the early universe, and its mass is bounded from below by present observations to a few keV [99].

The gravitino abundance produced thermally after the Big Bang is strongly diluted during inflation. However, after the reheating phase, during which a temperature T_R is reached, gravitinos are again produced through scattering processes of particles in the thermal bath. The gravitino abundance is then approximately proportional to T_R, see eq. (7.6). In a local supersymmetric model where the gravitino is not the LSP, its very weak couplings cause it to decay very slowly into whichever other particle is the LSP, with a lifetime of around [100]

$$\tau_{3/2} \approx \frac{1}{\alpha_{3/2}} \frac{M_p^2}{m_{3/2}^3}\,. \tag{7.3}$$

Here, $\alpha_{3/2}$ is dimensionless and $\mathcal{O}(1)$, meaning that the lifetime of the gravitino is

$$\tau_{3/2} \gtrsim 3.2\text{yrs}\left(\frac{100\text{GeV}}{m_{3/2}}\right)^3, \tag{7.4}$$

i.e. by far outlasting Big Bang Nucleosynthesis, which is estimated to have taken place seconds to minutes after the Big Bang. Gravitino decays taking place during or after Big Bang Nucleosynthesis, would release energetic decay products which could destroy the light nuclei produced [101, 102], see also section 7.3.2.

This problem has two possible solutions: The gravitino could be very heavy and decay sufficiently early, which would require a correspondingly large scale $\langle F \rangle$ of supersymmetry breaking, or its abundance could be kept low through an upper bound on T_R. A gravitino mass of $m_{3/2} \lesssim 5\,\text{TeV}$ amounts to an upper bound on the reheating temperature of $T_R \lesssim 10^{5-6}\,\text{GeV}$ [103], which would in turn rule out standard thermal leptogenesis. [96, 98]

[2] Sphaleron processes are beyond the scope of this thesis, but the interested reader can find more information in an excellent introduction by Buchmüller et al [96].

[3] at 95% confidence level, summing over the three flavours [21].

Table 7.1: Gauge couplings g_i, gaugino mass parameters M_i and values of associated constants for the gauge groups $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$, enumerated by the index i.

Gauge group	i	g_i	M_i	k_i	c_i	ω_i	β_i
$\mathrm{U}(1)_Y$	1	g'	M_1	1.266	11	0.018	11
$\mathrm{SU}(2)_L$	2	g	M_2	1.312	27	0.044	1
$\mathrm{SU}(3)_C$	3	g_s	M_3	1.271	72	0.117	−3

This quandary is known as the gravitino problem, and suggests that the gravitino should be the LSP, stable on cosmological scales and possibly the dark matter particle.

7.2.2 Gravitino production and abundance

After inflation has diluted away any initial gravitino population[4], the evolution of the gravitino number density $n_{3/2}$ per cosmic time t is described by the Boltzmann equation [104]

$$\begin{aligned} &\frac{dn_{3/2}}{dt} + 3Hn_{3/2} = C_{3/2} \\ &C_{3/2} = \sum_{i=1}^{3} \frac{3\zeta(3)T^6}{16\pi^3 M_P^2}\left(1 + \frac{M_i^2}{3m_{3/2}^2}\right) c_i g_i^2 \ln\left(\frac{k_i}{g_i}\right). \end{aligned} \tag{7.5}$$

Here, i denotes the gauge group, and the M_i the corresponding gaugino soft masses. As before, the term proportional to H represents dilution due to the cosmic expansion, while the right-hand side is the collision term and accounts for destruction and production in the thermal bath. The temperature T sets the scale for evaluation of the parameters in $C_{3/2}$, for which all parameters are given in table 7.1. From this, the thermal contribution to the gravitino relic density is [100, 105, 106]

$$\Omega_{3/2}^{\mathrm{th}} h^2 = \sum_{i=1}^{3} \omega_i g_i^2 \left(1 + \frac{M_i^2}{3m_{3/2}^2}\right) \ln\left(\frac{k_i}{g_i}\right) \left(\frac{100\mathrm{GeV}}{m_{3/2}}\right) \left(\frac{T_R}{10^{10}\mathrm{TeV}}\right). \tag{7.6}$$

The final gravitino abundance is constrained by dark matter relic density limits from observations. At the time of writing, an analysis of the anisotropies of the cosmic microwave background measured by the Planck space observatory settles the dark matter density at $\Omega_c h^2 = 0.120 \pm 0.001$ [21]. This again amounts to an effective constraint on T_R, as indicated by the above equation. However, as mentioned, leptogenesis requires a reheating temperature of $T_R \gtrsim 10^9$ GeV [96, 98], so in order to achieve this while avoiding a too high dark matter relic density, it seems suggestive to lift $m_{3/2}$.

The gravitino relic density also receives a non-thermal contribution[107, 108], the one from the decay of the NLSP. Denoting a general NLSP by χ, this is

$$\Omega_{3/2}^{\text{non-th}} h^2 = \frac{\Omega_\chi h^2}{m_\chi} m_{3/2}, \tag{7.7}$$

i.e. also dependent on the mass of the gravitino, and of course the abundance of the NLSP.

7.2.3 The neutralino and its abundance

Assuming a gravitino LSP, its extremely weak couplings imply a large NLSP lifetime, so large, indeed, that decays into the gravitino can be neglected when considering the freeze-out of the NLSP in the very early universe.

[4] Neglecting gravitino production from earlier epochs or inflaton decay. Such other scenarios could of course lead to sizeable contributions, but they are strongly model dependent and not discussed further here.

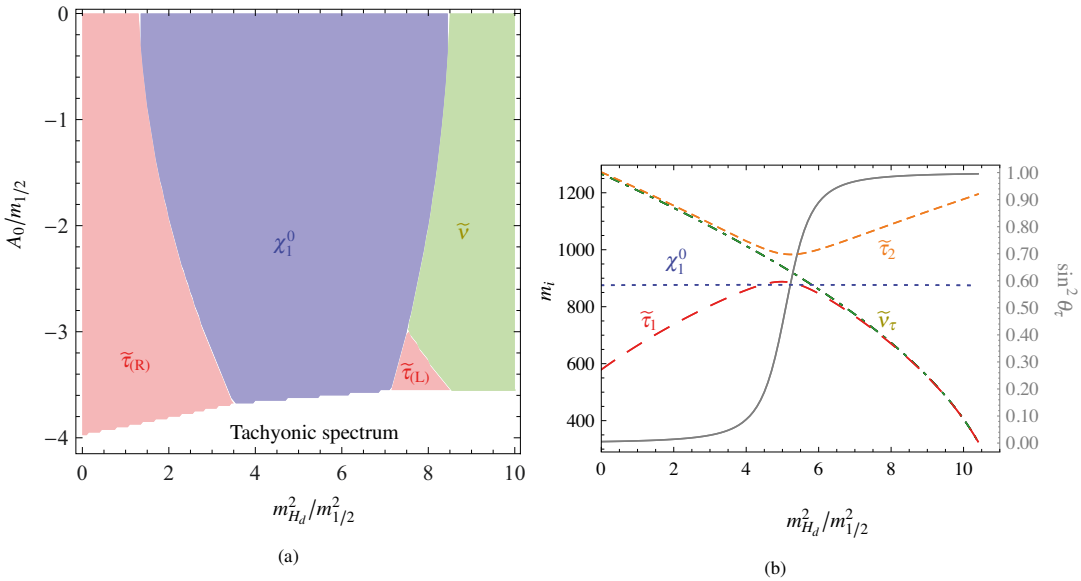


Figure 7.1: The identity of the NLSP in a supersymmetric model with a gravitino LSP, (a) for varying Higgs down soft mass and trilinear coupling, scaled to the common gaugino mass parameter, with $\tan\beta = 10$ and $m^2_{H_u} = 0$, and (b) the variation of the mass (m_i) of the NLSP candidates with the stau mixing angle, with $\tan\beta = 20$ and $m^2_{H_u} = 5\,\mathrm{TeV}^2$. The subscript '1' on the symbol of the neutralino indicates that this is the lightest of the neutralinos, and will be omitted in the following. The plots were produced for [1], where the identity of the NLSP across the parameter space is further discussed.

At some point during these early times, the NLSP thermalises and enters thermal equilibrium. As the temperature of the universe drops below the NLSP mass, the equilibrium number density of the NLSP decreases exponentially. Since the annihilation rate is proportional to the number density, the annihilation rate becomes inefficient, and thus the NLSP departs from thermal equilibrium and freezes out.

The point of the freeze-out depends on the annihilation cross section; a small cross section implies an early freeze out, and a large one a later. In other words: The larger the cross section, the smaller the final number density. This means that the NLSP relic density Ω_χ is determined solely by the thermal freeze-out, and independent of the reheating temperature, in contrast to the case of the gravitino.

If the lifetime of the NLSP is so large that decays take place during or after Big Bang Nucleosynthesis, the Standard Model particles produced in relation with these decays, primarily $\chi \to \widetilde{G} + X$ and some three-body decays, can affect the primordial abundance of light elements and contradict astrophysical observations. The larger the relic abundance of the NLSP, the closer the model comes to conflict observations, specifically bounds from Big Bang Nucleosynthesis. A higher gravitino mass implies a slower decay of the NLSP, again implying stringent constraints on Ω_χ.

In summary, there are strains from several angles when trying to construct a framework featuring gravitino dark matter — the solution to the gravitino problem present in all local supersymmetric models! — compatible with Big Bang Nucleosynthesis and thermal leptogenesis. Furthermore, the phenomenology of the scenario depends strongly on which particle is the NLSP, and the candidates for this role are the stau, the sneutrino and the neutralino, see fig. 7.1a.

Probably the most studied one of these candidates is the stau. In supersymmetric models with minimal gauge mediated SUSY breaking, which typically predict a heavy gravitino, the stau is naturally the NLSP. The stau also decays

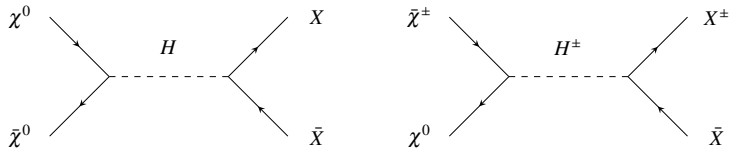


Figure 7.2: Given the right mass conditions, the neutralino χ^0 can (a) annihilate or (b) coannihilate with a near-mass chargino $\chi^\pm$ resonantly via a heavy Higgs into Standard Model particles, here denoted X.

via leptons, meaning that its decay carries with it few energetic hadrons to affect Big Bang Nucleosynthesis. However, since the stau is charged, typical stau relic abundances lead to an upper bound $T_R \lesssim 10^{8-9}\,\mathrm{GeV}$ [109, 110].

From a Big Bang Nucleosynthesis point of view, the sneutrino is an excellent candidate, since it evades these bounds even for a large reheating temperature, under a mass constraint of around 300 GeV [111]. However, from a model point of view, achieving a sneutrino NLSP is difficult, see fig. 7.1b, especially while keeping the mass of the lightest Higgs in the observed range, and the model explorable from a collider point of view. See also section 5.3 for a discussion on sneutrino detection.

The neutralino as NLSP candidate in a scenario such as the present was thoroughly considered in [112]. It is interesting to look closer at this scenario in a particular corner in parameter space, namely where a resonant decay of the neutralino via a heavy Higgs can help lower the neutralino relic density and thus allow for a high reheating temperature in these supersymmetric models.

7.2.4 Funneling the neutralino

In order for the neutralino to annihilate resonantly via a heavy Higgs state, see the Feynman graph depiction of the process in fig. 7.2a, the neutralino mass must be equal or slightly below half that of the heavy Higgs,

$$\frac{1}{2} m_{\chi^0} \lesssim m_H \,. \tag{7.8}$$

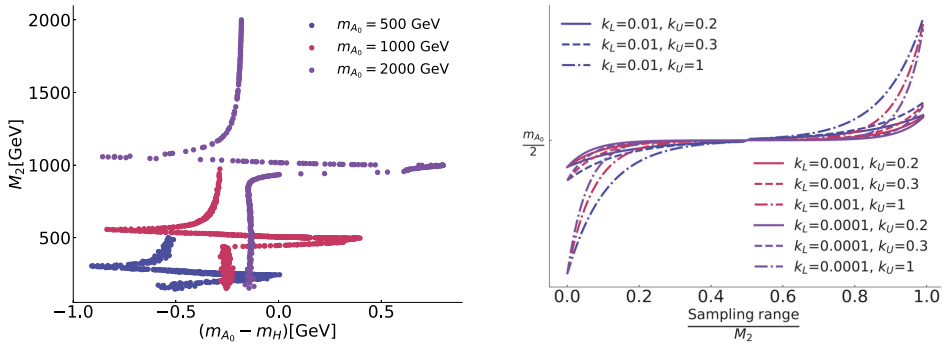
Moreover, if the neutralino has significant wino or higgsino admixture, the lightest chargino is expected to be close in mass to the neutralino, and so coannihilation processes can become relevant, see a depiction in fig. 7.2b. Which process is more efficient depends on the cross sections, and the final relic densities of either particle become codependent.

A technical challenge related to targeting this region within a specific model, is that spectrum generators yield the physical mass spectrum as output — not input — meaning that the region must be targeted indirectly, via the soft parameters in the supersymmetry Lagrangian.

Keeping in mind the neutralino and chargino mass matrices, the neutralino becomes light if at least one of (M_1, M_2, μ) is low, while the chargino mass relates similarly to M_2 and μ. The parameters which most directly affect the quantities of interest, are thus on the one hand the pseudoscalar Higgs pole mass m_{A_0}, and on the other the $SU(2)_L$ gaugino, or wino, mass M_2. Ridding the discussion of strong complications, the scalar spectrum can be pushed up beyond concern by adjusting the soft scalar masses, which in the following discussion is set to 15 TeV. Keeping all other input parameters — i.e. $M_1, M_3, \tan\beta, m_{A_0}$ and the aforementioned soft scalar masses — fixed, varying M_2 and μ around half the Higgs pole mass m_{A_0} and calculating the physical mass spectrum, yields the splits between the Higgs pole mass and the physical mass shown in fig. 7.3a. The variation of 1 GeV in this case is probably due to numerical uncertainties in the spectrum calculation, and in any event far below the theoretical uncertainty.

The values of M_2 which enter this calculation are obviously not linearly sampled, but densely sampled around half m_{A_0} and more sparsely on the edges. This is done using the sampling function

$$f_{sampling}(x) = k_L \exp\left\{ x \log\left(\frac{k_U}{k_L} \right) \right\} , \tag{7.9}$$



(a) (b)

Figure 7.3: (a) The split between the Higgs pole mass m_{A_0} and the physical mass m_H as calculated by `Softsusy`, versus the input soft mass parameter M_2. The different values for m_{A_0} are indicated on the plot, and M_2 is varied around half the Higgs pole mass according to the prior function eq. (7.9). The other input parameter values are $M_1 = 1000$, $M_3 = 2000$, $\tan\beta = 10$, $\mu = M_2$, $m_{A_0} = 2000$, all masses in GeV. (b) The function in eq. (7.9) wth coefficients $k_L = (10^{-2}, 10^{-3}, 10^{-4})$ and $k_U = (0.2, 0.3, 1.0)$.

which is shown in fig. 7.3b for different values if k_L and k_U. The sampling of M_2 for the results in fig. 7.3a is done using $k_L = 10^{-3}$ and $k_U = 0.3$

7.3 One scan to find them

In order to investigate how low the neutralino NLSP relic density can become — and consequently how allowed the parameter point is considering upper bounds from relic density and big bang nucleosynthesis — this region, coined the Higgs funnel region, is targeted in a parameter scan. The scan is of the type described in section 3.5.3, i.e. guided by the likelihoods of the physical observables, see section 3.2.1 for a discussion on likelihoods, and section 3.4.3 for hypothesis testing. The scan varies the input parameters, being the reheating temperature, the gravitino mass, the gaugino masses and the Higgs soft parameters, in an attempt to locate allowed regions for the neutralino NLSP scenario.

7.3.1 Reheating temperature and thermal relic density

To illustrate, consider first a very simple scan, which calculates only the thermal contribution to the gravitino relic density in eq. (7.6), and tries to maximise a Gaussian likelihood, see eq. (3.16), to the measured value of $\Omega_c h^2 = 0.120$. This scan has an easy job, but since the objective is to maximise the reheating temperature, the scan can be given a "fake" likelihood function which competes with the physical one, and incentivises the scan to pursue high values for T_R. Again, eq. (7.6) shows that there is a trade-off here, as the relic density increases with T_R.

The result from the scan trying to find the sweet-spot between these two contributions is shown in fig. 7.4. In the following, the $-2\log\lambda$ on the plots is short for $-2\log\mathcal{L}_{\text{scan}}$, i.e. the combined likelihood for each point, minus $-2\log\mathcal{L}_{\text{bf}}$, the overall best-fit value for the entire scan. As explained in section 3.4.1, this likelihood ratio follows a χ^2 distribution. The coloured regions represent n standard deviations σ for the χ^2 distribution with two degrees of

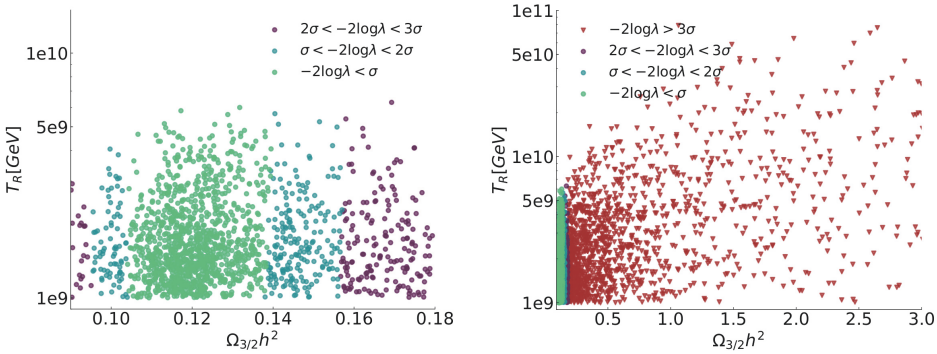


Figure 7.4: A parameter scan guided by two likelihood components; a Gaussian likelihood for the thermal contribution to the relic density of the gravitino, and an unphysical likelihood increasing with the reheating temperature. (a) Shows the points in the 1σ, 2σ and 3σ regions in green, blue and purple respectively, while (b) includes also the points beyond 3σ, shown in red, to illustrate the region the scan has investigated.

freedom, where the numerical values are

$$\begin{aligned} 1\sigma(\chi^2_{n=2}) &= 2.30 \\ 2\sigma(\chi^2_{n=2}) &= 6.18 \\ 3\sigma(\chi^2_{n=2}) &= 11.83\,. \end{aligned} \tag{7.10}$$

Figure 7.4a shows the one-, two- and three-sigma regions in colours, while fig. 7.4b includes part of the region beyond three sigma. The latter illustrates how the scan has attempted to keep the reheating temperature high, while struggling to achieve a sufficiently low thermal contribution to the gravitino relic density in the process. This particular scan had to settle for a maximum reheating temperature of $T_R^{\max} = 5.82 \cdot 10^9\,\text{GeV}$ in the one sigma region.

This reheating temperature is a very promising start, but several constraints have yet to be considered.

7.3.2 Big Bang Nucleosynthesis

Some of the more stringent bounds on the relic density of the formerly introduced neutralino NLSP, come from Big Bang Nucleosynthesis, abbreviated BBN. In short, BBN bounds are all about how much energy can be injected during and after the formation of light elements. Since the primordial abundances of these have been measured with high accuracy, they leave little room for processes in the early universe which result in more hadronic or electromagnetic energy. For instance, producing energetic photons in a late supersymmetric decay could cause photodissociation of the earliest forming nuclei.

The previously introduced parameter scan is extended to take constraints from BBN into account by implementing the bounds from [103], figures 11 and 12. Although the different channels in these plots yield varying limits, it is clear that the trend is to forbid large lifetimes unless combined with small abundances. The data used to produce the curves in these figures were not made available along with the publication [103], but they were digitalised as part of the work on *An open window for high reheating temperatures in gravitino dark matter scenarios* [4]. In the collaboration on this project, the BBN module for the scan is implemented by Jeriek Van den Abeele.

Extending the gravitino relic density likelihood to consider also the non-thermal contribution, i.e. that from decaying neutralino NLSPs, and adding a likelihood function representing BBN bounds, makes it significantly harder to find

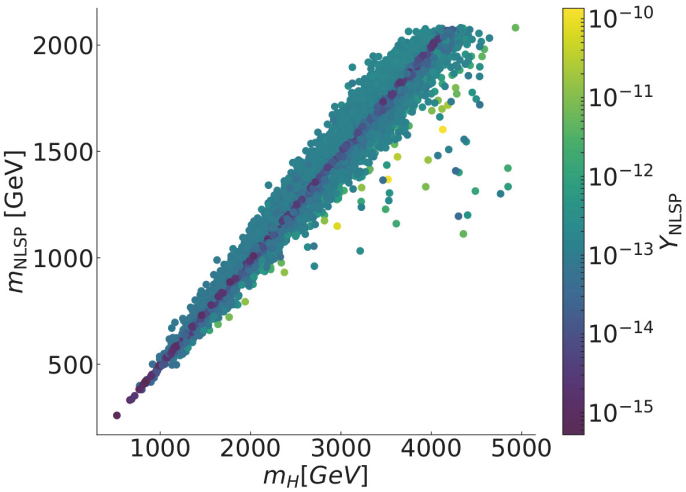


Figure 7.5: The mass of the heavy neutral Higgs versus the NLSP mass, with the colour indicating the NLSP yield, which becomes lower in the Higgs funnel region. The plot is produced using `Softsusy` and `MicrOMEGAS`.

allowed regions in parameter space featuring a high reheating temperature.

This is where the Higgs funnel region becomes important: Targeting the region described in eq. (7.8) yields significantly lower neutralino NLSP relic densities, see fig. 7.5. Here, the yield Y represents the comoving particle density, related to the relic density via

$$\Omega = \frac{\rho_0}{\rho_c} = \frac{mYs_0}{\rho_c}, \tag{7.11}$$

where ρ_0 is the density a particle would have had it not decayed, ρ_c is again the critical density and s_0 the entropy density of the universe at present. Numerical values are taken from [76].

The funnel region opens up a window for keeping reheating temperatures of $T_R \gtrsim 10^9$ GeV, when including also constraints from BBN, see fig. 7.6.

7.3.3 The final frontier: Collider limits

The only way to be sure that a model, here meaning a point in parameter space, is not excluded from a collider perspective, is to generate events for it and compare the simulation to experimental results. This is a time-consuming and computationally expensive process. Due to this, and because the BBN constraints described earlier are so strict, the computational strategy in the scan searching for models with a high reheating temperature, is to check collider constraints last, and only after cosmological constraints have been satisfied.

The collider module itself is divided into four parts which must be passed independently, and which each contribute to the total likelihood. If a likelihood contribution corresponds to a $\chi^2_{n=1}$ value exceeding $1.96\sigma_{\chi^2}$, corresponding to a 95% confidence level, the point is excluded. If the point is not excluded, the next check is invoked. The first of the collider checks deems a point worthy of proceeding if

- All physical gaugino masses are larger than 350 GeV

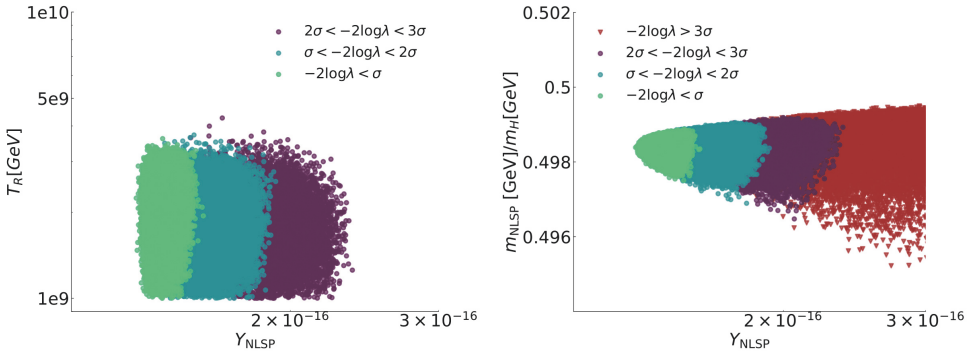


Figure 7.6: (a) Reheating temperatures above 10^9 GeV seem to be within reach even after taking into account constraints from BBN. The 1σ, 2σ and 3σ regions, indicated in green, blue and purple respectively, for the combined likelihood from gravitino relic density and BBN constraints, is found where the NLSP yield is low. (b) The NLSP yield is low when the NLSP mass is slightly below half that of the heavy Higgs. Here, some of the points beyond 3σ are again included, to illustrate how the likelihood decreases when moving away from the funnel region. The plots are produced using `Softsusy` and `MicrOMEGAS`.

- No gaugino lifetimes exceed 3.3×10^{-3} ns — corresponding to 1 mm propagation distance before decay.
- The combined gluino and chargino masses are located outside a polygon spanned by the exclusion limits from [113], figure 13b.

Any point which passes these constraints proceeds to the next check, which is based on CMS data from direct chargino searches. This module interpolates grid data on cross section exclusion limits from [114] to find an upper limit for the chargino production cross section from weak processes, σ_{limit}, based on the mass and lifetime of the lightest chargino.

The number of expected signal events from a direct search is

$$n_{signal} = n_{limit} \frac{\sigma_{signal}}{\sigma_{limit}}, \tag{7.12}$$

where σ_{signal} here represents the total chargino production cross section from weak processes, calculated by `MicrOMEGAS` for each parameter point. For the CMS search considered, $n_{bkg} = 6.5$ background events were expected, so a 95% confidence level puts the upper limit at $n_{limit} = 7.39$. These numbers define a Poisson likelihood function, see section 3.4.3 or eq. (B.9), used again to determine whether to exclude the point or to proceed.

While the aforementioned two collider checks are quick and simple, and only compare certain physical observables, the two remaining ones come closer to confronting the entire model with experimental results.

SModelS

`SModelS` [115] is "an automatic, public tool for interpreting simplified-model results from the LHC". It interprets LHC predictions from beyond Standard Model theories — including of course supersymmetry — in light of experimental constraints, by decomposing the model's predicted collider signatures into simplified model spectra, hence the name.

A simplified model is defined by an effective Lagrangian which describes a small number of particles and their interactions. The theoretical parameter space of supersymmetric models is immense, and simplified models reduce

the dimensionality to two to four sparticle masses and relevant branching ratios. The simplified model thus provides a framework for immediately evaluating searches, as these effective Lagrangian parameters are either themselves or directly related to collider physics observables.

The simplified model approach has been adopted by the collaborations ATLAS [116] and CMS [117], meaning that many current limits on sparticle masses are based on severe model assumptions, e.g. the LSP and NLSP masses along with some mass hierarchy, and a branching ratio of one for a specific process.

While a simplified model represents the limit of a more general model, the interpretation of a search result within a more general MSSM is, although possible, computationally expensive and not straightforward. This is where `SModelS` comes in: This python software decomposes a supersymmetric spectrum, as given by a spectrum generator, into simplified model topologies. These can then be relatively quickly tested against the limits set by ATLAS and CMS on simplified model spectra.

A simplified model's topology is described only by its shape and final states, the masses of the particles involved, and the cross section times branching ratio. Spins and colours of all involved particles, as well as all intermediate states are ignored. In this way, the complete model's signal is described in terms of simplified model elements, whose cross sections times branching ratios can be directly compared to simplified model limits from experiments. Finally, `SModelS` records the elements not covered by analyses as "missing topologies" cross sections, σ_{miss}.

The next part of the scan's collider module runs `SModelS` and uses its main outputs σ_{miss} and the r-value for each search in the `SModelS` database, defined as follows

$$r \equiv \frac{\text{Model prediction}}{\text{Upper limit}}\,. \tag{7.13}$$

`SModelS` has two ways of calculating this ratio: In the one case, it compares the simplified model prediction directly to the 95% confidence level limits on cross sections times branching ratio as provided by experimental collaborations. This is referred to by `SModelS` as upper limit, short UL, type results.

The other way involves so-called efficiency map, short EM, type results, in which the total number of signal events, i.e. cross section times branching ratio and efficiency including acceptance, is constrained inside a signal region. For each model and each signal region, `SModelS` uses a grid of simulated acceptance times efficiency values, which must have been provided by either an experimental collaboration or theory group, to calculate the actual number of signal events predicted by the model.

This means that internally, `SModelS` needs only one dataset per UL-type result, but one for each signal region in the case of EM-results. [5] When calculating an EM type result, `SModelS` sums up all the r-values from the whole map within a signal region, and returns the most restrictive r-value from all signal regions. Furthermore, `SModelS` provides a likelihood for the data given a signal strength, which is not possible for UL type results.

Returning now to the scan, an keeping in mind that UL-type results do not come with a likelihood, a decision gate and effective likelihood is implemented at this point: A rectified parabola loglikelihood function, see a conceptual sketch in fig. 7.7, represents the likelihood contribution consistently for all points considered by `SModelS`. This function is flat and equal to zero up until $r = 0.9$, and from this point decreases in the form of a parabola defined such that $-2\log\mathcal{L}(r=1)$ is the 95% confidence level, equal to 3.84 for one degree of freedom.

The subsequent decision is based on the missing cross section σ_{miss} and the r-value as

$$\text{if} \quad r > 1 \quad \Rightarrow \text{Excluded} \tag{7.14}$$

$$\text{elif} \quad \left.\begin{matrix} \sigma_{\text{miss}}\,\mathcal{L} < 3 \\ r < 1 \end{matrix}\right\} \quad \Rightarrow \text{Allowed} \tag{7.15}$$

$$\text{else} \quad \left.\begin{matrix} \sigma_{\text{miss}}\,\mathcal{L} \geq 3 \\ r < 1 \end{matrix}\right\} \quad \Rightarrow \text{needs further check} \tag{7.16}$$

using as integrated luminosity $\mathcal{L} = 39\,\text{fb}^{-1}$, corresponding to the value presently implemented in `CheckMATE`, and hence the limit of the subsequent and final check.

[5] Another `SModelS` technicality is that UL-type results can be generalised as EM-type results by defining the efficiency as one for a given element if the element appears in the UL constraint, and otherwise as zero. Although trivial, this allows the two types of results to be treated similarly [115].

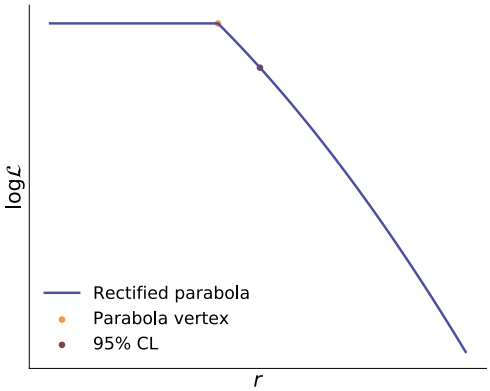


Figure 7.7: The rectified parabola function, constructed by defining a vertex at which the function changes from a straight line to a parbola, and one point on the parabola. In the present case, the vertex is at $r = 0.9$ with $(\log\mathcal{L})_{\max} = 0$, and the point on the parabola at $r = 1$ such that $-2\log\mathcal{L}(r = 1) = 3.84$.

To summarise: If the predicted cross section for a model exceeds the upper limit, the point is definitely excluded. If the r-value is inconclusive, but the missing cross section times the integrated luminosity results in fewer than three expected events, the point is allowed in the sense that it is impossible to exclude from current experimental limits. Lastly, if the r-value is inconclusive, but the number of expected events exceeds 3, the point requires further scrutiny.

This leads the discussion to the final part of the scan's collider module. In cases where tools using simplified models are insufficient, the last resort is to recast, meaning following as closely as possible the same procedure as the experimental collaborations. The first step is then to generate the events predicted by the model, and next performing a detector simulation on these. The objects returned by a detector simulation typically contain reconstructed photons, electrons, muons, jets and tracks, and the missing transverse momentum vector. These are then subjected to the same cuts as used in the experimental analysis for the corresponding search, and the limits on the observed number of events are applied in the relevant signal regions [118].

CheckMATE

The recasting tool used in the present project as well as in [1], is `CheckMATE` [118], and a significant update to the program was released in the time between these two projects. The main improvements from `CheckMATE` versions 1 to 2 are the implementation of many more experimental results, together with the integrated event generation by either `Pythia` [12] or `MadGraph` [13]. The latter means that the user can directly input a spectrum file to the program, and does not have to first generate large event files separately, store them and feed them to `CheckMATE` in a separate process. This greatly simplifies the workflow and potentially vastly decreases the need for storage space.

The workflow in this last part of the collider module is to provide the spectrum file to `CheckMATE`, which uses `MadGraph` to calculate matrix elements from the relevant Feynman diagrams, `Pythia` for hadronization, and lastly `Delphes` [119] for detector simulation, the latter being a time-consuming computational process. Thereafter, the detector level objects are passed to an internal method called an `AnalysisHandler`, whose responsibility is to impose quality requirements on each detector object and provide the necessary information to each analysis comparison.

Finally, `CheckMATE` collects the expected number of signal events from the event generation, and compares these to each experimental search result. The uncertainty on the expected number of signal events consists of the statistical

uncertainty from the event generation itself, and the systematic uncertainty on the estimated signal cross section. The former means that a larger number of simulated events yields smaller statistical uncertainty in the result, which in turn implies larger time consumption.

The main output of `CheckMATE` is whether a model is excluded or allowed, according to the highest sensitivity signal region of the analyses considered. Along with this output, `CheckMATE` also provides an r-value, defined in the same way as above. For consistency within the scan setup, this r-value is used in the same effective likelihood function as for `SModelS`.

The combined likelihood a point receives from the total relic density, BBN constraints and collider limits, represents how likely it is that it represents a model which may be realised in nature. The preliminary result, and what will hopefully form the grounds of the conclusion in *An open window for high reheating temperatures in gravitino dark matter scenarios* [4], is that the Higgs funnel region can indeed provide low enough neutralino yields to evade BBN bounds, in supersymmetric models not excluded by collider searches.

Conclusion

Throughout the past four years, during which I have been so fortunate to receive funding for literally trying answer questions, I have spent much time on supersymmetry and its phenomenology. Supersymmetry suffers from a complete lack of experimental support, along with all other candidate models for beyond Standard Model physics. The argument against supersymmetry and her siblings often goes as "We have been unable to verify supersymmetric models for the past decades, and tweaking model parameters only renders the model less natural".

A scientific motivation for naturalness would, from a frequentist point of view, require many universes whose physical constants we could observe, and the subsequent observation that some parameters are more common. This would give us a statistical motivation to expect those parameters in our own universe. Unfortunately, we do not have many universes from which to draw; this one seems to be all we have. Fortunately, frequentism is *not* all we have. From a Bayesian point of view, independent model evaluation does not make sense, while model *comparison* does. And compared to the Standard Model, there are quite a few supersymmetric models which look at least as, if not more, promising. My aim is neither to blandish nor disparage supersymmetric models or other beyond Standard Model theories, but rather to emphasize that in the present situation, the importance lies in developing powerful statistical methods for evaluating models and making sense of all the data particle physicists are fortunate enough to have access to.

In the first publication in this thesis, *Trilinear-augmented gaugino mediation* [1], we consider phenomenological constraints on the gaugino mediation model of supersymmetry breaking. We verify that the model allows for soft trilinear scalar interaction terms, which were previously assumed to vanish in this model. It turns out that these trilinear terms are crucial for achieving a Higgs mass of 125 GeV. The parameter space of this model with non-vanishing trilinear terms and non-universal Higgs masses is explored, and the identity of the NLSP throughout the parameter space is mapped. It turns out that the part of the parameter space with a neutralino NLSP is still unchallenged in terms of LHC sensitivity in missing energy searches.

The neutralino NLSP stays along for the ride into a later project: Chapter 7 describes how we investigate a scenario with gravitino dark matter, a neutralino NLSP and the possibility for achieving a high reheating temperature. This model is constrained by dark matter relic density observations, BBN bounds and collider searches, from which a likelihood is formed. This likelihood guides a parameter scan investigating the model's parameter space, and the computational challenges are significant. The project is still ongoing, under the working title *An open window for high reheating temperatures in gravitino dark matter scenarios* and the preliminary conclusion is that there is a possibility for achieving a sufficiently low neutralino yield in the region where the neutralino annihilates resonantly via a heavy Higgs, to accommodate cosmological constraints.

The topic of dark matter is also studied from a more model-independent viewpoint in chapter 6, in the form of gamma-ray signatures from dark matter annihilation into heavy mesons. Primarily excited meson states states which decay into their ground state by radiating off a monochromatic photon, can give rise to a signature which stands out from the astrophysical background. This project provided a contribution to *Smoking gun dark matter signatures in the MeV range* in the publication *Science with e-ASTROGAM: A space mission for MeV-GeV gamma-ray astrophysics* [3], and forms the basis for an ongoing project with working title *Gamma rays from the annihilation of dark matter into heavy mesons*.

During my work in particle physics, I have come across, and done my very best to learn, a method for statistical analysis which has impressed me in its power and efficiency. Machine learning algorithms can reduce the complexity of

and find new features in large datasets, which, when used correctly, can contribute significantly to solving problems in particle physics.

The field of machine learning has received an increased level of interest, to put it mildly, and advanced significantly over the past years. This is of course due to several factors, but one stands out: The improvement in computer hardware, which has opened up possibilities for fast calculations that were simply not there when the first machine learning algorithms were developed.

In the publication *Signal mixture estimation for degenerate heavy Higgses using a deep neural network* [2], we demonstrate that deep neural networks can discriminate between severely overlapping signal signatures. The result in this context is to greatly reduce the uncertainty in signal mixture parameter estimation in a two-Higgs doublet model with mass-degenerate neutral Higgses.

Machine learning is also a central tool in the ongoing project described in section 5.3, in which the possibility for using a machine learning algorithm to detect signals from a supersymmetric scenario with a sneutrino NLSP is investigated, and the main challenge is related to the very small mixture parameters expected. The working title for this project is *Enabling sneutrino detection in weak signal scenarios using machine learning methods*, and we look forward to finishing it in the near future.

Paper I

Trilinear-augmented gaugino mediation

Published for SISSA by Springer

Received: *February 7, 2017*
Accepted: *April 15, 2017*
Published: *May 2, 2017*

Trilinear-augmented gaugino mediation

Jan Heisig,[a] **Jörn Kersten,**[b] **Nick Murphy**[c] **and Inga Strümke**[b]

[a] *Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, 52056 Aachen, Germany*
[b] *Department of Physics and Technology, University of Bergen, 5020 Bergen, Norway*
[c] *CP*^{3}*-Origins, University of Southern Denmark, 5230 Odense M, Denmark*

E-mail: heisig@physik.rwth-aachen.de, joern.kersten@uib.no, murphy@cp3.sdu.dk, inga.strumke@uib.no

Abstract: We consider a gaugino-mediated supersymmetry breaking scenario where in addition to the gauginos the Higgs fields couple directly to the field that breaks supersymmetry. This yields non-vanishing trilinear scalar couplings in general, which can lead to large mixing in the stop sector providing a sufficiently large Higgs mass. Using the most recent release of FeynHiggs, we show the implications on the parameter space. Assuming a gravitino LSP, we find allowed points with a neutralino, sneutrino or stau NLSP. We test these points against the results of Run 1 of the LHC, considering in particular searches for heavy stable charged particles.

Keywords: Supersymmetry Phenomenology

ArXiv ePrint: 1701.02313

doi:10.1007/JHEP05(2017)003

Contents

JHEP05(2017)003

1 Introduction

Gaugino mediation [1, 2] is a mechanism for mediating supersymmetry breaking in a setup with extra spacetime dimensions, which avoids flavor problems by suppressing the soft sfermion masses at a high-energy scale. The original version of the model also yields suppressed trilinear scalar couplings, which is unfortunate since the measured Higgs mass [3] then requires a unified gaugino mass of $m_{1/2} \gtrsim 3\,\text{TeV}$ and thus very heavy sparticles [4].

However, a simple extension of the scenario does allow for non-vanishing trilinears and thus a lighter sparticle spectrum [5]. The couplings arise proportional to Yukawa couplings and thus do not lead to problematic flavor violation. We will investigate this possibility in detail in section 2, demonstrating explicitly how the trilinear couplings can be obtained.

In section 3, we study the parameter space of the extended setup. We show that the non-zero trilinears make it possible to reach the observed Higgs mass with sparticle masses that are accessible at the LHC. In gaugino mediation the gravitino can be the lightest supersymmetric particle (LSP) [6], making it a viable dark matter candidate [7].[1] We assume this scenario, in which case the next-to-lightest sparticle (NLSP) can be a stau, a tau sneutrino or a neutralino [8]. We determine the corresponding parts of the parameter space and constrain them by a careful analysis of LHC searches using data of the complete Run 1, in particular searches for long-lived heavy charged particles, extending the analysis in [5].

[1] Alternatively, another superweakly interacting particle such as the axino could be the LSP.

2 Gaugino-mediated supersymmetry breaking

2.1 General setup

The present work considers one out of a class of higher-dimensional models. There are in general D spacetime dimensions, $D-4$ of which are compact with volume V_{D-4}. This size determines the energy scale $M_c \equiv (1/V_{D-4})^{\frac{1}{D-4}}$ needed to resolve the compact dimensions, referred to as the *compactification scale*. Fields can either live in the whole D-dimensional space referred to as the *bulk* or be localized on $3+1$-dimensional *branes* that are located at different positions in the extra dimensions. The D-dimensional Lagrangian is [9]

$$\mathcal{L}_D = \mathcal{L}_{\text{bulk}}\left(\hat{\Phi}(x,y)\right) + \sum_j \delta^{(D-4)}(y-y_j)\,\mathcal{L}_j\left(\hat{\Phi}(x,y_j),\phi_j(x)\right), \tag{2.1}$$

where j runs over the branes, x are coordinates on the branes, y are coordinates in the bulk, $\hat{\Phi}$ is a bulk field[2] and ϕ_j is a field localized on the jth brane. Hats denote bulk fields with canonically normalized kinetic terms in D dimensions.

We consider a model with two branes: the MSSM brane, where the visible matter fields are localized, and the hidden brane with a chiral superfield S, which is a singlet under the Standard Model (SM) gauge groups. Supersymmetry (SUSY) is broken by the vacuum expectation value (VEV) $\langle F_S\rangle$ of the auxiliary field of S. The gauge and Higgs superfields propagate in the bulk. Therefore, they can couple directly to the SUSY-breaking field and obtain soft masses proportional to $\langle F_S\rangle$. In contrast, sfermion soft masses are strongly suppressed due to the separation between the MSSM and hidden brane, which avoids unacceptably large flavor-changing neutral currents [1, 2].

2.2 Trilinear couplings

The supersymmetric part of the MSSM Lagrangian contains both bulk fields and fields constrained to the visible brane,

$$\begin{aligned}\mathcal{L}_{\text{MSSM}} &= \mathcal{L}_{\text{bulk}} + \delta^{(D-4)}(y-y_1)\,\mathcal{L}_1 \\ &= \left[W(\hat{\Phi},\phi_1) + \frac{1}{4}\hat{\mathcal{W}}_\alpha\hat{\mathcal{W}}^\alpha\right]_F + \text{h.c.} + \left[K\left(\hat{\Phi},\hat{\Phi}^\dagger,\phi_1,\phi_1^\dagger,e^V\right)\right]_D,\end{aligned} \tag{2.2}$$

where W is the visible-sector superpotential, $\hat{\mathcal{W}}$ the field strength superfield and K the Kähler potential. Using the notation of equation (2.1), we have $j=1$ for the visible brane and will accordingly use $j=2$ for the hidden brane. On this brane, the gauge and Higgs superfields interact with the hidden-sector field S,

$$\begin{aligned}\mathcal{L}_2 = &\frac{1}{M^{D-3}}\left[\frac{h}{4}S\hat{\mathcal{W}}^\alpha\hat{\mathcal{W}}_\alpha\right]_F + \text{h.c.} \\ &+ \frac{1}{M^{D-3}}\left[S\left(a\hat{H}_u^\dagger\hat{H}_d^\dagger + b_u\hat{H}_u^\dagger\hat{H}_u + b_d\hat{H}_d^\dagger\hat{H}_d\right) + \text{h.c.}\right]_D \\ &+ \frac{1}{M^{D-2}}\left[S^\dagger S\left(c_u\hat{H}_u^\dagger\hat{H}_u + c_d\hat{H}_d^\dagger\hat{H}_d + (d\hat{H}_u\hat{H}_d + \text{h.c.})\right)\right]_D + \dots,\end{aligned} \tag{2.3}$$

[2] Strictly speaking, we use superfields of 4D $N=1$ supersymmetry. The higher-dimensional supersymmetry requires additional fields, which we do not write explicitly, since they are not relevant here.

JHEP05(2017)003

where h, a, $b_{u,d}$, $c_{u,d}$ and d are dimensionless couplings. The dots refer to terms containing only hidden-sector fields. Setting $b_{u,d} = 0$ reduces the present case to the one considered in [2]. Setting also $a = c_{u,d} = d = 0$, i.e., not placing the Higgs fields in the bulk, reduces our case to the one in [1]. Note that the localizations of S and the sfermions forbid terms like $S\bar{u}\hat{H}_u Q$ and $SQ^\dagger Q$, which would directly yield trilinear couplings and sfermion soft masses.

Interactions between the bulk fields and the hidden-sector field are non-renormalizable, so $\mathcal{L}_D$ describes an effective theory valid up to some fundamental scale M. To obtain the 4-dimensional effective theory valid below the compactification scale, we integrate over the extra dimensions and keep only the zero modes of the bulk fields, which are constant in the extra dimensions. The integration yields a volume factor V_{D-4} in the kinetic terms of the bulk fields, so we define fields with canonical kinetic terms in 4D by $\Phi \equiv \sqrt{V_{D-4}}\hat{\Phi}$. Thus, the part of the effective 4D Lagrangian describing the interactions of S with the visible sector is

$$\begin{aligned}\mathcal{L}_{D=4} \supset \frac{1}{V_{D-4}} \Bigg\{ & \frac{1}{M^{D-3}} \left[\frac{h}{4} S \mathcal{W}^\alpha \mathcal{W}_\alpha \right]_F + \text{h.c.} \\ & + \frac{1}{M^{D-3}} \left[S \left(a H_u^\dagger H_d^\dagger + b_u H_u^\dagger H_u + b_d H_d^\dagger H_d \right) + \text{h.c.} \right]_D \\ & + \frac{1}{M^{D-2}} \left[S^\dagger S \left(c_u H_u^\dagger H_u + c_d H_d^\dagger H_d + (d H_u H_d + \text{h.c.}) \right) \right]_D \Bigg\} . \end{aligned} \tag{2.4}$$

The first term generates gaugino masses [1, 2]. We assume a unified gauge theory above the compactification scale, so that there is a unified gaugino mass $m_{1/2}$. The remaining terms produce the $B\mu$-term, soft Higgs masses $m_{H_u}^2$ and $m_{H_d}^2$, and a contribution to the μ-term [2].

The terms proportional to b_u and b_d, which were not included in the original versions of gaugino mediation [1, 2], contribute to the soft Higgs masses and $B\mu$ as well. Most importantly, however, they yield trilinear scalar couplings [5]. This can be seen by absorbing them via the field redefinitions $H'_{u,d} \equiv H_{u,d}\left(1 + b_{u,d}\frac{S}{M}\right)$, from the general expressions for soft SUSY-breaking terms in the supergravity formalism, see e.g. [10, 11], or by integrating out the Higgs auxiliary fields. We find it instructive to show the latter calculation for our particular case.

First, the part of the Lagrangian (2.2) that contains the Higgs supermultiplets' auxiliary fields $F_{H_{u,d}}$ is

$$\begin{aligned}\mathcal{L}_{\text{MSSM}} \supset F_{H_u}^\dagger F_{H_u} + F_{H_d}^\dagger F_{H_d} + (\phi_{\bar{u}} y_u F_{H_u} \phi_Q - \phi_{\bar{d}} y_d F_{H_d} \phi_Q - \phi_{\bar{e}} y_e F_{H_d} \phi_L \\ + \mu F_{H_u} \phi_{H_d} + \mu \phi_{H_u} F_{H_d} + \text{h.c.}) \, , \end{aligned} \tag{2.5}$$

where ϕ_X denotes the scalar component of the superfield X. Adding the D-terms from equation (2.4) and employing the equations of motion $\partial\mathcal{L}/\partial F_{H_{u,d}}^\dagger = 0$ yields

$$F_{H_{u,d}} = -\frac{1}{V_{D-4} M^{D-3}} \left(b_{u,d} F_S \phi_{H_{u,d}} + b_{u,d} \phi_S F_{H_{u,d}} + b_{u,d}^* \phi_S^* F_{H_{u,d}} \right) + \dots \, , \tag{2.6}$$

where we have omitted terms that do not contribute to SUSY-breaking trilinears.[3] The

[3] Note that the term proportional to a contributes to the *supersymmetric* (scalar)^3 couplings. If the scalar component of S develops a VEV, the terms proportional to $c_{u,d}$ also contribute to the trilinears, but this contribution can be absorbed by a redefinition of $b_{u,d}$.

JHEP05(2017)003

solutions are thus

$$F_{H_{u,d}} = -\frac{\frac{b_{u,d} F_S \phi_{H_{u,d}}}{V_{D-4} M^{D-3}}}{1 + b_{u,d} \frac{\phi_S}{V_{D-4} M^{D-3}} + b^*_{u,d} \frac{\phi^*_S}{V_{D-4} M^{D-3}}} + \dots = -b_{u,d} \left(\frac{M_c}{M}\right)^{D-4} \frac{F_S}{M} \phi_{H_{u,d}} + \dots , \tag{2.7}$$

omitting irrelevant higher-order terms in ϕ_S and replacing the extra dimensions' volume by the compactification scale in the last step. Substituting F_{H_u} and F_{H_d} into the Lagrangian (2.5) and replacing F_S by its VEV finally gives rise to the desired trilinear terms,

$$\mathcal{L}_{\text{trilinear}} = \left(\frac{M_c}{M}\right)^{D-4} \frac{\langle F_S \rangle}{M} \left(-b_u \phi_{\bar{u}} y_u \phi_{H_u} \phi_Q + b_d \phi_{\bar{d}} y_d \phi_{H_d} \phi_Q + b_d \phi_{\bar{e}} y_e \phi_{H_d} \phi_L + \text{h.c.}\right). \tag{2.8}$$

Consequently, we obtain trilinear scalar couplings proportional to the SUSY-breaking VEV and the Yukawa matrices,

$$a_u = A_{u0}\, y_u \quad , \quad a_d = A_{d0}\, y_d \quad , \quad a_e = A_{d0}\, y_e \tag{2.9}$$

with

$$A_{u0} = \left(\frac{M_c}{M}\right)^{D-4} \frac{\langle F_S \rangle}{M} b_u \quad , \quad A_{d0} = \left(\frac{M_c}{M}\right)^{D-4} \frac{\langle F_S \rangle}{M} b_d \,. \tag{2.10}$$

Due to the proportionality of trilinear matrices and Yukawa matrices in the relations (2.9), these matrices are simultaneously diagonalized when changing to the super-CKM basis. Although the running to low energies leads to deviations from the exact proportionality, they are small enough to suppress flavor-changing neutral currents below the experimental upper limits.

Interestingly, the proportionality factors A_{u0} for the up-type squarks and A_{d0} for the down-type squarks and charged sleptons are different in general, in contrast to other simple setups for SUSY breaking like the Constrained MSSM or non-universal Higgs mass (NUHM) scenarios [12]. In the following we will restrict ourselves to the simplest possibility $A_{u0} = A_{d0} \equiv A_0$.

2.3 Constraints from naïve dimensional analysis

We will now estimate an upper limit on the trilinears, arguing that the couplings between the hidden-sector brane field S and the bulk fields can be constrained by naïve dimensional analysis (NDA) [9]. This discussion generalizes results of [8], where the specific case of a 6-dimensional model was considered, to an arbitrary number of dimensions.

We write the Lagrangian (2.4) in terms of dimensionless fields $\breve{H}_{u,d}$ and $\breve{S}$ defined by

$$H_{u,d} = \left(\frac{M^{D-2} V_{D-4}}{l_D / C}\right)^{1/2} \breve{H}_{u,d} \quad , \quad S = \left(\frac{M^2}{l_4 / C}\right)^{1/2} \breve{S} \,, \tag{2.11}$$

where $l_D = 2^D \pi^{D/2} \Gamma(\frac{D}{2})$ is the factor suppressing one-loop diagrams in D dimensions, and C is a group theory factor depending on the unified theory valid above M_c. The volume

JHEP05(2017)003

factor V_{D-4} ensures canonical kinetic terms in 4D for the zero modes of the bulk fields. In this way, we obtain for the part of the Lagrangian coupling S to the Higgs fields

$$\mathcal{L}_{D=4} \supset \frac{M^2}{l_4/C}\left\{\frac{\sqrt{Cl_4}}{l_D}\left[\check{S}\left(a\check{H}_u^\dagger\check{H}_d^\dagger + b_u\check{H}_u^\dagger\check{H}_u + b_d\check{H}_d^\dagger\check{H}_d\right) + \text{h.c.}\right]_D \right. \\ \left. + \frac{C}{l_D}\left[\check{S}^\dagger\check{S}\left(c_u\check{H}_u^\dagger\check{H}_u + c_d\check{H}_d^\dagger\check{H}_d + (d\check{H}_u\check{H}_d + \text{h.c.})\right)\right]_D\right\}. \tag{2.12}$$

According to NDA, the theory is weakly coupled below the cutoff scale M, if all couplings inside the curly brackets in equation (2.12) are smaller than one. This implies the constraints

$$\frac{\sqrt{Cl_4}}{l_D}\{|a|, |b_u|, |b_d|\} < 1\,, \\ \frac{C}{l_D}\{|c_u|, |c_d|, |d|\} < 1\,. \tag{2.13}$$

Combined with equation (2.10), they translate into the upper bound

$$|A_0| < \frac{\langle F_S\rangle}{M}\left(\frac{M_c}{M}\right)^{D-4}\frac{l_D}{\sqrt{Cl_4}} \tag{2.14}$$

on the trilinears. For comparison, the NDA constraint on the gaugino mass is [6]

$$m_{1/2} < \frac{\langle F_S\rangle}{M}\frac{1}{2}\left(\frac{M_c}{M}\right)^{D-4}\frac{l_D}{\sqrt{Cl_4}}\,. \tag{2.15}$$

Consequently, the ratio of the upper bounds is simply

$$\frac{|A_0|^{\max}}{m_{1/2}^{\max}} = 2\,. \tag{2.16}$$

If the limit on $m_{1/2}$ is saturated, it is thus possible for the trilinear couplings to be somewhat larger than the gaugino mass, but not by orders of magnitude.

3 Phenomenology of the model

Let us now explore the parameter space of gaugino mediation extended by trilinear couplings. As explained in section 2, the model contains the five free parameters $m_{1/2}$, $m_{H_u}^2$, $m_{H_d}^2$, A_0, and $B\mu$. The soft squark and slepton masses are negligibly small. This is a realization of the NUHM2 scenario [13] with the restriction $m_0 = 0$. These input parameters are boundary conditions at the compactification scale, which we identify with the scale of gauge coupling unification, $M_c \simeq 10^{16}\,\text{GeV}$. As usual, we trade $B\mu$ for $\tan\beta$ and use the measured Z mass to determine the absolute value of μ. We choose μ to be positive and restrict ourselves to negative values for A_0; changing the sign of both parameters would lead to a similar phenomenology.

One of the most important model restrictions is the Higgs mass required to match the value measured at the LHC, see section 3.1 for details. The allowed parameter space accommodates various choices of the lightest sparticle of the MSSM, discussed in section 3.2.

JHEP05(2017)003

It comprises the lightest neutralino, the tau sneutrino and the lighter stau. As the latter two are not phenomenologically viable dark matter candidates we assume here that the LSP is a non-MSSM sparticle with very weak interactions.[4] In the framework of supergravity, this could be the gravitino. In this case the lightest sparticle of the MSSM is the NLSP. Gaugino mediation allows for gravitino masses $m_{3/2} \gtrsim 10\,\text{GeV}$ [6], in which case the NLSP becomes stable on collider time-scales and the collider signature of the considered model vitally depends on the choice of NLSP. While a neutralino or sneutrino NLSP provides a signature containing missing transverse momentum, detector-stable staus provide a distinct signature of heavy stable charged particles (HSCPs), for which the LHC sensitivity is very high. LHC constraints for the respective signatures are discussed in section 3.3. Bounds from color or charge breaking minima of the scalar potential are briefly discussed in section 3.4. In section 3.5 we comment on the cosmological constraints on the model.

3.1 Higgs mass

One of the most important constraints on the parameter space is the experimentally observed Higgs mass of $125.09 \pm 0.24\,\text{GeV}$ [3]. The theoretical uncertainty of the Higgs mass prediction in the MSSM is on the order of $\sim 2\,\text{GeV}$ [15, 16]. As the theoretical error is large compared to the experimental one, we do not consider the latter. Furthermore, we assume that the lightest CP-even Higgs of the MSSM plays the role of the observed Higgs. Hence, we consider points with a theoretically predicted mass of the lightest CP-even Higgs in the rage $123\,\text{GeV} \lesssim m_h \lesssim 127\,\text{GeV}$ to be consistent with observations.

In order to compute the Higgs mass we proceed as follows. First we use SPheno 3.3.8 [17, 18] for the calculation of the sparticle masses and low-energy Lagrangian parameters. The output from SPheno is then used as input to FeynHiggs 2.12.2 [15, 16, 19–23], which we use to more accurately calculate the lightest Higgs pole mass. Both programs incorporate two-loop diagrams in the calculation of m_h. However, FeynHiggs 2.12.2 includes a more complete treatment of the calculation, including momentum dependent two-loop QCD contributions [16], leading three-loop contributions [15] and additionally, by combining an effective field theory approach with the fixed-order calculation, it incorporates up to NNLL contributions resummed to all orders [23]. This treatment can significantly reduce the theoretical uncertainties with respect to the pure fixed-order calculation, in particular for large $M_{\text{susy}} \equiv \sqrt{m_{\tilde{t}_1} m_{\tilde{t}_2}}$ [23, 24].

The result for the Higgs mass[5] is shown in figure 1, where the left panel shows the contour for which $m_h = 125.09\,\text{GeV}$ in the A_0-$m_{1/2}$ plane. The darker and lighter shaded

[4] For the case that a neutralino is the lightest sparticle of the MSSM it could itself be the LSP and hence identified with the dark matter particle. In this case constraints from direct and indirect detection as well as from the thermal relic density could be applied in order to narrow down the viable part of the parameter space. See e.g. [14] for a global fit within the (general) NUHM2 scenario taking into account dark matter observables for a neutralino LSP.

[5] We used the most recent results available in [25] for the Standard Model input parameters relevant for the scans. The values used in both SPheno and FeynHiggs are

$$G_F = 1.166379 \cdot 10^{-5}\,\text{GeV} \qquad m_b(m_b) = 4.18\,\text{GeV}\ (\text{SM}\,\overline{\text{MS}})$$
$$m_Z = 91.18760\,\text{GeV} \qquad m_\tau = 1.77686\,\text{GeV}$$
$$\alpha_s(M_z) = 1.181 \cdot 10^{-1}\ (\text{SM}\,\overline{\text{MS}}) \qquad m_t = 1.732 \cdot 10^2\,\text{GeV (pole mass)}.$$

JHEP05(2017)003

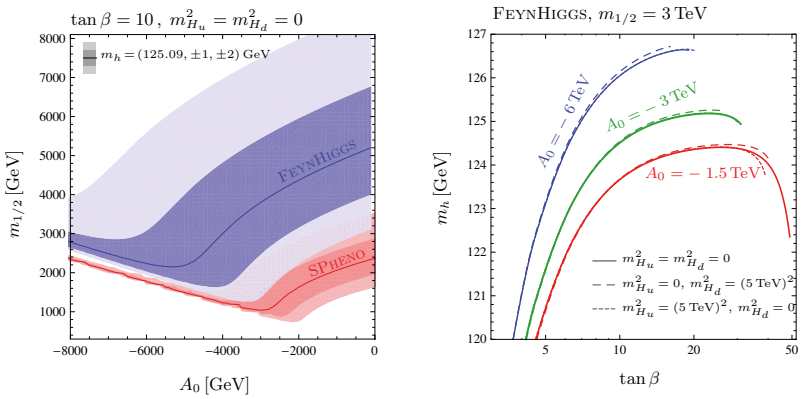


Figure 1. Left panel: contours of the Higgs mass computed by SPheno (red curve) and FeynHiggs (blue curve) in the A_0-$m_{1/2}$ plane. The solid lines denote the contour where $m_h = 125.09\,\text{GeV}$ whereas the corresponding darker and lighter shaded areas around them denote a deviation of ±1 and $\pm2\,\text{GeV}$, respectively. Right panel: dependence of the Higgs mass, m_h, computed by FeynHiggs, on $\tan\beta$ for $m_{1/2} = 3\,\text{TeV}$ and three choices of the trilinear coupling $A_0 = -1.5\,\text{TeV}$ (red curves), $A_0 = -3\,\text{TeV}$ (green curves), $A_0 = -6\,\text{TeV}$ (red curves) as well as for three choices of the Higgs soft mass parameters $m^2_{H_u} = m^2_{H_d} = 0$ (solid curves), $m^2_{H_u} = 0$, $m^2_{H_d} = (5\,\text{TeV})^2$ (long-dashed curves), $m^2_{H_u} = (5\,\text{TeV})^2$, $m^2_{H_d} = 0$ (short-dashed curves).

regions around it denote the ±1 and $\pm2\,\text{GeV}$ bands respectively. As mentioned above, we use the Higgs mass as computed by FeynHiggs, represented by the blue curve and bands on the plot. The right panel shows the Higgs mass dependence on $\tan\beta$, $m^2_{H_u}$ and $m^2_{H_d}$ for a fixed value of $m_{1/2}$ and three choices of A_0.

For $\tan\beta = 10$ and vanishing A_0, very large values of $m_{1/2}$ on the order of 6 TeV are needed to achieve a suitable Higgs mass of 125 GeV. With growing negative A_0, the required $m_{1/2}$ drops to a minimum around $m_{1/2} \simeq 2\,\text{TeV}$, beyond which the Higgs mass rises again. This minimum corresponds to the maximal mixing scenario, where $|X_t| = |A_t - \mu\cot\beta| \sim \sqrt{6}M_{\text{susy}}$, see [26] for a detailed discussion. This result shows that only with a non-zero trilinear coupling A_0, a Higgs mass of around 125 GeV can be obtained with $m_{1/2}$ such as to obtain a sufficiently light spectrum to be observable in upcoming collider experiments. See further discussion in section 3.3.

The ±1 and $\pm2\,\text{GeV}$ bands span a large range, reflecting the relatively large uncertainty in the required value of $m_{1/2}$ between 3 and 8 TeV. However, this uncertainty band shrinks significantly for large negative A_0.

The dependence on $\tan\beta$ is shown in the right panel of figure 1. Both very small and very large values of $\tan\beta$ cause the Higgs mass to drop drastically, making it hard to achieve the correct Higgs mass even for very large $m_{1/2}$. Note that for large $\tan\beta$ and large negative A_0, the spectrum acquires tachyonic states. Therefore, not all curves extend to $\tan\beta = 50$.

JHEP05(2017)003

The influence of the Higgs soft masses $m^2_{H_u}$ and $m^2_{H_d}$ on the Higgs mass is small throughout the explored parameter space. The most significant effect arises for large $\tan\beta$, cf. the solid and dashed curves in the right panel of figure 1.

The Higgs mass contour as computed by SPheno, presented by the red curve and shaded bands in the left panel of figure 1, is included for comparison.[6] The required Higgs mass is reached with considerably smaller $m_{1/2}$ for a given A_0, as the SPheno result for m_h is typically around 3 GeV larger than the one from FeynHiggs. In particular for large M_{susy}, NNLL resummation can yield important corrections that significantly contribute to the difference between the results obtained by the two codes, see e.g. [16, 23, 24] for details.

3.2 Particle spectrum

The phenomenology of the model regarding collider searches, astrophysics and cosmology strongly depends on the nature of the NLSP. As mentioned above, we compute the sparticle spectrum with SPheno. In the considered parameter space, we encounter three possible candidates for the NLSP: the neutralino, the sneutrino, or the lighter stau, which can be predominantly left- or right handed. Figure 2 shows several projections of the parameter space in the plane $m^2_{H_d}/m^2_{1/2}$-$A_0/m_{1/2}$. We have rescaled m_{H_d} and A_0 by $m_{1/2}$ as the nature of the NLSP is almost independent of the overall mass scale that is governed mostly by $m_{1/2}$. In other words, for fixed ratios $A_0/m_{1/2}$, $m^2_{H_d}/m^2_{1/2}$ and $m^2_{H_u}/m^2_{1/2}$, the sparticle spectrum is mainly shifted with $m_{1/2}$ and the shown projections remain approximately unchanged.

The results summarized in figure 2 demonstrate the relationship between the Higgs soft masses and the NLSP. As the ratio $r \equiv (m^2_{H_u} - m^2_{H_d})/m^2_{1/2}$ becomes more negative, the NLSP can shift from the stau, to the neutralino and finally to the sneutrino, depending on the value of $\tan\beta$ and A_0. If $\tan\beta$ is relatively large and A_0 is large and negative, only a stau NLSP is possible. Interestingly, the stau NLSP is also observed to shift through regions of right-chirality, large mixing and left-chirality with decreasing r (cf. the gray solid curve in the plots of figure 3, showing the stau mixing angle). In addition, figure 2 depicts the NLSP sensitivity to the value of $\tan\beta$, showing that the stau NLSP region grows with $\tan\beta$. In fact, for $\tan\beta \gtrsim 30$, the entire region contains only a stau NLSP. We also find that some of the regions of interest contain unphysical tachyonic spectra, meaning negative soft-masses squared. This occurs when A_0 has a large negative value compared to $m_{1/2}$, and becomes more frequent with increasing $\tan\beta$.

We would like to explain some of this behavior in a rough analytical manner, beginning with the chirality switch of the stau. This can be understood from analyzing the one-loop RGE's for the third generation leptonic soft masses [11]

$$16\pi^2 \frac{d}{dt} m^2_{L_3} = \chi_\tau - 6g_2^2|M_2|^2 - \frac{6}{5}g_1^2|M_1|^2 - \frac{3}{5}g_1^2\Sigma \tag{3.1a}$$

$$16\pi^2 \frac{d}{dt} m^2_{\tilde{e}_3} = 2\chi_\tau - \frac{24}{5}g_1^2|M_1|^2 + \frac{6}{5}g_1^2\Sigma, \tag{3.1b}$$

[6] For definiteness we also show ± 2 GeV bands for the SPheno predicition. However, the actual uncertainty might be larger [24].

JHEP05(2017)003

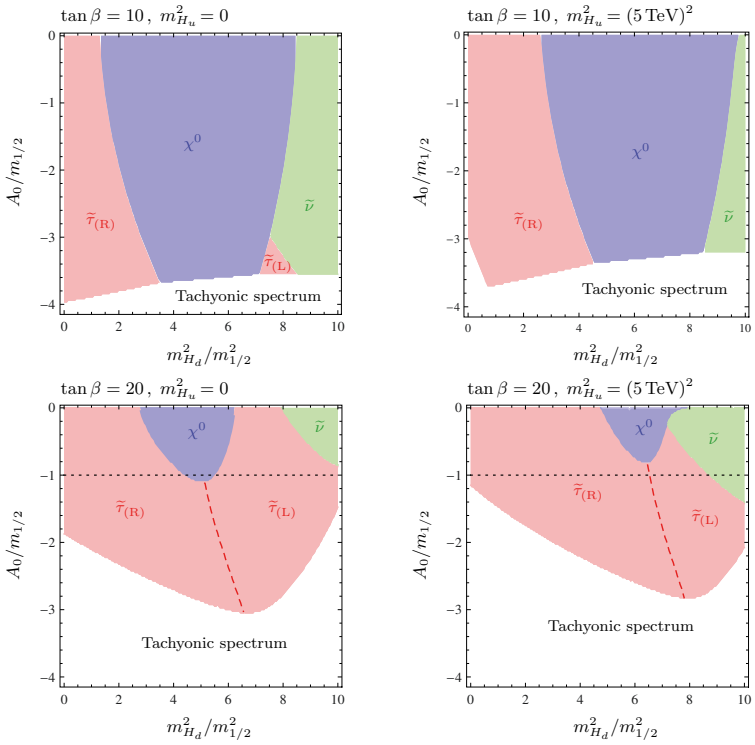


Figure 2. Regions characterized by a stau (red), neutralino (blue) and sneutrino (green) NLSP in the $m^2_{H_d}/m^2_{1/2}$-$A_0/m_{1/2}$ plane for four choices of $\tan\beta$ and $m^2_{H_u}$. All panels have $m_{1/2} = 2\,\text{TeV}$. In the white region below, we run into a tachyonic region, i.e., negative soft masses squared. The red dashed curve indicates the transition from a predominantly right- to left-handed stau NLSP, i.e., the contour $\sin^2\theta_\tau = 1/2$. The black dotted lines in the lower plots denote the slices in parameter space that are considered in figure 3.

where

$$\begin{aligned}\chi_\tau &\equiv 2|y_\tau|^2(m^2_{H_d} + m^2_{L_3} + m^2_{\bar{e}_3}) + 2|a_\tau|^2 \\ \Sigma &\equiv m^2_{H_u} - m^2_{H_d} + \text{Tr}[m^2_Q - m^2_L - 2m^2_{\bar{u}} + m^2_{\bar{d}} + m^2_{\bar{e}}].\end{aligned} \tag{3.2}$$

For $m^2_{H_d} \gg m_{1/2}$, we can neglect the gaugino masses in the above formula, and the running will depend mostly on the Σ parameter. From equation (3.2), one sees that for very large $m^2_{H_d}$, this value is negative, and will therefore lower the value of the left-chiral soft mass term but increase the size of the right-chiral term. Therefore, the NSLP will become more left-chiral with increasing $m^2_{H_d}$. For larger values of $m^2_{H_u}$, the absolute value of the Σ term is smaller, and the progression from right- to left-chirality happens at larger values of $m^2_{H_d}$.

JHEP05(2017)003

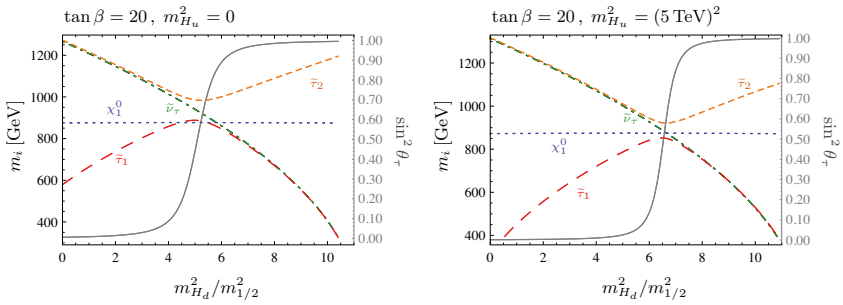


Figure 3. Sparticle masses $m_{\widetilde{\tau}_1}$ (red long-dashed curve), $m_{\widetilde{\tau}_2}$ (orange short-dashed curve), $m_{\widetilde{\nu}_\tau}$ (green dot-dashed curve) and $m_{\chi_1^0}$ (blue dotted curve) as a function of $m^2_{H_d}/m^2_{1/2}$ for two choices of $\tan\beta$ and $m^2_{H_u}$. The stau mixing angle is indicated through the gray solid line showing $\sin^2\theta_\tau$, labelled on the right axis.

Regions where the sneutrino becomes the LSP are also determined by equations (3.1). Again, these regions occur in the limit $m^2_{H_d} \gg m_{1/2}$, so we can make the same approximation and assume that the stau is mostly left-chiral. When the stau is mostly left-chiral, it is a delicate matter which of the two particles becomes the NLSP. The sneutrino mass is completely determined by equation (3.1a), as there are no right-chiral neutrinos in the MSSM, whereas there is mixing in the stau sector. The off-diagonal elements in the stau mixing matrix, which are A_0 and $\tan\beta$ dependent, push the eigenvalue down. However, the diagonal elements, which are predominantly dependent on the soft masses $m^2_{L_3}$ and $m^2_{\bar{e}_3}$, but also depend on the "hyperfine splitting" arising from EWSB, increase the eigenvalues. In figure 3 we show the masses of the staus, the tau sneutrino and the neutralino for the two slices denoted by the black dotted lines in the lower panels of figure 2. It reveals the small mass difference between $\widetilde{\tau}_1$ and $\widetilde{\nu}_\tau$ for large $m^2_{H_d}/m^2_{1/2}$.

The $\tan\beta$ and A_0 dependence can be understood by first noting that the neutralino mass is pushed up with $\tan\beta$, and larger values of A_0 push the third generation leptonic soft masses down by increasing χ_τ. This explains the shrinking neutralino region seen in the lower panels of figure 2. Large values of A_0 also increase mixing in the stau sector, pushing down the smallest eigenvalue of the stau mass matrix, implying the sneutrino LSP region should also shrink with larger $\tan\beta$.

3.3 Tests at colliders

Heavy stable charged particles. The lighter stau is the NLSP for a large part of the considered parameter space in our model. In order to determine the 95% CL exclusion limits from collider searches for HSCPs, we first compute the total cross section for the production of sparticles with Pythia 6 [27]. For points with $\sigma^{\text{tot}}_{8\,\text{TEV}} > 1/\mathcal{L}^{\text{int}}_{8\,\text{TEV}}$, i.e. for an expected total signal of more than one event we perform a Monte Carlo simulation of the signal at the 8 TeV LHC with the MadGraph5_aMC@NLO event generator [28]. We generate 10 k events for each point in the model parameter space, taking into account

JHEP05(2017)003

all possible sparticle production channels. The decay, showering and hadronization is performed with Pythia 6 [27]. We do not perform a detector simulation. Instead we determine the signal efficiencies with the method introduced in ref. [29], which allows for the direct analysis of the hadron-level events on the basis of the kinematic properties of isolated HSCP candidates. In order to identify isolated HSCP candidates we first impose the isolation criteria

$$\left(\sum_{i}^{\substack{\text{charged particles} \\ \Delta R<0.3}} {p_{\mathrm{T}}}^{i} \right) < 50\,\mathrm{GeV} \tag{3.3}$$

and

$$\left(\sum_{i}^{\substack{\text{visible particles} \\ \Delta R<0.3}} \frac{E^i}{|\boldsymbol{p}|} \right) < 0.3\,, \tag{3.4}$$

where the sums include all charged and visible particles, respectively, in a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ around the direction of the HSCP candidate, ${p_{\mathrm{T}}}^{i}$ denotes their transverse momenta and E^i their energy. Muons are not considered as visible particles as their energy deposition in the calorimeter is small. $|\boldsymbol{p}|$ is the magnitude of the three-momentum of the HSCP candidate. The HSCP candidate itself is not included in either sum.

We compute the signal efficiency by averaging the probabilities for events to pass the on- and off-line selection criteria [29],

$$\epsilon = \frac{1}{N} \sum_{i}^{N} P^{(n)}_{\mathrm{on},\,i} \times P^{(n)}_{\mathrm{off},\,i}\,, \tag{3.5}$$

where the sum runs over all N generated events i. For events containing one or two HSCP candidates the probabilities are given by

$$P^{(1)}_{\mathrm{on/off},\,i} = P_{\mathrm{on/off}}(\boldsymbol{k}^1_i) \tag{3.6}$$

or

$$P^{(2)}_{\mathrm{on/off},\,i} = P_{\mathrm{on/off}}(\boldsymbol{k}^1_i) + P_{\mathrm{on/off}}(\boldsymbol{k}^2_i) - P_{\mathrm{on/off}}(\boldsymbol{k}^1_i) P_{\mathrm{on/off}}(\boldsymbol{k}^2_i)\,, \tag{3.7}$$

respectively, where $\boldsymbol{k}^{1,2}_i$ are the kinematical vectors of the HSCP candidates in the ith event. $\boldsymbol{k} = (\eta, p_{\mathrm{T}}, \beta)$ contains the candidate's pseudo-rapidity, η, transverse momentum, p_{T}, and velocity, β.

The CMS analysis [29] requires a minimum reconstructed mass, m_{rec}, for the candidate. The probabilities $P_{\mathrm{on/off}}(\boldsymbol{k})$ are provided for four distinct mass cuts

$$m_{\mathrm{rec}} > 0,\ 100,\ 200,\ 300\,\mathrm{GeV}\,,$$

which we here consider to be four different signal regions. Due to detector resolution effects, the reconstructed mass is typically $m_{\mathrm{rec}} \simeq 0.6 m_{\mathrm{HSCP}}$ [29]. Hence, we set the efficiencies to zero if $0.6 m_{\mathrm{HSCP}}$ is below the respective mass cut of the signal region.

JHEP05(2017)003

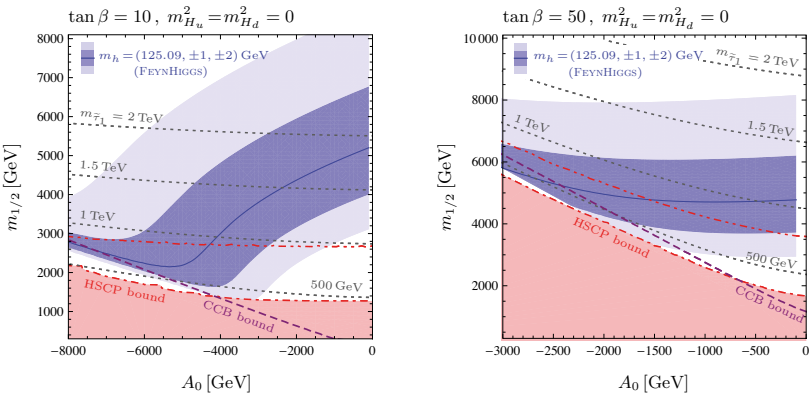


Figure 4. Contours of $m_h = 125.09\,\text{GeV}$ computed by FeynHiggs (blue solid curve) in the A_0-$m_{1/2}$ plane, as well as constraints from searches for heavy stable charged particles (HSCP) at the 8 TeV LHC (red shaded region below the red dot-dashed curve). Projections for the 13 TeV LHC at $300\,\text{fb}^{-1}$ are indicated by the red dot-dot-dashed curve. The purple dashed line represents the strongest of the CCB constraints from equations (3.9)–(3.10). The grey dotted curves show the contours of the lighter stau mass $m_{\tilde{\tau}_1}$. For $\tan\beta = 50$ and $-A_0 \gtrsim 2.3\,\text{TeV}$ the HSCP limit (dot-dashed curve) extents into the region of a tachyonic spectrum, in this region this limit is only an extrapolation.

This prescription is also used in ref. [30], where it is validated by reproducing the efficiencies and cross section upper limits for the gauge mediated supersymmetry breaking model from the full CMS detector simulation [29] with a relative error below 5%.

The resulting limits are shown in figure 4, projected onto the A_0-$m_{1/2}$ plane for two slices in parameter space, where $m^2_{H_u} = m^2_{H_d} = 0$, and $\tan\beta = 10$ (left panel) and $\tan\beta = 50$ (right panel). Both choices are characterized by a stau NLSP in the entire considered parameter plane. The considered CMS search for HSCPs at the 8 TeV LHC excludes the region below the red dot-dashed line (red shaded region) at 95% CL. The exclusion reach depends strongly on the overall sparticle mass spectrum, which is indicated by drawing several contours for the mass of the stau NLSP. The exclusion limits turn out to cut at around $m_{\tilde{\tau}_1} \gtrsim 400\,\text{GeV}$ with a mild dependence on the other parameters. This translates into a limit on $m_{1/2}$ between 1 and 2 TeV for $\tan\beta = 10$ in the considered region of A_0, but can be much larger for large $\tan\beta$, as shown in the right panel. The existing limit only touches the $-2\,\text{GeV}$ band regarding the Higgs mass, and leaves most of the parameter space that provides a Higgs mass in the range $123\,\text{GeV} \lesssim m_h \lesssim 127\,\text{GeV}$ unchallenged.

The 13 TeV LHC runs have pursued searches for heavy stable charged particles, and (preliminary) results from an integrated luminosity of $2.5\,\text{fb}^{-1}$ [31] ($12.9\,\text{fb}^{-1}$ [32]) have been released. For the 13 TeV searches, no on-/off-line probabilities (as in ref. [29]), have been provided, such that these searches cannot be easily reinterpreted. We do, however, expect to obtain a meaningful estimate of the 13 TeV sensitivity as described in the following.

JHEP05(2017)003

The signal efficiencies for the 8 TeV LHC increase with increasing $m_{\tilde{\tau}_1}$ for the tested points, and are only mildly dependent on the other parameters within the considered model. In particular, we found that the efficiency is always above 0.5 for $m_{\tilde{\tau}_1} > 350\,\text{GeV}$, and above 0.6 for $m_{\tilde{\tau}_1} > 500\,\text{GeV}$ in our scan. Assuming a similar detector performance, the efficiencies at the 13 TeV LHC for a certain stau mass will to first approximation be the same as for the 8 TeV efficiency, for a mass that is smaller by a factor of 8/13. Hence, for the 13 TeV LHC we assume an efficiency of 0.5, which is expected to provide a mostly conservative estimate for stau masses above 600 GeV. Furthermore, as for $m_{\text{rec}} > 200\,\text{GeV}$ the signal region is typically background-free [29, 31] we require 3 signal events in the signal region supporting a 95% CL exclusion limit. In this way we estimated the projected sensitivity for $300\,\text{fb}^{-1}$ at 13 TeV, for which we computed the production cross sections with Pythia 6 [27], see the red dot-dot-dashed curves in figure 4. The projected exclusion reach cuts into a larger portion of the parameter space providing the correct Higgs mass. In particular, the maximal mixing scenario for moderate values for $\tan\beta$ can be tested. With $300\,\text{fb}^{-1}$, stau masses up to around 1 TeV could be tested.

Note that performing the same estimate for the analysis at $2.5\,\text{fb}^{-1}$ ($12.9\,\text{fb}^{-1}$) provides an estimated limit very close to (slightly above) the 8 TeV limit, which we do not show in figure 4 for the sake of better readability.

Missing energy signatures. As discussed in section 3.2, a high enough $m^2_{H_d}$ relative to $m^2_{H_u}$ and $m_{1/2}$ results in a neutralino or even sneutrino NLSP. If present in collision events, neutral NLSPs lead to a missing transverse energy (MET) signature at the LHC.

In order to test the compatibility with current LHC results, we perform a Monte Carlo simulation with the MadGraph5_aMC@NLO event generator [28] for the 8 TeV LHC. We generate 20k events. The decay, showering and hadronization is performed by Pythia 6 [27]. The results are used as input to CheckMate 1 [33],[7] allowing us to simultaneously test the signal against various LHC searches for missing transverse energy.

We test our model against all ATLAS analyses implemented in CheckMate 1 [38–58]. These analyses search for final states containing a significant amount of missing transverse energy, in addition to jets or leptons. The signal is compared to experimental limits in the respective signal regions of the analysis at 95% CL. The most sensitive region from all the analyses is used to conclude whether the model can be excluded or not. Among the points that provide a Higgs mass $m_h > 123\,\text{GeV}$, we tested the lighter part of the spectrum, i.e., $m_{1/2} \leq 3$ TeV for various slices in parameter space regarding $\tan\beta$, A_0 and Higgs soft masses. We found that even for the lightest spectra the signal falls below the exclusion limits by at least an order of magnitude. Since the spectrum becomes heavier for larger values of $m_{1/2}$, we expect no sensitivity of searches for MET in the region $m_h > 123\,\text{GeV}$. The analysis which most frequently has the largest signal region is the search for direct stop pair production in final states with two leptons [38].

[7] CheckMate is built upon a number of external tools. The detector simulation is based on Delphes 3 [34], which incorporates FastJet [35, 36] using the Anti-kt jet algorithm [37].

JHEP05(2017)003

3.4 Charge and color breaking

In addition to the collider constraints, we investigate whether and in which regions of parameter space the current model is limited by charge- and color-breaking minima of the scalar potential. The MSSM contains 26 scalars, most of which carry electric or color charge. Hence, there is a danger of introducing charge- and color-breaking (CCB), depending on their VEVs where the scalar potential has its minimum. Due to the large number of scalars in the theory, the scalar potential is very complex, limiting an analytical approach to only considering certain rays in field space. It is common to investigate directions in field space where the VEVs of the Higgses and $\tilde{\tau}_{L/R}$ or $\tilde{t}_{L/R}$ have the same value, and to neglect the D-term of the potential, which is a gauge interaction and positive for non-zero values of the scalar fields, as well as loop corrections. Based on criteria for CCB as found in [59–61], we use the same condition as [5] for the stop trilinear coupling, namely

$$A_t^2 < 3(m_{H_u}^2 + |\mu|^2 + m_{Q_3}^2 + m_{\bar{u}_3}^2)\,. \tag{3.8}$$

By analogy, we take the bound on the stau trilinear to be

$$A_\tau^2 < 3(m_{H_d}^2 + |\mu|^2 + m_{L_3}^2 + m_{\bar{e}_3}^2)\,. \tag{3.9}$$

For large $\tan\beta$, one can derive an upper bound on the product $\mu\tan\beta$ requiring the standard electroweak vacuum to be stable or metastable with a lifetime larger than the age of the universe [62–65]. We use [65],

$$\begin{aligned}|\mu\tan\beta_{\text{eff}}| < {} & 56.9\sqrt{m_{L_3}m_{\bar{e}_3}} + 57.1\,(m_{L_3} + 1.03\,m_{\bar{e}_3}) - 1.28\times10^4\,\text{GeV} \\ & + \frac{1.67\times10^6\,\text{GeV}^2}{m_{L_3}+m_{\bar{e}_3}} - 6.41\times10^7\,\text{GeV}^3\left(\frac{1}{m_{L_3}^2} + \frac{0.983}{m_{\bar{e}_3}^2}\right),\end{aligned} \tag{3.10}$$

where $\tan\beta_{\text{eff}} \equiv \tan\beta/(1+\Delta_\tau)$ with

$$\Delta_\tau \simeq -\frac{3g^2}{32\pi^2}\mu\tan\beta\, M_2\, I(m_{\tilde{\nu}_\tau}, M_2, \mu) + \frac{g'^2}{16\pi^2}\mu\tan\beta\, M_1\, I(m_{\tilde{\tau}_1}, m_{\tilde{\tau}_2}, M_1)\,, \tag{3.11}$$

and

$$I(a,b,c) = \frac{1}{(a^2-b^2)(b^2-c^2)(a^2-c^2)}\left(a^2b^2\log\frac{a^2}{b^2} + b^2c^2\log\frac{b^2}{c^2} + c^2a^2\log\frac{c^2}{a^2}\right). \tag{3.12}$$

These bounds are superimposed in figure 4, where we show the most constraining bound from equations (3.9)–(3.10). For $\tan\beta = 10$, the region below the purple dashed line violates equation (3.9), while for $\tan\beta = 50$ it violates equation (3.10). For large negative A_0, the CCB bound cuts into the part of the parameter space that provides the correct Higgs mass.

Note that we impose these bounds as a first estimate, indicating the region where CCB constraints *might* exclude points in the parameter space. It has been shown [66, 67] that these bounds are useful, but not entirely reliable in determining vacuum stability when more sophisticated analyses are performed. We leave a detailed numerical analysis of the vacuum stability utilizing Vevacious [68] for future work.

JHEP05(2017)003

3.5 Cosmological constraints

Scenarios with long-lived NLSPs are subject to constraints from big bang nucleosynthesis (BBN) because the presence and late decays of the NLSPs can change the primordial abundances of light elements [69–71]. In our case, the NLSP decays comparatively early on BBN timescales due to the relatively heavy sparticle spectrum. For example, for gaugino mediation with a stau NLSP a lower bound of $m_{\widetilde{\tau}} \gtrsim 400\,\text{GeV}$ was found in [72], which roughly coincides with the lower limit from HSCP searches. Therefore we do not perform a detailed analysis here.

Another constraint we did not include is the non-thermal production of gravitino dark matter by NLSP decays, which may not exceed the observed dark matter density. This is interesting from a theoretical point of view because it leads to an *upper* bound on the sparticle masses but less relevant for phenomenology, since the constraint becomes relevant only for very large values of $m_{1/2}$ [72], which are far beyond the reach of the LHC.

4 Conclusions

We have considered phenomenological constraints on the gaugino mediation model of supersymmetry breaking. First, we verified that the model allows for soft trilinear scalar interaction terms. These terms were originally assumed to vanish in gaugino mediation and play a crucial role in achieving a Higgs mass in agreement with the observed value of 125 GeV. The trilinear matrices are proportional to the Yukawa coupling matrices, thus avoiding flavor problems. The proportionality factor can be different for up- and down-type sfermions.

Second, we explored the phenomenological consequences of non-vanishing trilinears. The first constraint we discussed is the experimentally observed Higgs mass, calculating the low-energy parameters and the sparticle spectrum with SPheno and the Higgs mass with FeynHiggs. We determined the parameter space regions where the Higgs mass lies within the LHC limits. Large negative trilinears are required to obtain an acceptable Higgs mass if the SUSY scale is to be kept near the reach of the LHC. We also observe that FeynHiggs 2.12.2 — incorporating important NNLL contributions — predict a Higgs mass around 3 GeV lower compared to the SPheno calculation in the parameter regions considered.

We also considered the phenomenological implications of the non-universal soft Higgs masses. We found that these parameters mainly affect which sparticle becomes the NLSP (we assume a gravitino LSP and that the lightest MSSM sparticle is the NLSP). Values of the ratio $r \equiv (m_{H_u}^2 - m_{H_d}^2)/m_{1/2}^2$ near zero correspond to a stau NLSP. As r is pushed to larger negative values, the NLSP can become the neutralino and eventually the tau sneutrino. This behavior also depends on A_0 and $\tan\beta$. For sufficiently large $|A_0|$ and $\tan\beta$, the composition of the stau NLSP changes from mainly $\widetilde{\tau}_\text{R}$ to mainly $\widetilde{\tau}_\text{L}$ as r becomes large and negative, passing through regions with large mixing.

Proceeding to investigate the LHC sensitivity of the scenario, we found that for a neutral NLSP, the viable part of parameter space is not challenged by missing energy searches. However, for a stau NLSP, the corresponding searches for heavy stable charged

JHEP05(2017)003

particles become sensitive and cut into the region where $123\,\text{GeV} \lesssim m_h \lesssim 127\,\text{GeV}$. The projection for an integrated luminosity of $300\,\text{fb}^{-1}$ reaches a large portion of this part of parameter space, especially in the maximal-mixing scenario.

Finally, we indicate in which regions of parameter space the model might be limited by charge- and color-breaking minima of the scalar potential by using (semi-)analytic estimates for the CCB conditions. It turns out that only a small part of the allowed Higgs mass region is in conflict with these CCB bounds.

Acknowledgments

We would like to thank Sven Heinemeyer, Ben O'Leary, Werner Porod, Florian Staub, Jamie Tattersall, and Alexander Voigt for helpful discussions. Special thanks are due to Felix Brümmer for pointing out to us the possibility of non-zero trilinear couplings in gaugino mediation. We acknowledge support by the German Research Foundation (DFG) through the research unit "New physics at the LHC".

References

[1] D.E. Kaplan, G.D. Kribs and M. Schmaltz, *Supersymmetry breaking through transparent extra dimensions*, *Phys. Rev.* **D 62** (2000) 035010 [hep-ph/9911293] [INSPIRE].

[2] Z. Chacko, M.A. Luty, A.E. Nelson and E. Ponton, *Gaugino mediated supersymmetry breaking*, *JHEP* **01** (2000) 003 [hep-ph/9911323] [INSPIRE].

[3] ATLAS and CMS collaborations, *Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments*, *Phys. Rev. Lett.* **114** (2015) 191803 [arXiv:1503.07589] [INSPIRE].

[4] R. Kitano, R. Motono and M. Nagai, *MSSM without free parameters*, *Phys. Rev.* **D 94** (2016) 115016 [arXiv:1605.08227] [INSPIRE].

[5] F. Brummer, S. Kraml and S. Kulkarni, *Anatomy of maximal stop mixing in the MSSM*, *JHEP* **08** (2012) 089 [arXiv:1204.5977] [INSPIRE].

[6] W. Buchmüller, K. Hamaguchi and J. Kersten, *The gravitino in gaugino mediation*, *Phys. Lett.* **B 632** (2006) 366 [hep-ph/0506105] [INSPIRE].

[7] H. Pagels and J.R. Primack, *Supersymmetry, cosmology and new TeV physics*, *Phys. Rev. Lett.* **48** (1982) 223 [INSPIRE].

[8] W. Buchmüller, J. Kersten and K. Schmidt-Hoberg, *Squarks and sleptons between branes and bulk*, *JHEP* **02** (2006) 069 [hep-ph/0512152] [INSPIRE].

[9] Z. Chacko, M.A. Luty and E. Ponton, *Massive higher dimensional gauge fields as messengers of supersymmetry breaking*, *JHEP* **07** (2000) 036 [hep-ph/9909248] [INSPIRE].

[10] A. Brignole, L.E. Ibáñez and C. Muñoz, *Soft supersymmetry breaking terms from supergravity and superstring models*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 125 [hep-ph/9707209] [INSPIRE].

JHEP05(2017)003

[11] S.P. Martin, *A Supersymmetry primer*, World Scientific, Singapore (2011), hep-ph/9709356 [INSPIRE].

[12] D. Matalliotakis and H.P. Nilles, *Implications of nonuniversality of soft terms in supersymmetric grand unified theories*, *Nucl. Phys.* **B 435** (1995) 115 [hep-ph/9407251] [INSPIRE].

[13] J.R. Ellis, K.A. Olive and Y. Santoso, *The MSSM parameter space with nonuniversal Higgs masses*, *Phys. Lett.* **B 539** (2002) 107 [hep-ph/0204192] [INSPIRE].

[14] O. Buchmueller et al., *The NUHM2 after LHC Run 1*, *Eur. Phys. J.* **C 74** (2014) 3212 [arXiv:1408.4060] [INSPIRE].

[15] T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak and G. Weiglein, *High-precision predictions for the light CP-even Higgs boson mass of the minimal supersymmetric standard model*, *Phys. Rev. Lett.* **112** (2014) 141801 [arXiv:1312.4937] [INSPIRE].

[16] S. Borowka, T. Hahn, S. Heinemeyer, G. Heinrich and W. Hollik, *Momentum-dependent two-loop QCD corrections to the neutral Higgs-boson masses in the MSSM*, *Eur. Phys. J.* **C 74** (2014) 2994 [arXiv:1404.7074] [INSPIRE].

[17] W. Porod, *SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e^+e^- colliders*, *Comput. Phys. Commun.* **153** (2003) 275 [hep-ph/0301101] [INSPIRE].

[18] W. Porod and F. Staub, *SPheno 3.1: Extensions including flavour, CP-phases and models beyond the MSSM*, *Comput. Phys. Commun.* **183** (2012) 2458 [arXiv:1104.1573].

[19] M. Frank, T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak and G. Weiglein, *The Higgs boson masses and mixings of the complex MSSM in the Feynman-diagrammatic approach*, *JHEP* **02** (2007) 047 [hep-ph/0611326] [INSPIRE].

[20] G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich and G. Weiglein, *Towards high precision predictions for the MSSM Higgs sector*, *Eur. Phys. J.* **C 28** (2003) 133 [hep-ph/0212020] [INSPIRE].

[21] S. Heinemeyer, W. Hollik and G. Weiglein, *The Masses of the neutral CP-even Higgs bosons in the MSSM: accurate analysis at the two loop level*, *Eur. Phys. J.* **C 9** (1999) 343 [hep-ph/9812472] [INSPIRE].

[22] S. Heinemeyer, W. Hollik and G. Weiglein, *FeynHiggs: a program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM*, *Comput. Phys. Commun.* **124** (2000) 76 [hep-ph/9812320].

[23] H. Bahl and W. Hollik, *Precise prediction for the light MSSM Higgs boson mass combining effective field theory and fixed-order calculations*, *Eur. Phys. J.* **C 76** (2016) 499 [arXiv:1608.01880] [INSPIRE].

[24] P. Athron, J.-h. Park, T. Steudtner, D. Stöckinger and A. Voigt, *Precise Higgs mass calculations in (non-)minimal supersymmetry at both high and low scales*, *JHEP* **01** (2017) 079 [arXiv:1609.00371] [INSPIRE].

[25] Particle Data Group collaboration, K.A. Olive et al., *Review of particle physics*, *Chin. Phys.* **C 38** (2014) 090001 [INSPIRE].

[26] M. Carena et al., *Reconciling the two loop diagrammatic and effective field theory computations of the mass of the lightest CP-even Higgs boson in the MSSM*, *Nucl. Phys.* **B 580** (2000) 29 [hep-ph/0001002] [INSPIRE].

[27] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

JHEP05(2017)003

[28] 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*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[29] CMS collaboration, *Constraints on the pMSSM, AMSB model and on other models from the search for long-lived charged particles in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, *Eur. Phys. J.* **C 75** (2015) 325 [arXiv:1502.02522] [INSPIRE].

[30] J. Heisig, A. Lessa and L. Quertenmont, *Simplified models for exotic BSM searches*, *JHEP* **12** (2015) 087 [arXiv:1509.00473] [INSPIRE].

[31] CMS collaboration, *Search for long-lived charged particles in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Rev.* **D 94** (2016) 112004 [arXiv:1609.08382] [INSPIRE].

[32] CMS collaboration, *Search for heavy stable charged particles with* 12.9 fb^{-1} *of 2016 data*, CMS-PAS-EXO-16-036 (2016).

[33] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall and J.S. Kim, *CheckMATE: confronting your favourite new physics model with LHC data*, *Comput. Phys. Commun.* **187** (2015) 227 [arXiv:1312.2591] [INSPIRE].

[34] DELPHES 3 collaboration, J. de Favereau et al., *DELPHES 3, a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[35] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[36] M. Cacciari and G.P. Salam, *Dispelling the* N^3 *myth for the* k_t *jet-finder*, *Phys. Lett.* **B 641** (2006) 57 [hep-ph/0512210] [INSPIRE].

[37] M. Cacciari, G.P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[38] 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*, *JHEP* **06** (2014) 124 [arXiv:1403.4853] [INSPIRE].

[39] ATLAS collaboration, *Search for strongly produced supersymmetric particles in decays with two leptons at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2013-089 (2013).

[40] ATLAS collaboration, *Search for direct-slepton and direct-chargino production in final states with two opposite-sign leptons, missing transverse momentum and no jets in* 20 fb^{-1} *of pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, ATLAS-CONF-2013-049 (2013).

[41] ATLAS collaboration, *Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum and* 20.3 fb^{-1} *of* $\sqrt{s} = 8$ *TeV proton-proton collision data*, ATLAS-CONF-2013-047 (2013).

[42] ATLAS collaboration, *Search for direct production of charginos and neutralinos in events with three leptons and missing transverse momentum in* 21 fb^{-1} *of pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, ATLAS-CONF-2013-035 (2013).

[43] ATLAS collaboration, *Search for direct production of the top squark in the all-hadronic* $t\bar{t}$*+ Etmiss final state in* 21 fb^{-1} *of pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, ATLAS-CONF-2013-024 (2013).

[44] ATLAS collaboration, *Search for new phenomena in monojet plus missing transverse momentum final states using* 10 fb^{-1} *of pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector at the LHC*, ATLAS-CONF-2012-147 (2012).

JHEP05(2017)003

[45] ATLAS collaboration, *Search for supersymmetry at $\sqrt{s} = 8$ TeV in final states with jets, missing transverse momentum and one isolated lepton*, ATLAS-CONF-2012-104 (2012).

[46] 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*, *Eur. Phys. J.* **C 75** (2015) 299 [arXiv:1502.01518] [INSPIRE].

[47] ATLAS collaboration, *ATLAS Run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider*, *Eur. Phys. J.* **C 75** (2015) 510 [arXiv:1506.08616] [INSPIRE].

[48] ATLAS collaboration, *Search for new phenomena in events with a photon and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Rev.* **D 91** (2015) 012008 [arXiv:1411.1559] [INSPIRE].

[49] 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*, *Phys. Rev.* **D 90** (2014) 052008 [arXiv:1407.0608] [INSPIRE].

[50] 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*, *JHEP* **11** (2014) 118 [arXiv:1407.0583] [INSPIRE].

[51] 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*, *JHEP* **09** (2014) 176 [arXiv:1405.7875] [INSPIRE].

[52] 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*, *JHEP* **06** (2014) 035 [arXiv:1404.2500] [INSPIRE].

[53] 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*, *Eur. Phys. J.* **C 74** (2014) 2883 [arXiv:1403.5222] [INSPIRE].

[54] 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*, *JHEP* **04** (2014) 169 [arXiv:1402.7029] [INSPIRE].

[55] 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*, *JHEP* **10** (2013) 189 [arXiv:1308.2631] [INSPIRE].

[56] 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*, *JHEP* **10** (2013) 130 [*Erratum ibid.* **01** (2014) 109] [arXiv:1308.1841] [INSPIRE].

[57] ATLAS collaboration, *Search for supersymmetry in events containing a same-flavour opposite-sign dilepton pair, jets and large missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, *Eur. Phys. J.* **C 75** (2015) 318 [arXiv:1503.03290] [INSPIRE].

[58] ATLAS collaboration, *Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets using* 20.1 fb^{-1} *of pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, ATLAS-CONF-2013-061 (2013).

[59] C. Kounnas, A.B. Lahanas, D.V. Nanopoulos and M. Quirós, *Low-energy behavior of realistic locally supersymmetric grand unified theories*, *Nucl. Phys.* **B 236** (1984) 438 [INSPIRE].

JHEP05(2017)003

[60] J.P. Derendinger and C.A. Savoy, *Quantum effects and* SU(2) × U(1) *breaking in supergravity gauge theories*, *Nucl. Phys.* **B 237** (1984) 307 [INSPIRE].

[61] J.M. Frere, D.R.T. Jones and S. Raby, *Fermion masses and induction of the weak scale by supergravity*, *Nucl. Phys.* **B 222** (1983) 11 [INSPIRE].

[62] R. Rattazzi and U. Sarid, *Large* tan β *in gauge mediated SUSY breaking models*, *Nucl. Phys.* **B 501** (1997) 297 [hep-ph/9612464] [INSPIRE].

[63] J. Hisano and S. Sugiyama, *Charge-breaking constraints on left-right mixing of stau's*, *Phys. Lett.* **B 696** (2011) 92 [*Erratum ibid.* **B 719** (2013) 472] [arXiv:1011.0260] [INSPIRE].

[64] M. Carena, S. Gori, I. Low, N.R. Shah and C.E.M. Wagner, *Vacuum stability and higgs diphoton decays in the MSSM*, *JHEP* **02** (2013) 114 [arXiv:1211.6136] [INSPIRE].

[65] T. Kitahara and T. Yoshinaga, *Stau with large mass difference and enhancement of the Higgs to diphoton decay rate in the MSSM*, *JHEP* **05** (2013) 035 [arXiv:1303.0461] [INSPIRE].

[66] J.E. Camargo-Molina, B. O'Leary, W. Porod and F. Staub, *Stability of the CMSSM against sfermion VEVs*, *JHEP* **12** (2013) 103 [arXiv:1309.7212] [INSPIRE].

[67] D. Chowdhury, R.M. Godbole, K.A. Mohan and S.K. Vempati, *Charge and color breaking constraints in MSSM after the Higgs discovery at LHC*, *JHEP* **02** (2014) 110 [arXiv:1310.1932] [INSPIRE].

[68] J.E. Camargo-Molina, B. O'Leary, W. Porod and F. Staub, *Vevacious: a tool for finding the global minima of one-loop effective potentials with many scalars*, *Eur. Phys. J.* **C 73** (2013) 2588 [arXiv:1307.1477] [INSPIRE].

[69] T. Moroi, H. Murayama and M. Yamaguchi, *Cosmological constraints on the light stable gravitino*, *Phys. Lett.* **B 303** (1993) 289 [INSPIRE].

[70] M. Pospelov, *Particle physics catalysis of thermal Big Bang nucleosynthesis*, *Phys. Rev. Lett.* **98** (2007) 231301 [hep-ph/0605215] [INSPIRE].

[71] M. Kawasaki, K. Kohri, T. Moroi and A. Yotsuyanagi, *Big-Bang nucleosynthesis and gravitino*, *Phys. Rev.* **D 78** (2008) 065011 [arXiv:0804.3745] [INSPIRE].

[72] J. Kersten and K. Schmidt-Hoberg, *The gravitino-stau scenario after catalyzed BBN*, *JCAP* **01** (2008) 011 [arXiv:0710.4528] [INSPIRE].

JHEP05(2017)003

Paper II

Signal mixture estimation for degenerate heavy Higgses using a deep neural network

Eur. Phys. J. C (2018) 78:1010
https://doi.org/10.1140/epjc/s10052-018-6455-z

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Signal mixture estimation for degenerate heavy Higgses using a deep neural network

Anders Kvellestad[1,2,a], **Steffen Maeland**[3,b], **Inga Strümke**[3,c]

[1] Department of Physics, University of Oslo, 0316 Oslo, Norway
[2] Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
[3] Department of Physics and Technology, University of Bergen, 5020 Bergen, Norway

Received: 9 May 2018 / Accepted: 15 November 2018

Abstract If a new signal is established in future LHC data, a next question will be to determine the signal composition, in particular whether the signal is due to multiple near-degenerate states. We investigate the performance of a deep learning approach to signal mixture estimation for the challenging scenario of a ditau signal coming from a pair of degenerate Higgs bosons of opposite CP charge. This constitutes a parameter estimation problem for a mixture model with highly overlapping features. We use an unbinned maximum likelihood fit to a neural network output, and compare the results to mixture estimation via a fit to a single kinematic variable. For our benchmark scenarios we find a $\sim 20\%$ improvement in the estimate uncertainty.

1 Introduction

Machine learning techniques have already proven useful in particle physics, especially for separating signal from background events in analyses of LHC data. More recently, *deep learning* methods, such as multi-layer neural networks, have been shown to perform very well, due to their ability to learn complex non-linear correlations in high-dimensional data [1–3]. In this paper we study the performance of a deep neural network classifier, but rather than classifying signal vs. background we focus on estimating the mixture of different signal classes in a dataset. This is motivated by the not-unlikely scenario where a new (and possibly broad) resonance is discovered in future LHC data, but limited statistics makes the interpretation difficult, in particular the question of whether the signal is due to multiple degenerate states. In such a scenario it will clearly be important to squeeze as much information as possible from the available data.

While the approach studied here is general, we take a Two-Higgs-Doublet Model (THDM) as our example scenario. In these models the Higgs sector of the Standard Model (SM) is extended with an additional $SU(2)$ doublet, predicting the existence of a pair of charged scalars ($H^\pm$) and three neutral scalars (h, H, A), one of which should be the observed 125 GeV Higgs. Several more extensive frameworks for New Physics predict a Higgs sector with the structure of a THDM, the prime example being the Minimal Supersymmetric Standard Model (MSSM). A further motivation for THDMs comes from the fact that the extended scalar sector can allow for additional sources of CP violation and a strongly first-order electroweak phase transition, as required for electroweak baryogenesis [4–7]. For a recent study of this, see [8].

We associate the light scalar h with the observed 125 GeV Higgs and take the heavier scalars H, A and $H^\pm$ to be mass degenerate. The focus of our study is on the ditau LHC signal from decays of the neutral states H and A, which in this case are indistinguishable save for their opposite CP charges. Searches for heavy neutral Higgses in ditau final states are carried out by both the ATLAS and CMS collaborations, see [9,10] for recent results.

The remainder of this paper is structured as follows. In Sect. 2 we motivate why it is reasonable to expect a certain level of mass-degeneracy among the new scalars in THDMs and present our example THDM scenario. The technical setup for our analysis is given in Sect. 3. Here we define our signal models, describe the procedure for Monte Carlo event generation and detail the neural network layout and training. In Sect. 4 we demonstrate H/A signal mixture estimation using the method of fitting a single kinematic variable. The result serves as our baseline for judging the performance of the deep learning approach. Our main results are presented in

[a] e-mail: anders.kvellestad@fys.uio.no
[b] e-mail: steffen.maeland@uib.no
[c] e-mail: inga.strumke@uib.no

Published online: 12 December 2018

Sect. 5. Here we estimate the signal mixture via a maximum likelihood fit to the output distribution from a network trained to separate H and A ditau events. The results are compared to those from Sect. 4. We state our conclusions in Sect. 6.

2 Theory and motivation

The starting point for our study is a THDM scenario where $m_H \approx m_A$. Our main motivation for this choice is to obtain a challenging test case for signal mixture estimation. However, there are also physical reasons to expect the H and A states to have similar masses. After requiring that the scalar potential has a minimum in accordance with electroweak symmetry breaking, we are left with a model with only two mass scales, $v \approx 246\,\text{GeV}$ and a free mass parameter μ, to control the four masses m_h, m_H, m_A and $m_{H^\pm}$. From the point of view of the general THDM parameter space, the least fine-tuned way to align the light state h with SM predictions, as favoured by LHC Higgs data, is to move towards simultaneous decoupling of the three heavier states by increasing μ, leaving v to set the scale for $m_h = 125\,\text{GeV}$ [11]. This points to a scenario where $|m_H - m_A| \lesssim 100\,\text{GeV}$, and quite possibly much smaller, depending on the quartic couplings of the scalar potential.[1]

Further motivation for a small H–A mass difference can be found in less general realisations of THDMs. For the type-II THDM in the MSSM the quartic couplings are fixed by the squares of the SM gauge couplings, resulting in the tree-level prediction that $m_H - m_A \lesssim 10\,\text{GeV}$ for $m_A \sim 400\,\text{GeV}$ and $\tan\beta \sim 1$, and decreasing further with increasing $\tan\beta$ or m_A [13]. Another well-motivated scenario predicting closely degenerate H and A states is the $SO(5)$-based Maximally Symmetric THDM [14].

When mass degenerate, the H and A appear identical except for their CP charge. If the properties of the light h deviates from SM predictions, this difference in CP charge can manifest as non-zero ZZ and WW couplings for H, while for the CP-odd A the Zh coupling is available. However, these couplings all vanish in the perfect SM-alignment limit we assume here. Yet the CP nature of H and A is still expressed as spin correlations in fermionic decay modes, impacting the kinematics of subsequent decays. Here we study the channels $H \to \tau\tau$ and $A \to \tau\tau$. Methods for reconstructing spin correlations in ditau decays of the 125 GeV Higgs have been investigated in detail [15–18], providing a good baseline for comparison. The use of neural networks to optimize CP measurements for the 125 GeV state is studied in [19,20].

2.1 Benchmark scenario

Two-Higgs-Doublet Models are classified in different types based on the structure of the Yukawa sector. We choose a benchmark scenario within the CP-conserving lepton-specific THDM, with $m_H = m_A = m_{H^\pm} = 450\,\text{GeV}$. In this model, the quarks couple to one of the Higgs doublets and the leptons to the other. This enables large branching ratios for $H/A \to \tau\tau$, even for masses above the 350 GeV threshold for $H/A \to t\bar{t}$.

By varying the remaining THDM parameters we can obtain a wide range of ditau signal strengths for the H and A states at 450 GeV. In Appendix A we illustrate how $\sigma(pp \to A) \times \mathcal{B}(A \to \tau\tau)$ and $\sigma(pp \to H) \times \mathcal{B}(H \to \tau\tau)$ vary across the high-mass region of the lepton-specific THDM parameter space. For $m_H = m_A \approx 450\,\text{GeV}$, we find that the ditau signal strengths can reach up to $\sigma(pp \to H) \times \mathcal{B}(H \to \tau\tau) \approx 34\,\text{fb}$ and $\sigma(pp \to A) \times \mathcal{B}(A \to \tau\tau) \approx 54\,\text{fb}$ in 13 TeV proton–proton collisions. This includes production via gluon–gluon fusion and bottom-quark annihilation, with cross sections evaluated at NLO using SusHi 1.6.1 [21–27] and branching ratios obtained from 2HDMC 1.7.0 [28].

For comparison, in Appendix A we also show the result of a similar scan of the type-I THDM. In this model all fermions couple to only one of the two Higgs doublets. Compared to the lepton-specific THDM, the ditau signal in type-I THDM suffers a much stronger suppression from the $H/A \to t\bar{t}$ channel.

As further described in Sects. 3 and 4, the mixture estimation techniques we study require each tau to decay through the $\tau^\pm \to \pi^\pm\pi^0\nu$ channel, which has a branching ratio of 25%. However, the neural network method we employ can be extended to include other tau decay modes as well, by implementing the "impact parameter method" in [18] in addition to the "ρ decay-plane method" used here.

If we only assume the $\tau^\pm \to \pi^\pm\pi^0\nu$ decay channel and an acceptance times efficiency of 5%–10% for the signal selection, our example scenarios predict no more than ~ 100 signal events for the anticipated 300 fb^{-1} dataset at the end of Run 3. However, as the model scan in Appendix A shows, considering slightly lower benchmark masses can provide an order of magnitude increase in the predicted cross-section. Also, extending the method to include more tau decay channels can greatly increase the statistics available to the analysis discussed here. Still, the large backgrounds in the ditau channel, e.g. from "fake QCD taus", implies that a signal mixture estimation study for the THDM benchmark scenario we present here likely will require the improved statistics of the full High-Luminosity LHC dataset.

[1] A large H–A mass difference in this decoupling scenario relies on $\mathcal{O}(1)$ quartic couplings. We note that when loop corrections are taken into account, the viability of such scenarios can be significantly more restricted than what tree-level results suggest [12].

We do not include a third mixture component representing ditau backgrounds for our benchmark study. Clearly, the inclusion of backgrounds will increase the uncertainty in the estimated $H/A \to \tau\tau$ signal mixture. However, as we discuss in more detail in Sect. 5.1, the mixture estimate obtained from the neural network approach we study here is likely to be less affected by backgrounds than traditional mixture estimation from fitting a single kinematic variable.

For our further discussions we define the parameter α as the ratio of the $A \to \tau\tau$ signal strength to the total ditau signal strength,

$$\alpha \equiv \frac{\sigma(pp \to A) \times \mathcal{B}(A \to \tau\tau)}{\sigma(pp \to A) \times \mathcal{B}(A \to \tau\tau) + \sigma(pp \to H) \times \mathcal{B}(H \to \tau\tau)}. \tag{1}$$

This is the parameter we seek to determine in our signal mixture estimation.[2] The parameter region of our benchmark scenario predicts values of α between 0.5 and 0.7. To allow for some further variation in the assumptions, we will in our tests use α values of 0.5, 0.7 and 0.9.

3 Analysis setup

3.1 Event generation

We generate 13 TeV pp Monte Carlo events for this study using Pythia 8.219 [29,30]. Only gluon-gluon fusion and bottom-quark annihilation are considered, as these are the dominant H/A production modes at the LHC.[3] For our analysis we select opposite-sign taus decaying to $\pi^{\pm}\pi^0\nu$, which is the decay mode with the highest branching ratio. In order to roughly match recent LHC searches for $H/A \to \tau\tau$, taus are required to have visible transverse momentum p_T larger than 40 GeV and pseudorapidity less than 2.1. Further, we require the taus to be separated by $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} > 0.5$, and that there are no more than two taus in the event which pass the p_T selection. Events with muons or electrons with $p_T > 20$ GeV are rejected.

Detector effects are taken into account by randomly smearing the directions and energies of the outgoing pions, following the procedure described in [18]: Each track is deflected by a random polar angle θ, which is drawn from a Gaussian distribution with width σ_θ, so that the smeared track lies within a cone around the true track direction. For charged pions a value of $\sigma_\theta = 1$ mrad is used, while the energy resolution is $\Delta E/E = 5\%$. For neutral pions, we use $\sigma_\theta = 0.025/\sqrt{12}$ rad and $\Delta E/E = 10\%$. To gauge the impact of such detector effects on our results, we repeat the main analyses in Sects. 4 and 5 for simulated data with and without detector smearing.

3.2 Network input features

For the neural signal mixture estimation in Sect. 5, we train a network to separate $H \to \tau\tau$ events from $A \to \tau\tau$ events. The four-momenta of the visible tau decay products ($\pi^{\pm}$ and π^0) constitute the most basic kinematic input features to our network. The momenta are boosted back to the visible ditau rest frame (the zero-momentum frame for the four pions) and rotated so that the visible taus are back-to-back along the z-axis. The system is then rotated a second time, now around the z-axis, so that the x-component of the π^+ is zero. This is done in order to align all events to a common orientation, as the azimuthal angle around the z-axis carries no physics information.

In addition to the pion momenta, the network is trained on missing transverse energy (E_T^{miss}); the invariant mass of the four-pion system (m_{vis}); the transverse mass (m_T^{tot}); the impact parameter vectors of the charged pions, which help constrain the neutrino directions; the pion energy ratios $\Upsilon^{\pm}$, defined as

$$\Upsilon^{\pm} = \frac{E_{\pi^{\pm}} - E_{\pi^0}}{E_{\pi^{\pm}} + E_{\pi^0}}; \tag{2}$$

and the angle φ^* between the tau decay planes. For φ^* we follow the definition in [18],[4] which uses the direction $\hat{\mathbf{p}}_{\perp}^{(0)\pm}$ of the π^0 transverse to the direction $\hat{\mathbf{p}}^{\pm}$ of the corresponding $\pi^{\pm}$, to form an intermediate observable $\varphi \in [0, \pi)$ and a CP-odd triple correlation product $\mathcal{O}^*$,

$$\varphi = \arccos\left(\hat{\mathbf{p}}_{\perp}^{(0)+} \cdot \hat{\mathbf{p}}_{\perp}^{(0)-}\right) \quad \text{and} \tag{3}$$

$$\mathcal{O}^* = \hat{\mathbf{p}}^{+} \cdot \left(\hat{\mathbf{p}}_{\perp}^{(0)+} \times \hat{\mathbf{p}}_{\perp}^{(0)-}\right). \tag{4}$$

From these, we can define an angle continuous on the interval $[0, 2\pi)$:

$$\varphi' = \begin{cases} \varphi & \text{if } \mathcal{O}^* \geq 0 \\ 2\pi - \varphi & \text{if } \mathcal{O}^* < 0 \end{cases}. \tag{5}$$

[2] In linking this theory quantity directly with the H/A event mixture in the datasets we simulate, we make the approximation that the acceptance times efficiency is equal for $H \to \tau\tau$ and $A \to \tau\tau$ events.

[3] The magnitudes of the up-type and down-type Yukawa couplings have the same $\tan\beta$ dependence in both the lepton-specific and the type-I THDM. Gluon–gluon fusion through a top loop is therefore by far the most important production channel for the scenarios considered here.

[4] Our definition of φ^* only differs from that in [18] in that we define φ^* in the $\pi^+\pi^0\pi^-\pi^0$ zero-momentum frame, whereas the $\pi^+\pi^-$ zero-momentum frame is used in [18].

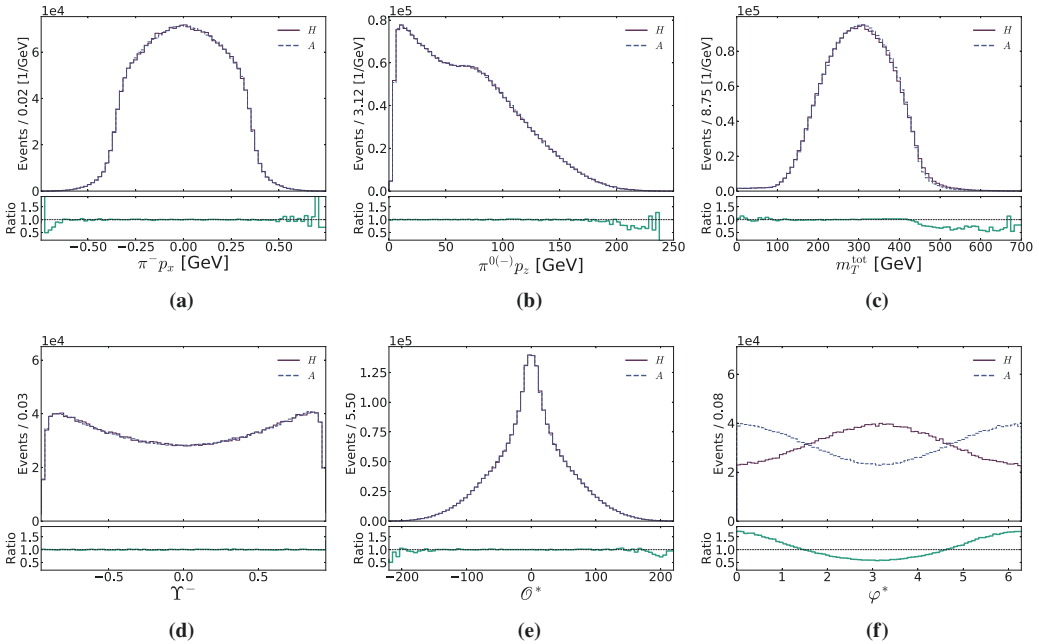


Fig. 1 Distributions for some kinematic features in $H \to \tau\tau \to (\pi^+\pi^0\nu)(\pi^-\pi^0\nu)$ events (solid purple line) and $A \to \tau\tau \to (\pi^+\pi^0\nu)(\pi^-\pi^0\nu)$ events (dashed blue line), assuming $m_H = m_A = 450$ GeV. The quantities in **a** and **b** are momentum components of the π^- and π^0 from the τ^- decay, after each event has been boosted back to the visible ditau restframe and rotated such that the taus are back-to-back in the z direction and the x-component of the π^+ momentum is zero. **c** The transverse mass m_T^{tot}, defined in [31]. The observables Υ (**d**) and $\mathcal{O}^*$ (**e**), defined in Eqs. (2) and (4), respectively, are required for the computation of φ^*, along with the momentum vectors of the tau decay products. The distribution of φ^* is shown in (**f**). The green graph below each plot shows the ratio of the A-event and H-event distributions

The distribution of φ' depends on the sign of the product $\Upsilon^+\Upsilon^-$; in the case of $\Upsilon^+\Upsilon^- \geq 0$, the distribution is phase-shifted by π relative to the case of $\Upsilon^+\Upsilon^- < 0$. To incorporate this into a single consistent CP-sensitive observable, we define φ^* as

$$\varphi^* = \begin{cases} \varphi' & \text{if } \Upsilon^+\Upsilon^- \geq 0 \\ (\varphi' + \pi) \bmod 2\pi & \text{if } \Upsilon^+\Upsilon^- < 0 \end{cases}. \tag{6}$$

Before being input to the network, all feature distributions are standardised to have zero mean and unit variance. A selection of the feature distributions in the training data is shown in Fig. 1. The univariate feature distributions are severely overlapping for H and A events, indicating that the classification task is very challenging. The one feature which stands out here is φ^*, which is the basis for the single-variable mixture estimation described in Sect. 4.

For the results presented in Sect. 5 we use a network trained on all features discussed above. However, features such as φ^* and m_T are derived from the basic pion momenta that the network also has access to. These "high-level" features can in principle be inferred by the network itself from the "low-level" pion momenta. To briefly investigate this we repeat the network training with varying subsets of the input features, starting with only the pion four-momenta and sequentially adding φ^*, $\Upsilon^\pm$ and the remaining features. For all networks we obtain ROC AUC scores of ~ 0.630. While a full statistical comparison of the resulting networks is beyond the scope of our study, this indicates that the network is itself able to extract the relevant information from high-dimensional correlations between the pion momenta, making the explicit inclusion of the high-level inputs mostly redundant. We note that this observation is in agreement with the results of [1,2].

It is still interesting to investigate how much of the discriminatory power can be captured by the high-level features alone. For this we train several classifiers on high-level features only, adding a new set of features for each classifier. The first classifier is trained only on φ^* and achieves a ROC AUC score of ~ 0.605. When Υ^+ and Υ^- are included as input

features the performance improves to a score of ~ 0.618. This improvement can be understood qualitatively from the fact that the difference between the $\Upsilon^{\pm}$-conditional φ^* distributions for H and A events increases with $|\Upsilon^{\pm}|$. Adding E_T^{miss}, m_T, $\pi^{\pm}$ impact parameter vectors and $\mathcal{O}^*$ raises the ROC AUC score to ~ 0.620, and finally including m_{vis} further increases the score to ~ 0.623, which seems to be the limit for our network when trained on high-level features only. This indicates that φ^* and $\Upsilon^{\pm}$ together capture most of the sensitivity, but that the neural network is able to extract from the pion four-momenta some additional information which is not contained in the high-level quantities. Similar behaviour was seen in [19] in a study focusing on the CP-nature of the 125 GeV Higgs.

3.3 Network layout

In this study we employ a fully-connected feed-forward network. The input layer has 26 nodes, followed by 500 nodes in the first hidden layer, 1000 nodes in the second hidden layer, and 100 nodes in the final hidden layer. These have leaky ReLU [32] activation functions, and dropout [33] is applied with a dropping probability of 0.375. No further regularisation is imposed. All network weights are initialised from a normal distribution, following the He procedure [34]. The output layer has a softmax activation function, and we apply batch normalisation [35] between all layers. The weights are optimised using Adam [36] with cross-entropy loss and an initial learning rate of 0.03. 20% of the training data are set aside to validate the model performance during training. If there is no improvement of the loss on the validation data for ten consecutive epochs, the learning rate is reduced by a factor ten. The network is trained for 100 epochs or until no improvement is observed during 15 epochs, whichever occurs first. The neural network implementation is done using the Keras [37] and TensorFlow [38] frameworks.

4 The φ^* method

Traditional approaches for separating CP-even and -odd decays are based on the angle φ^* between the tau decay planes, as defined in Eq. (6). The φ^* distribution for H and A events can be seen in Fig. 2a. The CP-sensitive parameter in this distribution is the phase of the sinusoidal curve, which is shifted by π radians between the H and A hypotheses. We note that the distributions overlap across the full φ^* range, hence no absolute event separation is possible based on this variable.

Using the simplified notation $p(\varphi^*|A) \equiv p_A(\varphi^*)$ and $p(\varphi^*|H) \equiv p_H(\varphi^*)$, the φ^* distribution for H/A signal data can be expressed as a simple mixture model,

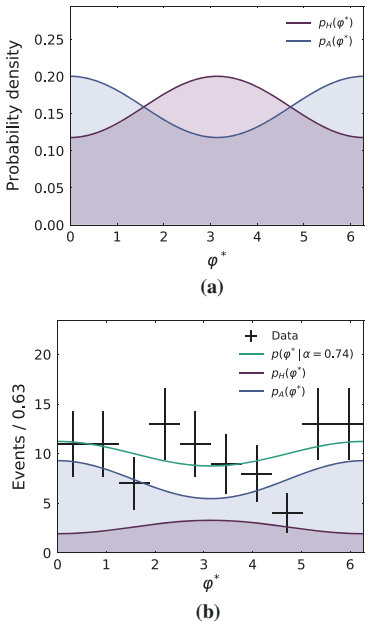


Fig. 2 **a** The probability density for φ^* in H events ($p_H(\varphi^*)$) and A events ($p_A(\varphi^*)$). **b** A fit of the mixture model $p(\varphi^*|\alpha) = \alpha p_A(\varphi^*) + (1-\alpha)p_H(\varphi^*)$ to a test dataset. Data points are shown in black, while the fitted model (normalized to 100 events) is shown in green. For this dataset the best-fit α value is $\hat{\alpha} = 0.74$

$$\begin{aligned} p(\varphi^*|\alpha) &= \alpha p_A(\varphi^*) + (1-\alpha)p_H(\varphi^*) \\ &= \alpha\big(a\cos\varphi^* + c\big) + \big(1-\alpha\big)\big(a\cos(\varphi^* + \pi) + c\big) \\ &= \alpha\, a\cos\varphi^* + (1-\alpha)\, a\cos(\varphi^* + \pi) + c, \end{aligned} \tag{7}$$

where we fix the amplitude a and offset c to $a = 0.041$ and $c = 0.159$, obtained from a separate fit to H and A training data. This leaves us with a model for the φ^* distribution where α is the only free parameter. Given a dataset $\{\varphi_i^*\}$ with N events, we can now obtain an estimate $\hat{\alpha}$ for α by maximising the likelihood function

$$\hat{\alpha} = \arg\max_{\alpha} \prod_{i=1}^{N} p(\varphi_i^*|\alpha). \tag{8}$$

We demonstrate this method in Fig. 2 for a dataset with 100 H/A Pythia events, generated using a model with a true α of 0.7. The pdfs $p_H(\varphi^*)$ and $p_A(\varphi^*)$ are shown in Fig. 2a, while the fit result is shown in Fig. 2b. For this example the best-fit α estimate comes out at $\hat{\alpha} = 0.74$.

To demonstrate the statistical performance of this estimator we repeat the fit using 10,000 independent test sets with 100 Pythia events each, generated with true α values

of 0.5, 0.7 and 0.9. The resulting distributions of α estimates are shown in Fig. 4a, where the purple (green) distributions depict results without (with) detector effects. By fitting a Gaussian to each distribution we find the spread in the estimates to be $\sigma_\alpha = 0.27$ ($\sigma_\alpha^{\rm det} \simeq 0.45$) when detector smearing is omitted (included). Further, the estimator is mean-unbiased for all three cases. Note that to demonstrate the unbiasedness we have allowed the fit to vary α beyond the physically valid range of [0, 1].

5 The neural network method

When estimating some parameter θ using collider data we ideally want to make use of the multivariate density $p(\mathbf{x}|\theta)$ for the complete set of event features $\mathbf{x}$.[5] However, it is typically infeasible to evaluate this density directly for a given $\mathbf{x}$. A common approach is then to construct a new variable $y(\mathbf{x})$ and base the parameter estimation on the simpler, univariate distribution $p(y(\mathbf{x})|\theta)$, as exemplified by the φ^* fit in Sect. 4.

The performance of such a univariate approach depends on how well the distribution $p(y(\mathbf{x})|\theta)$ retains the sensitivity to θ found in the underlying distribution $p(\mathbf{x}|\theta)$. In the special case where the map $y(\mathbf{x})$ is the output from a trained classifier, it can be shown that using $p(y(\mathbf{x})|\theta)$ to estimate θ in the ideal limit is equivalent to using the full data distribution $p(\mathbf{x}|\theta)$. Here we briefly review this argument before applying the classifier approach to our mixture estimation problem.

After training on θ-labeled data, a classifier that minimizes a suitably chosen error function will approximate a decision function $s(\mathbf{x})$ that is a strictly monotonic function of the density ratio $p(\mathbf{x}|\theta)/p(\mathbf{x}|\theta')$ [39].[6] As shown in [40], the monotonicity of $s(\mathbf{x})$ ensures that density ratios based on the multivariate distribution $p(\mathbf{x}|\theta)$ and the univariate distribution $p(s(\mathbf{x})|\theta)$ are equivalent,

$$\frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta')} = \frac{p(s(\mathbf{x})|\theta)}{p(s(\mathbf{x})|\theta')}. \tag{9}$$

If we now take θ' to be a fixed value such that the support of $p(\mathbf{x}|\theta')$ covers the support of $p(\mathbf{x}|\theta)$,[7] the maximum likelihood estimator for θ based on $p(\mathbf{x}|\theta)$ can be rewritten as follows [40]:

$$\begin{aligned}\hat{\theta} &= \arg\max_\theta \prod_{i=1}^N p(\mathbf{x}_i|\theta) \\ &= \arg\max_\theta \prod_{i=1}^N \frac{p(\mathbf{x}_i|\theta)}{p(\mathbf{x}_i|\theta')} \\ &= \arg\max_\theta \prod_{i=1}^N \frac{p(s(\mathbf{x}_i)|\theta)}{p(s(\mathbf{x}_i)|\theta')} \\ &= \arg\max_\theta \prod_{i=1}^N p(s(\mathbf{x}_i)|\theta).\end{aligned} \tag{10}$$

Hence, if the classifier output $y(\mathbf{x})$ provides a reasonable approximation of $s(\mathbf{x})$ we can expect the maximum likelihood estimator based on $p(y(\mathbf{x})|\theta)$ to exhibit similar performance to an estimator based on $p(\mathbf{x}|\theta)$. The main drawbacks of this approach are the complications associated with training the classifier, and that the physics underlying the parameter sensitivity may remain hidden from view.

We now apply this classifier approach to our H/A mixture estimation problem. The maximum likelihood estimator for the mixture parameter α is then given by

$$\begin{aligned}\hat{\alpha} &= \arg\max_\alpha \prod_{i=1}^N p(y(\mathbf{x}_i)|\alpha) \\ &= \arg\max_\alpha \prod_{i=1}^N \Big[\alpha p_A(y(\mathbf{x}_i)) + (1-\alpha) p_H(y(\mathbf{x}_i))\Big],\end{aligned} \tag{11}$$

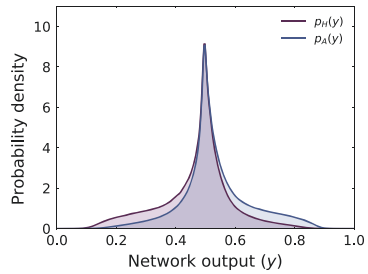
where we have expressed the overall network output distribution $p(y|\alpha)$ as a mixture of the pure-class distributions $p(y|A) \equiv p_A(y)$ and $p(y|H) \equiv p_H(y)$. We use a network trained on a balanced set of H and A events. The network is trained to associate outputs $y = 0$ and $y = 1$ with H and A events, respectively. By applying this network to another labeled dataset of equal size to the training set, we construct templates for the probability densities $p_H(y)$ and $p_A(y)$ in Eq. (11) using a nonparametric kernel density estimation method (KDE) [41]. The resulting templates are shown in Fig. 3a. We note that the pdfs do not span the entire allowed range $y \in [0, 1]$. This is expected, since the CP nature of a single event cannot be determined with complete certainty. Proper determination of the pdf shapes in the extremities – where the sensitivity is highest – requires a sufficient amount of data, which is why we devote a similarly sized data set to the template creation as to the network training.

Given a set of unlabeled data we can now estimate α by carrying out the maximization in Eq. (11) as an unbinned maximum-likelihood fit. The resulting fit to the same example dataset as used for the φ^* fit in Fig. 2b is shown in Fig. 3b. The best-fit α estimate in this case is $\hat{\alpha} = 0.67$.

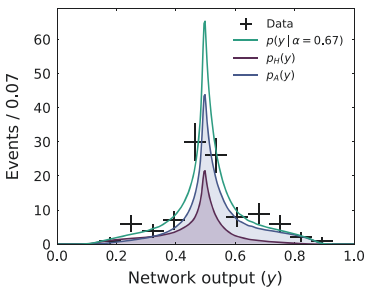
[5] Here θ represents an arbitrary model parameter, not necessarily a simple mixture parameter.

[6] In general the decision function can depend directly on the parameter values θ and θ': $s = s(\mathbf{x}; \theta, \theta')$. However, this is not the case for a mixture estimation problem like the one considered here, where $\mathbf{x}$ represents a single draw from one of the mixture model components (kinematic data from a single H or A event) and the parameter of interest is the unknown component mixture (α) of the complete dataset $\{\mathbf{x}_i\}$.

[7] This is trivially satisfied for any choice $\theta' \in (0, 1)$ when θ represents the mixture parameter of a simple two-component mixture model.



(a)



(b)

Fig. 3 **a** KDE estimate for the distribution of the network output y for H events ($p_H(y)$) and A events ($p_A(y)$), given a balanced network. **b** A fit of the mixture model $p(y|\alpha) = \alpha p_A(y) + (1-\alpha)p_H(y)$ to the same example dataset as used in Fig. 2b. Data points are shown in black and the fitted model in green. The best-fit α value is $\hat{\alpha} = 0.67$

5.1 Results

We can now compare the performance of the neural network method with that of the φ^* method of Sect. 4. To this end, we apply the network method to the same test sets as used in Fig. 4a, i.e. 10,000 datasets of 100 Pythia events each, for each of the three scenarios $\alpha = 0.5, 0.7, 0.9$. The analysis is repeated with network training and test sets with and without detector smearing. The results are given in Fig. 4b, for easy comparison with the corresponding results of the φ^* method in Fig. 4a. We fit each distribution of α estimates with a Gaussian and summarize the fit parameters in Table 1.

As for the φ^* method, we find that detector smearing significantly impacts the width of the α distribution, which increases from $\sigma_\alpha = 0.21$ to $\sigma_\alpha^{\text{det}} = 0.37$ upon inclusion of detector effects. Yet, the network approach consistently outperforms the φ^* method, as σ_α and $\sigma_\alpha^{\text{det}}$ are reduced by $\sim 22\%$ and $\sim 18\%$, respectively, compared to the φ^* results. So while the absolute widths of the α distributions in Fig. 4 illustrate that a dataset of 100 events is probably too small to obtain an accurate α estimate, the comparison with the φ^* results indicates that the relative performance gain offered by the network approach is relatively robust against detector smearing.

Similar to the φ^* method, the network method provides a mean-unbiased estimator. In order to demonstrate this we allow α to vary outside the physical range [0, 1] in our fits. However, for $\alpha > 1$, the combined mixture model $p(y|\alpha) = \alpha p_A(y) + (1-\alpha)p_H(y)$ will become negative for y values that satisfy $p_A(y)/p_H(y) < (\alpha-1)/\alpha$. This we do not allow in our fits, and in such cases we lower the α estimate until $p(y|\alpha)$ is non-negative everywhere. This choice explains the slight deviation from Gaussianity in the region around $\hat{\alpha} = 1.2$ in the bottom right plot.[8]

Figure 5 shows the distributions of α estimates for the cases of 20 events per test dataset (top row) and 500 events per test dataset (bottom row), where all sets have been generated with $\alpha = 0.7$ and no detector smearing has been included. Compared to the results with 100 events per set, σ_α for both fit methods increase (decrease) by approximately a factor $\sqrt{5}$ for the case with 20 (500) events per set, as expected from the factor 5 decrease (increase) in statistics. Thus, the relative accuracy improvement of the neural network approach over the φ^* method remains approximately the same: 30% for the 20-events case, and 25% for the 500-events case. However, the absolute spread of estimates in the 20-events case shows that this is clearly not enough statistics to obtain a useful estimate of α.

As a cross-check of the behaviour of the network fit method, we plot in Fig. 6a the distribution of the log-likelihood ratio $-2\ln(L(\alpha = 0.7)/L(\hat{\alpha}))$ for all test datasets of our benchmark point with $\alpha = 0.7$. According to Wilks' theorem [42], the distribution of this statistic should tend towards a χ^2 distribution with one degree of freedom. By overlaying a χ^2 distribution in Fig. 6a we see that this is indeed the case. Thus, confidence intervals constructed from the log-likelihood ratio for a neural network fit should have the expected coverage. In Fig. 6b we show the log-likelihood ratio curves for the example dataset used in Figs. 2b and 3b. The narrowing of the log-likelihood parabola for the network method again illustrates the increase in precision over the φ^* method.

For this study we focus only on the separation of two signal classes, not the separation of signal from background. Of course, a realistic dataset is likely to contain a significant fraction of background events. For the signal scenario studied here, the most important backgrounds are due to "fake taus" from QCD production, single Z production ($pp \to Z \to \tau\tau$), double Z and W production ($pp \to ZZ/WZ/WW \to \tau\tau + X$) and top pair production ($t\bar{t} \to WbWb \to \tau\tau + X$). While such backgrounds

[8] The same effect is not seen for the φ^* fits, as the ratio $p_A(\varphi^*)/p_H(\varphi^*) \geq 0.59$ for all φ^*, and none of the test sets prefer an α value as large as $1/(1-0.59) \approx 2.4$.

Table 1 Summary of α estimation on 10,000 independent test sets with 100 events in each set, using the φ^* fit and the neural network (NN) fit methods

True mixture parameter α	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
α Estimates (φ^* method, no detector smearing)	0.50 ± 0.27	0.71 ± 0.27	0.90 ± 0.27
α Estimates (φ^* method, with detector smearing)	0.50 ± 0.45	0.70 ± 0.46	0.90 ± 0.45
α Estimates (NN method, no detector smearing)	0.50 ± 0.21	0.70 ± 0.21	0.90 ± 0.21
α Estimates (NN method, with detector smearing)	0.48 ± 0.37	0.68 ± 0.37	0.88 ± 0.37

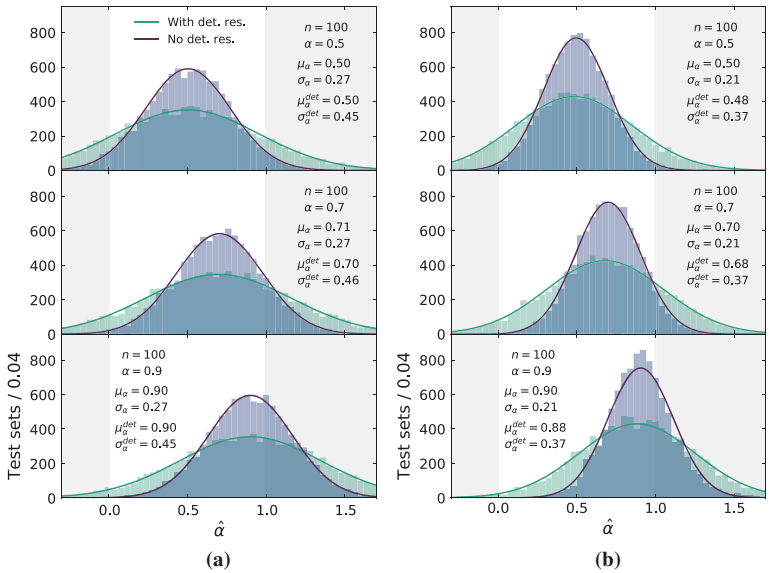


Fig. 4 Comparison of the distributions of α estimates using **a** the φ^* method and **b** the neural network method, for test sets generated with $\alpha = 0.5$ (top), $\alpha = 0.7$ (middle) and $\alpha = 0.9$ (bottom). The slight deviation from Gaussianity seen around $\hat{\alpha} = 1.2$ in the bottom right plot is due to the fact that we let α vary beyond [0, 1] in our fits, but still demand that that the mixture model $p(y|\alpha)$ is always non-negative. See the text for further details

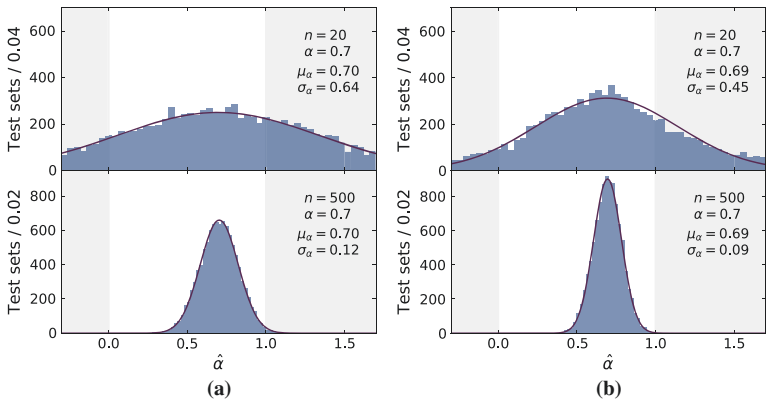


Fig. 5 Comparison of the distributions of α estimates using **a** the φ^* method and **b** the neural network method, for test sets generated with $\alpha = 0.7$. The top row shows results for test sets containing 20 events each, while the bottom row corresponds to test sets with 500 events each. The deviation from the Gaussian distribution seen at high α in the upper right plot is due to the same effect as discussed for Fig. 4

will degrade the absolute accuracy in the signal mixture estimate, it is likely to impact the φ^* method more severely than the neural network method. With one or several background components in the mixture model, the network's ability to extract information from the many-dimensional kinematic space should allow it to differentiate the background components from the signal components better than what is possible with the φ^* variable alone. We therefore expect a similar

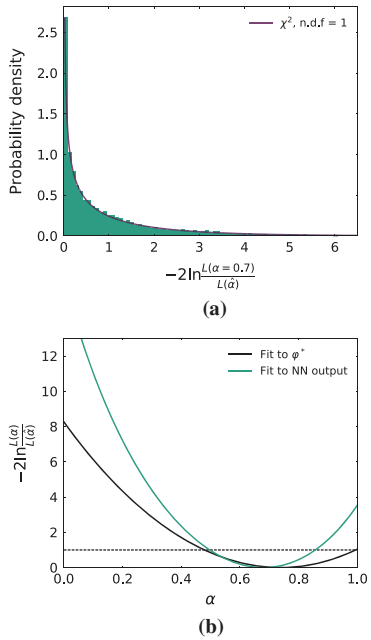


Fig. 6 **a** Distribution of the log-likelihood ratio $-2\ln(L(\alpha = 0.7)/L(\hat{\alpha}))$ for the 10,000 test sets generated with $\alpha = 0.7$. Overlaid is a χ^2 distribution for one degree of freedom. **b** Comparison of the log-likelihood ratio curves for the test dataset from Figs. 2b and 3b, using the network method (green) and the φ^* method (black). Intersection with the horizontal dashed line at $-2\ln(L(\alpha)/L(\hat{\alpha})) = 1$ illustrates the 1σ confidence intervals, which for this example are [0.48, 1.0] for the φ^* method and [0.50, 0.86] for the neural network method

or better relative performance of the network method in the presence of background, compared to the results we have presented here. There are two ways to extend the network method to take into account additional components in the mixture model: either by implementing a multi-class classifier, or by training multiple binary classifiers on pairwise combinations of the model components. Based on [43] we expect the latter approach would give the best performance.

6 Conclusions

Estimating the component weights in mixture models with largely overlapping kinematics is a generic problem in high-energy physics. In this paper we have investigated how a deep neural network approach can improve signal mixture estimates in the challenging scenario of a ditau LHC signal coming from a pair of heavy, degenerate Higgs bosons of opposite CP charge. This is a theoretically well-motivated scenario within both general and more constrained Two-Higgs-Doublet Models.

We have studied a benchmark scenario with degenerate H and A states at $m_H = m_A = 450\,\text{GeV}$. For this case we find that the neural network approach provides a $\sim 20\%$ reduction in the uncertainty of signal mixture estimates, compared to estimates based on fitting the single most discriminating kinematic variable (φ^*). However, the improved accuracy of the neural network approach comes with a greater computational complexity.

The network method we have studied here can be extended to include additional mixture components, such as one or several background processes, either by training a multi-class classifier or by training multiple binary classifiers. To increase the available statistics, the method can also be extended to work with a wider range of tau decay modes, for instance by using the "impact parameter method" described in [18].

The code used to generate events, train the network and run the maximum likelihood estimates will be made available on gitlab.com/BSML after publication.

Acknowledgements We would like to thank Andrey Ustyuzhanin and Maxim Borisyak for helpful discussions during MLHEP 2017, and Kyle Cranmer for comments and discussions during Spåtind 2018. We also thank Ørjan Dale for helpful comments. S.M. and I.S. thank the Theory Section at the Department of Physics, University of Oslo, for the kind hospitality during the completion of this work. This work was supported by the Research Council of Norway through the FRIPRO grant 230546/F20.

Appendix A Supplementary figures

A simple scan of the high-mass parameter regions of the (SM-aligned) lepton-specific and type-I THDMs is performed to illustrate the parameter dependence of the ditau signal strengths $\sigma(pp \to H) \times \mathcal{B}(H \to \tau\tau)$ and $\sigma(pp \to A) \times \mathcal{B}(A \to \tau\tau)$, as well as the mixture parameter α. The results are shown in Fig. 7. The parameters $m_H = m_A = m_{H^\pm}$, $\tan\beta$ and m_{12}^2 are varied in the scan, while we fix the light Higgs mass $m_h = 125\,\text{GeV}$ and the neutral scalar mixing parameter $\sin(\beta - \alpha') = 1$ to ensure perfect SM alignment for the light state h. The NLO cross sections are calculated with SusHi 1.6.1, while branching ratios are calculated using 2HDMC 1.7.0. We test the parameter points against constraints from the various collider searches for Higgs bosons using HiggsBounds 4.3.1 [44–48], while the-

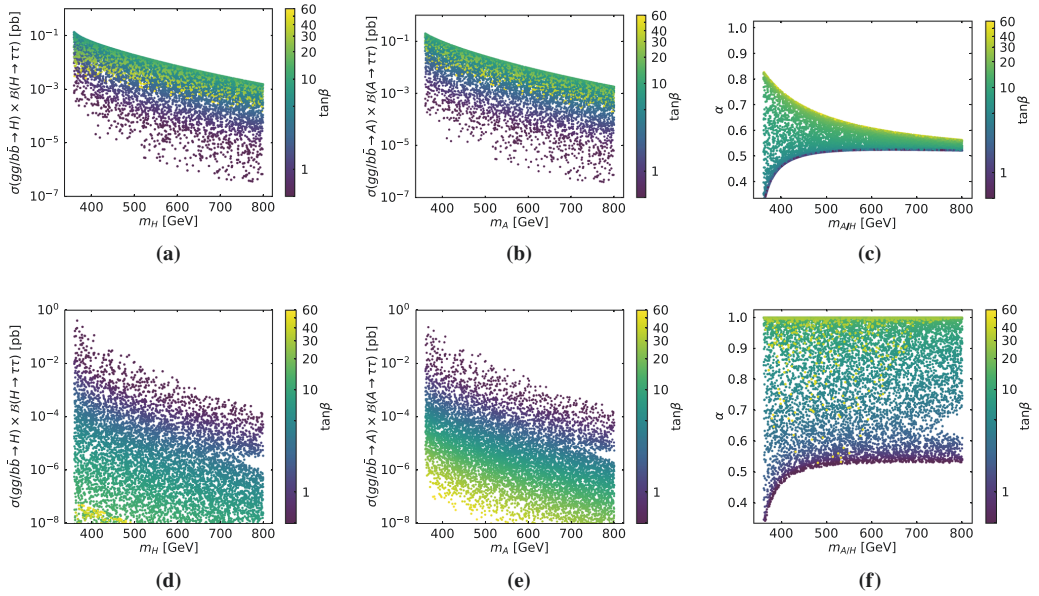


Fig. 7 *Top row, lepton-specific THDM:* **a** Signal strength for $pp \to H \to \tau\tau$, as a function of m_H and $\tan\beta$. **b** Similar result for $pp \to A \to \tau\tau$. **c** The ratio of the signal strength of $pp \to A \to \tau\tau$ to the total ditau signal strength, as defined in Eq. (1). *Bottom row:* Corresponding results within the type-I THDM

oretical constraints are checked with 2HDMC. Constraints from flavour physics, in particular $\mathcal{B}(b \to s\gamma)$, disfavour parameter regions at very low $\tan\beta$ in the type-I and lepton-specific THDMs. These constraints were not included in the simple scan.

References

1. P. Baldi, P. Sadowski, D. Whiteson, Nat. Commun. **5**, 4308 (2014). https://doi.org/10.1038/ncomms5308
2. P. Baldi, P. Sadowski, D. Whiteson, Phys. Rev. Lett. **114**(11), 111801 (2015). https://doi.org/10.1103/PhysRevLett.114.111801
3. A. Farbin, PoS **ICHEP2016**, 180 (2016)
4. V.A. Kuzmin, V.A. Rubakov, M.E. Shaposhnikov, Phys. Lett. B **155**, 36 (1985). https://doi.org/10.1016/0370-2693(85)91028-7
5. G.W. Anderson, L.J. Hall, Phys. Rev. D **45**, 2685 (1992). https://doi.org/10.1103/PhysRevD.45.2685
6. N. Turok, J. Zadrozny, Nucl. Phys. B **358**, 471 (1991). https://doi.org/10.1016/0550-3213(91)90356-3
7. N. Turok, J. Zadrozny, Nucl. Phys. B **369**, 729 (1992). https://doi.org/10.1016/0550-3213(92)90284-I
8. J.O. Andersen, T. Gorda, A. Helset, L. Niemi, T.V.I. Tenkanen, A. Tranberg, A. Vuorinen, D.J. Weir, Phys. Rev. Lett. **121**(19), 191802 (2018). https://doi.org/10.1103/PhysRevLett.121.191802
9. M. Aaboud, JHEP **01**, 055 (2018). https://doi.org/10.1007/JHEP01(2018)055
10. A.M. Sirunyan, JHEP **09**, 007 (2018). https://doi.org/10.1007/JHEP09(2018)007
11. J. Bernon, J.F. Gunion, H.E. Haber, Y. Jiang, S. Kraml, Phys. Rev. D **92**(7), 075004 (2015). https://doi.org/10.1103/PhysRevD.92.075004
12. A. Haarr, A. Kvellestad, T.C. Petersen (2016) preprint. arXiv:1611.05757
13. S.P. Martin (1997). https://doi.org/10.1142/9789812839657_0001. https://doi.org/10.1142/9789814307505_0001. [Adv. Ser. Direct. High Energy Phys. **18**, 1 (1998)]
14. P.S. Bhupal Dev, A. Pilaftsis, JHEP **12**, 024 (2014). https://doi.org/10.1007/JHEP11(2015)147. https://doi.org/10.1007/JHEP12(2014)024. [Erratum: JHEP **11**, 147 (2015)]
15. J.R. Dell'Aquila, C.A. Nelson, Nucl. Phys. B **320**(1), 61 (1989). https://doi.org/10.1016/0550-3213(89)90211-3
16. M. Kramer, J.H. Kuhn, M.L. Stong, P.M. Zerwas, Z. Phys. C **64**, 21 (1994). https://doi.org/10.1007/BF01557231
17. G.R. Bower, T. Pierzchala, Z. Was, M. Worek, Phys. Lett. B **543**, 227 (2002). https://doi.org/10.1016/S0370-2693(02)02445-0
18. S. Berge, W. Bernreuther, S. Kirchner, Phys. Rev. D **92**, 096012 (2015). https://doi.org/10.1103/PhysRevD.92.096012
19. R. Józefowicz, E. Richter-Was, Z. Was, Phys. Rev. D **94**(9), 093001 (2016). https://doi.org/10.1103/PhysRevD.94.093001
20. E. Barberio, B. Le, E. Richter-Was, Z. Was, D. Zanzi, J. Zaremba, Phys. Rev. D **96**(7), 073002 (2017). https://doi.org/10.1103/PhysRevD.96.073002
21. R.V. Harlander, S. Liebler, H. Mantler, Comput. Phys. Commun. **184**, 1605 (2013). https://doi.org/10.1016/j.cpc.2013.02.006

22. R.V. Harlander, S. Liebler, H. Mantler, Comput. Phys. Commun. **212**, 239 (2017). https://doi.org/10.1016/j.cpc.2016.10.015
23. R.V. Harlander, W.B. Kilgore, Phys. Rev. Lett. **88**, 201801 (2002). https://doi.org/10.1103/PhysRevLett.88.201801
24. R.V. Harlander, W.B. Kilgore, Phys. Rev. D **68**, 013001 (2003). https://doi.org/10.1103/PhysRevD.68.013001
25. S. Actis, G. Passarino, C. Sturm, S. Uccirati, Phys. Lett. B **670**, 12 (2008). https://doi.org/10.1016/j.physletb.2008.10.018
26. R. Harlander, P. Kant, JHEP **12**, 015 (2005). https://doi.org/10.1088/1126-6708/2005/12/015
27. K.G. Chetyrkin, J.H. Kuhn, M. Steinhauser, Comput. Phys. Commun. **133**, 43 (2000). https://doi.org/10.1016/S0010-4655(00)00155-7
28. D. Eriksson, J. Rathsman, O. Stal, Comput. Phys. Commun. **181**, 189 (2010). https://doi.org/10.1016/j.cpc.2009.09.011
29. T. Sjostrand, S. Mrenna, P.Z. Skands, JHEP **05**, 026 (2006). https://doi.org/10.1088/1126-6708/2006/05/026
30. T. Sjostrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008). https://doi.org/10.1016/j.cpc.2008.01.036
31. C. Patrignani, Chin. Phys. C **40**(10), 100001 (2016). https://doi.org/10.1088/1674-1137/40/10/100001
32. A.L. Maas, A.Y. Hannun, A.Y. Ng, in *International conference on machine learning*, vol. 30, p. 3 (2013)
33. N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, J. Mach. Learn. Res. **15**(1), 1929 (2014)
34. K. He, X. Zhang, S. Ren, J. Sun, CoRR arXiv:1502.01852 (2015)
35. S. Ioffe, C. Szegedy, CoRR arXiv:1502.03167 (2015)
36. D.P. Kingma, J. Ba, CoRR arXiv:1412.6980 (2014)
37. F. Chollet, et al. Keras. https://github.com/fchollet/keras (2015)
38. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems (2015). Software available from http://tensorflow.org/
39. J. Friedman, T. Hastie, R. Tibshirani, Ann. Stat. **28**(2), 337 (2000). https://doi.org/10.1214/aos/1016218223
40. K. Cranmer, J. Pavez, G. Louppe (2015) preprint. arXiv:1506.02169
41. K.S. Cranmer, Comput. Phys. Commun. **136**, 198 (2001). https://doi.org/10.1016/S0010-4655(00)00243-5
42. S.S. Wilks, Ann. Math. Stat. **9**(1), 60 (1938). https://doi.org/10.1214/aoms/1177732360
43. K. Cranmer, J. Pavez, G. Louppe, W.K. Brooks, J. Phys. Conf. Ser. **762**(1), 012034 (2016). https://doi.org/10.1088/1742-6596/762/1/012034
44. P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, K.E. Williams, Comput. Phys. Commun. **181**, 138 (2010). https://doi.org/10.1016/j.cpc.2009.09.003
45. P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, K.E. Williams, Comput. Phys. Commun. **182**, 2605 (2011). https://doi.org/10.1016/j.cpc.2011.07.015
46. P. Bechtle, O. Brein, S. Heinemeyer, O. Stal, T. Stefaniak, G. Weiglein, K. Williams, PoS **CHARGED2012**, 024 (2012)
47. P. Bechtle, O. Brein, S. Heinemeyer, O. Stal, T. Stefaniak, G. Weiglein, K.E. Williams, Eur. Phys. J. C **74**(3), 2693 (2014). https://doi.org/10.1140/epjc/s10052-013-2693-2
48. P. Bechtle, S. Heinemeyer, O. Stal, T. Stefaniak, G. Weiglein, Eur. Phys. J. C **75**(9), 421 (2015). https://doi.org/10.1140/epjc/s10052-015-3650-z

Paper III

Smoking gun dark matter signatures in the MeV range

Contribution to
Science with e-ASTROGAM: A space mission for MeV - GeV gamma-ray astrophysics [3]

Smoking gun dark matter signatures in the MeV range

Torsten Bringmann[1], Andrzej Hryczuk[1], Are Raklev[1], Inga Strümke[2], and Jeriek Van den Abeele[1]

[1]Department of Physics, University of Oslo, Box 1048, NO-0371 Oslo, Norway

[2]University of Bergen, Institute for Physics and Technology, Postboks 7803, N-5020 Bergen, Norway

October 10, 2017

Science questions – The nature of the astrophysically and cosmologically observed dark matter (DM) in the universe [1] is one of the big open scientific questions today. While it is clear that DM must be non-baryonic, the list of possible explanations in terms of a new particle is long [2]. Given that all evidence for DM so far is of gravitational origin, any non-gravitational DM signal would be a major breakthrough towards determining the identity of those DM particles.

Among the most favourite DM candidates are weakly interacting massive particles (WIMPs), with masses and coupling strengths at the electroweak scale. Besides the fact that many of these are theoretically very well motivated, such as the supersymmetric neutralino [3], an attractive feature of this class of candidates is that the observed DM abundance today can straight-forwardly be explained by the thermal production of WIMPs in the early universe. In recent years however – triggered not the least by the lasting absence of any undisputed WIMP signals, despite immense experimental efforts – the focus of the community has started to shift beyond WIMPs as the main DM paradigm.

For example, it was pointed out that thermal production is also an attractive option for smaller DM masses [4]. Other relevant DM models with (sub-)GeV masses include light gravitino DM [5], inelastic DM [6], light scalar DM [7] or secluded DM [8]. Models in this mass range have received significant interest because they could have easily escaped the ever more stringent constraints from direct DM detection experiments (for a suggestion of how to overcome the lack of sensitivity of traditional methods in this mass range, see e.g. Ref. [10]). From the indirect detection perspective, an intriguing feature of such models is furthermore that the center-of-mass energy, and hence the energy of final state quarks, is at the same mass scale as standard model hadronic states. As we argue in this contribution, this can lead to a potentially rich phenomenology in MeV gamma rays that may allow to draw far-reaching conclusions about the nature of the DM particles and the underlying theory.

Importance of gamma-ray observations – Gamma rays from both decaying and annihilating DM have sometimes been argued to be the *golden channel* of indirect DM searches [9] because they directly point back to their sources and hence provide the potentially most accurate way to probe the astronomically observed DM distribution *in situ*. Furthermore, they may carry distinct spectral features that can both act as 'smoking gun' signals for the particle nature of DM and convey further detailed information about the nature of these particles.

Motivated by the WIMP case, the main focus has traditionally been on spectral features in the 100 GeV – TeV range, with relevant limits presented e.g. in Ref. [11]; also exotic line contributions in the keV range have been scrutinized in detail, where a signal could be expected from decaying sterile neutrino DM [12]. Here, we point out that also the largely neglected MeV range is very well motivated in this respect (for earlier work, see Refs. [13, 14]), and hence ideally suited for searches with e-ASTROGAM.

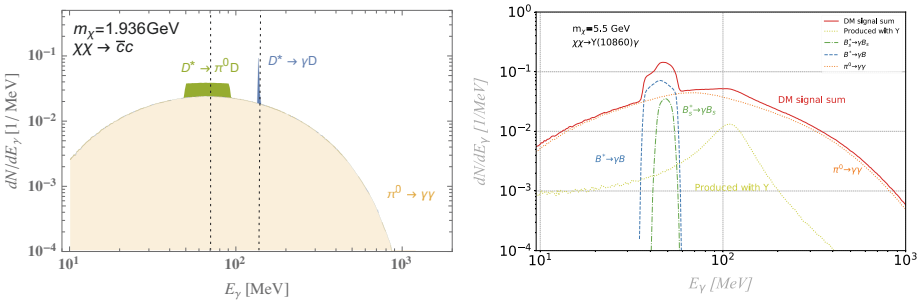


Figure 1: **Left:** Example of the expected gamma-ray spectrum for DM annihilation into charm quarks, with a DM mass m_χ just above the kinematic threshold to produce D-mesons. The sharp spectral features result from the indicated meson transitions, while the background is mostly due to $\pi^0 \to \gamma\gamma$. For more details, see Ref. [14]. **Right:** Gamma-ray spectrum from DM annihilation through the quarkonium channel $\chi\chi \to \Upsilon(10860)\gamma$. The three visible spectral features are due to two different meson transitions and the photon produced in conjunction with the quarkonium. For more details, see Ref. [18].

In fact, gamma-ray and cosmic microwave background observations already put significant constraints on light DM candidates, and e-ASTROGAM would imply an additional boost in sensitivity [15]. As we show here, hadronic final states from DM decay or annihilation could furthermore lead to a plethora of potential smoking-gun signatures for a DM signal in MeV gamma rays that only a dedicated mission like e-ASTROGAM may be able to detect.

Expected results with e-ASTROGAM – Among the various processes that could potentially lead to spectral features in MeV gamma rays (see also Ref. [14] for an overview, and Ref. [16] for further examples), we will focus here on standard model meson transitions and quarkonium resonances. We consider a center-of-mass energy of the annihilating DM pair, or DM mass in the case of decaying DM, that is close to the threshold for the production of (excited) heavy mesons. The de-excitation of excited meson states in the final state, via the emission of a photon or neutral pion, will then generate box-like signatures (which in the case of photon emission can be almost monochromatic).

For illustration, we show in Fig. 1 an example where DM is assumed to annihilate dominantly into $\bar{c}c$ pairs. In this example, both types of de-excitation processes lead to spectral features that are clearly visible above the standard 'background' part of the signal, resulting from decaying neutral pions that are copiously produced in fragmentations and decays of heavier mesons. Implementing a realistic modelling of the expected astrophysical background, we have shown that the sensitivity of e-ASTROGAM to this DM annihilation channel improves by a factor of up to about 2 by taking into account these spectral features, compared to using the standard pion bump as a signal template [14]. For $\bar{b}b$ final states, the effect can be twice as large. We note that the exact form and location of these spectral features are very specific for each final state. This allows, in principle, a highly accurate reconstruction not only of the DM mass but also of the branching ratios for the DM decay or annihilation channels.

The possibility of MeV gamma-ray features from annihilation into heavy meson pairs also raises the issue of contributions from quarkonia. Either through the process $\chi\chi \to (\bar{Q}Q)\gamma$, where gamma-rays are produced both directly and through subsequent decay into (excited) heavy mesons, or heavy-meson production enhanced by a quarkonium resonance $\chi\chi \to (\bar{Q}Q) \to M_A M_B$, where M_A and M_B are two heavy mesons with radiative decays. An example of the resulting expected spectrum for a DM mass $m_\chi = 5.5$ GeV and the channel $\chi\chi \to \Upsilon(10860)\gamma$ is shown in Fig. 1 (right). Here structures from three processes, $B^* \to \gamma B$, $B_s^* \to \gamma B_s$, and direct production in the annihilation, can all be identified. Notably, such a signal would also exist in the annihilation of

sub-GeV DM into light quarkonium states, e.g. $\chi\chi \to \eta^{(\prime)}\gamma$, with subsequent decay of the $\eta^{(\prime)}$ into photon pairs.

Furthermore, it is well known experimentally that for heavy-meson production at e^+e^- colliders, quarkonium resonances can be dominant near threshold [17]. We have explored dark matter annihilation through the related vector currents $\chi\Gamma^\mu\chi\bar{Q}\gamma_\mu Q$. Using collider data as input to our model we observe significant enhancement of the MeV features [18] due to these resonances. We also find that the existence and dominance of different processes is highly dependent on the structure of the DM-quark interaction and the nature of the DM particle, e.g. as seen in the well-known suppression of the vector current for Majorana or scalar DM [13].

In conclusion, the sensitivity gap in the MeV range explored by e-ASTROGAM is a window of opportunity to detect new physics – not only by confirming the particle nature of DM, but with the additional potential of closing in on some of its detailed properties, like the DM particle's mass, its branching ratios to quark final states and, to some degree, its underlying interaction structure.

References

[1] P. A. R. Ade *et al.* [Planck Collaboration], Astron. Astrophys. **594**, A13 (2016) [arXiv:1502.01589 [astro-ph.CO]].

[2] J. L. Feng, Ann. Rev. Astron. Astrophys. **48**, 495 (2010) [arXiv:1003.0904 [astro-ph.CO]].

[3] G. Jungman, M. Kamionkowski and K. Griest, Phys. Rept. **267**, 195 (1996) [hep-ph/9506380].

[4] J. L. Feng and J. Kumar, Phys. Rev. Lett. **101**, 231301 (2008) [arXiv:0803.4196 [hep-ph]].

[5] F. Takayama and M. Yamaguchi, Phys. Lett. B **485** (2000) 388 doi:10.1016/S0370-2693(00)00726-7 [hep-ph/0005214].

[6] D. Tucker-Smith and N. Weiner, Phys. Rev. D **64** (2001) 043502 doi:10.1103/PhysRevD.64.043502 [hep-ph/0101138].

[7] C. Boehm and P. Fayet, Nucl. Phys. B **683** (2004) 219 doi:10.1016/j.nuclphysb.2004.01.015 [hep-ph/0305261].

[8] M. Pospelov, A. Ritz and M. B. Voloshin, Phys. Lett. B **662** (2008) 53 doi:10.1016/j.physletb.2008.02.052 [arXiv:0711.4866 [hep-ph]].

[9] T. Bringmann and C. Weniger, Phys. Dark Univ. **1**, 194 (2012) [arXiv:1208.5481 [hep-ph]].

[10] R. Essig, J. Mardon and T. Volansky, Phys. Rev. D **85**, 076007 (2012) [arXiv:1108.5383 [hep-ph]].

[11] M. Ackermann *et al.* [Fermi-LAT Collaboration], Phys. Rev. D **91**, no. 12, 122002 (2015) [arXiv:1506.00013 [astro-ph.HE]]; A. Abramowski *et al.* [H.E.S.S. Collaboration], Phys. Rev. Lett. **110**, 041301 (2013) [arXiv:1301.1173 [astro-ph.HE]].

[12] For the much discussed potential discovery of such a line, see E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein and S. W. Randall, Astrophys. J. **789**, 13 (2014) [arXiv:1402.2301 [astro-ph.CO]]; A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi and J. Franse, Phys. Rev. Lett. **113**, 251301 (2014) [arXiv:1402.4119 [astro-ph.CO]].

[13] M. Srednicki, S. Theisen and J. Silk, Phys. Rev. Lett. **56**, 263 (1986) Erratum: [Phys. Rev. Lett. **56**, 1883 (1986)]; S. Rudaz, Phys. Rev. Lett. **56**, 2128 (1986); L. Bergstrom, Nucl. Phys. B **325**, 647 (1989).

[14] T. Bringmann, A. Galea, A. Hryczuk and C. Weniger, Phys. Rev. D **95**, no. 4, 043002 (2017) [arXiv:1610.04613 [hep-ph]].

[15] See, e.g., I. A. Antonelli *et al.* and R. Bartels *et al.* (this contribution).

[16] V. Brdar, J. Kopp, J. Liu, A. Merle and X. Wang (this contribution).

[17] B. Aubert *et al.* [BaBar Collaboration], Phys. Rev. Lett. **102** (2009) 012001 [arXiv:0809.4120 [hep-ex]].

[18] A. Raklev, I. Strümke, J. Van den Abeele, *in progress* (2017).

Appendices

Appendix A

Gaugino mediated calculations

This appendix contains calculations upon which the results in section 4.5 and [1] are based.

$$d^2\theta = -\frac{1}{4}d\theta^\alpha d\theta^\beta \varepsilon_{\alpha\beta}\,, \qquad d^2\theta^\dagger = -\frac{1}{4}d\theta^\dagger_{\dot\alpha} d\theta^\dagger_{\dot\beta}\varepsilon^{\dot\alpha\dot\beta} \tag{A.1}$$

A.1 Equations of motion for Higgs auxiliary field in the framework with a source field

The equations of motion for F are in general

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu F)}\right) = \frac{\partial\mathcal{L}}{\partial F}\,, \tag{A.2}$$

and since F here denotes the auxiliary field of a chiral supermultiplet, it is not a physical field and thus does not propagate. This means that there exist no kinetic terms for F or $F^\dagger$, i.e.

$$\frac{\partial\mathscr{L}}{\partial F} = \frac{\partial\mathscr{L}}{\partial F^\dagger} = 0\,, \tag{A.3}$$

which will be used later. In the following, as explained in section 4.5, M is the cutoff up to which the effective theory is valid, and the factor $\frac{1}{V_{D-4}}$ appears upon integrating over all dimensions higher than 4. The terms in the Lagrangian containing Higgs fields come from

- The superpotential
$$W = \bar u y_u Q H_u - \bar d y_d Q H_d - \bar e y_e L H_d + \mu H_u H_d + \lambda S H_u H_d \tag{A.4}$$
where the term with the S must be accompanied by a factor $\frac{1}{M^{D-3}}$ in the Lagrangian,
- The Kähler potential
$$\begin{aligned} K = {} & H_u^\dagger e^V H_u + H_d^\dagger e^V H_d + \frac{1}{M^{D-3}}\left[S\left(aH_u^\dagger H_d^\dagger + b_u H_u^\dagger H_u + b_d H_d^\dagger H_d\right) + \text{h.c.}\right] \\ & + \frac{1}{M^{D-2}}\left[S^\dagger S\left(c_u H_u^\dagger H_u + c_d H_d^\dagger H_d + \left(dH_u H_d + \text{h.c.}\right)\right)\right. \\ & \left. + SS\left(fH_u^\dagger H_d^\dagger + e_u H_u^\dagger H_u + e_d H_d^\dagger H_d\right) + \text{h.c.}\right] + \ldots \end{aligned} \tag{A.5}$$
where the dots denote higher orders in S. The first two terms are the only contributions from the MSSM Kähler potential, see eq. (4.25), where it does not matter which vector supermultiplet V is, since the interesting contribution from the D-term will pick out only the Higgs auxiliary fields, see below or eq. (4.9.11) in [44].

Collecting all terms in the Lagrangian containing F_H, up to first order in S, and not writing out the volume factor or the M to avoid clutter,

$$\begin{aligned}
\mathscr{L}_{F_H} = &\overbrace{y_u F_{H_u}\phi_{\bar{u}}\phi_Q - y_d F_{H_d}\phi_{\bar{d}}\phi_Q - y_e F_{H_d}\phi_{\bar{e}}\phi_L + \mu F_{H_d}\phi_{H_u} + \mu F_{H_u}\phi_{H_d} + \text{h.c.}}^{[W_{\text{MSSM}}]_F + \text{h.c.}} \\
&+ \overbrace{\lambda\phi_s\left(F_{H_u}\phi_{H_d} + \phi_{H_u}F_{H_d}\right) + \lambda^\dagger\phi_s^\dagger\left(F_{H_u}^\dagger\phi_{H_d}^\dagger + \phi_{H_u}^\dagger F_{H_d}^\dagger\right)}^{[\lambda S H_u H_d]_F + \text{h.c.}} \\
&+ \overbrace{F_{H_u}^\dagger F_{H_u} + F_{H_d}^\dagger F_{H_d}}^{[H^\dagger e^V H]_D} \\
&+ a^\dagger\overbrace{F_s^\dagger F_{H_d}\phi_{H_u} + F_s^\dagger\phi_{H_d}F_{H_u}}^{[S^\dagger H_u H_d]_D} + a\overbrace{F_s F_{H_d}^\dagger\phi_{H_u}^\dagger + F_s\phi_{H_d}^\dagger F_{H_u}^\dagger}^{[S H_u^\dagger H_d^\dagger]_D} \\
&+ b_u\overbrace{\left(\phi_s F_{H_u}F_{H_u}^\dagger + F_s\phi_{H_u}F_{H_u}^\dagger + \psi_{H_u}\psi_s F_{H_u}^\dagger\right)}^{[S H_u^\dagger H_u]_D} + b_u^\dagger\overbrace{\left(\phi_s^\dagger F_{H_u}F_{H_u}^\dagger + F_s^\dagger\phi_{H_u}^\dagger F_{H_u} + \psi_{H_u}^\dagger\psi_s^\dagger F_{H_u}\right)}^{[S^\dagger H_u^\dagger H_u]_D} \\
&+ b_d\overbrace{\left(\phi_s F_{H_d}F_{H_d}^\dagger + F_s\phi_{H_d}F_{H_d}^\dagger + \psi_{H_d}\psi_s F_{H_d}^\dagger\right)}^{[S H_d^\dagger H_d]_D} + b_d^\dagger\overbrace{\left(\phi_s^\dagger F_{H_d}F_{H_d}^\dagger + F_s^\dagger\phi_{H_d}^\dagger F_{H_d} + \psi_{H_d}^\dagger\psi_s^\dagger F_{H_d}\right)}^{[S^\dagger H_d^\dagger H_d]_D}.
\end{aligned} \tag{A.6}$$

The equations of motion for H_u are thus

$$\begin{aligned}
\frac{\partial\mathscr{L}}{\partial F_{H_u}} &= \overbrace{\underbrace{\mu\phi_{H_d} + y_u\phi_{\bar{u}}\phi_Q}_{\equiv\omega^u}}^{[W_{\text{MSSM}}]_F} + \overbrace{\lambda\phi_s\phi_{H_d}}^{[\lambda S H_u H_d]_F} + \overbrace{F_{H_u}^\dagger}^{[H_u^\dagger e^V H_u]_D} + \overbrace{a^\dagger F_s^\dagger\phi_{H_d}}^{[S^\dagger H_u H_d]_D} + \overbrace{b_u\phi_s F_{H_u}^\dagger}^{[S H_u^\dagger H_u]_D} + b_u^\dagger\overbrace{\left(\phi_{H_u}^\dagger F_s^\dagger + \psi_{H_u}^\dagger\psi_s^\dagger + \phi_s^\dagger F_{H_u}^\dagger\right)}^{[S^\dagger H_u^\dagger H_u]_D} = 0\,, \\
\frac{\partial\mathscr{L}}{\partial F_{H_u}^\dagger} &= \omega_{\text{h.c.}}^u + \lambda^\dagger\phi_s^\dagger\phi_{H_d}^\dagger + F_{H_u} + aF_s\phi_{H_d}^\dagger + b_u^\dagger\phi_s^\dagger F_{H_u} + b_u\left(\phi_{H_u}F_s + \psi_{H_u}\psi_s + \phi_s F_{H_u}\right) = 0\,,
\end{aligned} \tag{A.7}$$

and correspondingly for H_d. The solutions are

$$\begin{aligned}
F_{H_{u,d}}^\dagger &= -\frac{F_s^\dagger\left(b_u\phi_{H_{u,d}}^\dagger + a\phi_{H_{d,u}}\right) + b_u\psi_{H_{u,d}}^\dagger\psi_s^\dagger + \omega^{u,d} + \lambda\phi_s\phi_{H_{d,u}}}{1 + \phi_s + \phi_s^\dagger}\,, \\
F_{H_{u,d}} &= -\frac{F_s\left(b_u\phi_{H_{u,d}} + a^\dagger\phi_{H_{d,u}}^\dagger\right) + b_u\psi_{H_{u,d}}\psi_s + \omega_{\text{h.c.}}^{u,d} + \lambda^\dagger\phi_s^\dagger\phi_{H_{d,u}}^\dagger}{1 + \phi_s + \phi_s^\dagger}
\end{aligned} \tag{A.8}$$

The term proportional to a contributes to the μ term (and thus, among other things, to the supersymmetric three-scalar interactions $M_{in}^* y^{jkn}\phi^{*i}\phi_j\phi_k$ present in the MSSM) and does not break SUSY. Together with $b_{u,d}$ it also contributes to $B\mu$, see eq. (7) of [120], which breaks SUSY, but this contribution vanishes for $b_{u,d} = 0$, so the SUSY breaking stems from the $b_{u,d}$ terms. The ϕ_S dependence of the numerator stems solely from the superpotential term $[\lambda S H_u H_d]_F$. It gives rise to a ϕ^4 interaction, which is not relevant for the present discussion. Neither is the part containing source brane fermions ψ_S. The solutions, with volume factors and EFT cutoff written out, keeping only the relevant parts for the present discussion, letting the auxiliary field of S acquire a VEV and setting its other components to zero finally yields

$$F_{H_{u,d}} = -\frac{b_{u,d}\langle F_S\rangle\phi_{H_{u,d}}}{V_{D-4}M^{D-3}} = -b_{u,d}\left(\frac{M_c}{M}\right)^{D-4}\frac{\langle F_S\rangle}{M}\,\phi_{H_{u,d}}\,, \tag{A.9}$$

replacing the extra dimensions' volume by the compactification scale in the last step.

A.2 Regarding scalar interaction terms

Substituting F_{H_u} and F_{H_d} from eq. (A.9) in superpotential part of the Lagrangian yields the trilinear terms, eq. (4.57) and repeated here for convenience

$$\mathscr{L}_{\text{trilinear}} = \left(\frac{M_c}{M}\right)^{D-4} \frac{\langle F_S\rangle}{M}\left(-b_u\phi_{\bar{u}}y_u\phi_{H_u}\phi_Q + b_d\phi_{\bar{d}}y_d\phi_{H_d}\phi_Q + b_d\phi_{\bar{e}}y_e\phi_{H_d}\phi_L + \text{h.c.}\right). \tag{A.10}$$

These are (scalar)^3 interactions that break SUSY. This can be seen by direct comparison with $\mathscr{L}_{\text{soft}}$ or simply by realising that these terms contain the SUSY breaking source field.

A.2.1 Supersymmetric three-scalar interactions

In the transition from eq. (A.8) to eq. (A.9), the $\omega^{u,d}$ term was ommitted. Including it, i.e. inserting $F_{H_{u,d}} = \omega^{u,d}_{\text{h.c.}}$ and $F^\dagger_{H_{u,d}} = \omega^{u,d}$ in the superpotential in the Lagrangian yields

$$\begin{aligned}\mathscr{L} &\supset y_u\left(\omega^u_{\text{h.c.}}\phi_{\bar{u}}\phi_Q + \omega^u\phi^\dagger_{\bar{u}}\phi^\dagger_Q\right) + \dots \\ &= y_u\left(\mu^\dagger\phi^\dagger_{H_d}\phi_{\bar{u}}\phi_Q + \mu\phi_{H_d}\phi^\dagger_{\bar{u}}\phi^\dagger_Q\right) + \dots\end{aligned} \tag{A.11}$$

which means that these interactions were already present in the MSSM and thus don't break SUSY, see for instance [44] fig. 3.2, and the corresponding eq. (3.2.18).

A.2.2 Higher order scalar interactions

Other terms what were discarded in the transition from eq. (A.8) to eq. (A.9), are the ones containing a and λ. Keeping them would modify eq. (A.9) to

$$F_{H_{u,d}} = -\left(\frac{M_c}{M}\right)^{D-4}\left(\frac{\langle F_S\rangle}{M}(b_{u,d}\phi_{H_{u,d}} + a^\dagger\phi^\dagger_{H_{d,u}}) - \frac{\phi^\dagger_s}{M}\lambda^\dagger\phi^\dagger_{H_{d,u}}\right), \tag{A.12}$$

and the above Lagrangian to

$$\begin{aligned}\mathscr{L} \supset \left(\frac{M_c}{M}\right)^{D-4}\Bigg\{&\frac{\langle F_S\rangle}{M}\Big(-b_u\phi_{H_u}\phi_{\bar{u}}y_u\phi_Q - a\phi^\dagger_{H_d}\phi_{\bar{u}}y_u\phi_Q \\ &+b_d s\phi_{H_d}\phi_{\bar{d}}y_d\phi_Q + a\phi^\dagger_{H_u}\phi_{\bar{d}}y_d\phi_Q \\ &+b_d\phi_{H_d}\phi_{\bar{e}}y_e\phi_L + a\phi^\dagger_{H_u}\phi_{\bar{e}}y_e\phi_L \\ &-\mu(b_u+b_d)\phi_{H_u}\phi_{H_d}\Big) \\ &-\frac{\lambda^\dagger\phi^\dagger_S}{M}\Big(+\phi^\dagger_{H_d}\phi_{\bar{u}}y_u\phi_Q + \phi^\dagger_{H_u}\phi_{\bar{d}}y_d\phi_Q + \phi^\dagger_{H_u}\phi_{\bar{d}}y_e\phi_Q \\ &+\mu(|\phi_{H_d}|^2 + |\phi_{H_u}|^2)\Big) \\ &+a^\dagger(|\phi_{H_u}|^2 + |\phi_{H_d}|^2)\Bigg\} + \text{h.c.},\end{aligned} \tag{A.13}$$

i.e. including (scalar)^4 interactions, as announced earlier.

A.2.3 Trilinears from field redefinition

Another way to derive the trilinear terms is a holomorphic field redefinition, as remarked by [121] as well as the referee for [1].

The referee remarked that this "shows that once the $b_{u,d}$ terms are included, there is no symmetry that protects the theory from explicit trilinear terms, so omitting them in the first place appears to be ad-hoc." However, explicit trilinears were not omitted due to a symmetry but due to the extra-dimensional setup — the model is defined in D dimensions, where, e.g., $S\bar{u}QH_u$ cannot appear in the superpotential because S and the squarks are localised on different branes.

The field redefinition is expressed as follows (not distinguishing between H_u and H_d)

$$H = H'\left(1 - \frac{b}{M}S\right), \tag{A.14}$$

which reformulates the Kähler potential in terms of H' as

$$\begin{aligned}
K &= H^\dagger H + \frac{b}{M}(SH^\dagger H + S^\dagger H^\dagger H) + \frac{c}{M^2}S^\dagger S H^\dagger H + \frac{e}{M^2}(SSH^\dagger H + S^\dagger S^\dagger H^\dagger H) + \dots \\
&= H'^\dagger\left(1 - \frac{b}{M}S^\dagger\right)H'\left(1 - \frac{b}{M}S\right) + \frac{b}{M}\left(SH^\dagger\left(1 - \frac{b}{M}S^\dagger\right)H\left(1 - \frac{b}{M}S\right) + S^\dagger H^\dagger\left(1 - \frac{b}{M}S^\dagger\right)H\left(1 - \frac{b}{M}S\right)\right) + \dots \\
&= H'^\dagger H' - \frac{b}{M}S^\dagger H'^\dagger H' - \frac{b}{M}SH'^\dagger H' + \left(\frac{b}{M}\right)^2 S^\dagger S H^\dagger H + \frac{b}{M}\left\{SH'^\dagger H' - \frac{b}{M}SSH'^\dagger H' - \frac{b}{M}S^\dagger S H'^\dagger H'\right. \\
&\quad \left. + \left(\frac{b}{M}\right)^2 SSS^\dagger H'^\dagger H' + S^\dagger H'^\dagger H' - \frac{b}{M}S^\dagger S^\dagger H'^\dagger H' - \frac{b}{M}S^\dagger S H'^\dagger H' + \left(\frac{b}{M}\right)^2 SS^\dagger S^\dagger H'^\dagger H'\right\} + \dots \\
&= H'^\dagger H' - \left(\frac{b}{M}\right)^2 S^\dagger S H'^\dagger H' + \frac{b}{M}\left\{-\frac{b}{M}\left(SS + S^\dagger S^\dagger\right) + \left(\frac{b}{M}\right)^2\left(S^\dagger SS + S^\dagger S^\dagger S\right)\right\}H'^\dagger H' + \dots,
\end{aligned} \tag{A.15}$$

so the $SH^\dagger H$ terms have vanished, and the coefficients (to lowest order) are modified as follows

$$\begin{aligned}
\frac{c}{M^2} &\to \frac{c}{M^2}\left(1 - \left(\frac{b}{M}\right)^2\right) \\
\frac{e}{M^2} &\to \frac{e}{M^2}\left(1 - \left(\frac{b}{M}\right)^2\right).
\end{aligned} \tag{A.16}$$

Doing the same field redefinition in the superpotential yields

$$\begin{aligned}
W_{\text{MSSM}} &= \bar{u}y_u QH_u - \bar{d}y_d QH_d - \bar{e}y_e LH_d + \mu H_u H_d \\
&= \bar{u}y_u QH_u\left(1 - \frac{b_u}{M}S\right) + \dots \\
&= \bar{u}y_u QH'_u - \bar{u}\frac{b_u y_u}{M}QH'_u S + \cdots.
\end{aligned} \tag{A.17}$$

Taking the F-term of this and letting S acquire a VEV, yields the trilinear term $\frac{\langle F_S\rangle}{M}\phi_u y_u \phi_{H_u}\phi_Q$.

Appendix B

Special functions

Error function

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx \tag{B.1}$$

Incomplete gamma function

$$\gamma(a,x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt \tag{B.2}$$

Relation between the two

$$\mathrm{erf}(x) = \gamma\left(\frac{1}{2}, x^2\right) \tag{B.3}$$

Gaussian cumulative distribution function relation

$$\Phi(z) = \frac{1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)}{2} \tag{B.4}$$

The significance

$$Z = -2 \ln \frac{\mathcal{L}_{S+B}}{\mathcal{L}_B} \equiv -2 \ln \mathcal{L}, \tag{B.5}$$

asymptotically follows a χ^2 distribution if the null hypothesis is correct (or a non-central one if the alternate hypothesis is), and given the χ^2 cumulative distribution function

$$cdf(x;n) = \gamma\left(\frac{n}{2}, \frac{x}{2}\right) \tag{B.6}$$

for n degrees of freedom (here: 1), an expression for the the p-value is

$$\begin{aligned} p &= 1 - cdf(Z) \\ &= 1 - \gamma\left(\frac{1}{2}, \frac{Z}{2}\right). \end{aligned} \tag{B.7}$$

From this follows

$$1 - p = \gamma\left(\frac{1}{2}, \frac{Z}{2}\right), \tag{B.8}$$

and

$$\begin{aligned} \frac{Z}{2} &= \gamma^{-1}\left(\frac{1}{2}, 1 - p\right) \\ \Rightarrow \log \mathcal{L} &= -\gamma^{-1}\left(\frac{1}{2}, 1 - p\right). \end{aligned} \tag{B.9}$$

Bibliography

[1] J. Heisig, J. Kersten, N. Murphy, and I. Strümke, *Trilinear-augmented gaugino mediation*. Journal of High Energy Physics **2017** (2017) no. 5, 3. https://doi.org/10.1007/JHEP05(2017)003.

[2] A. Kvellestad, S. Maeland, and I. Strümke, *Signal mixture estimation for degenerate heavy Higgses using a deep neural network*. The European Physical Journal C **78** (2018) no. 12, 1010. https://doi.org/10.1140/epjc/s10052-018-6455-z.

[3] A. D. Angelis, ..., I. Strümke, *et al.*, *Science with e-ASTROGAM: A space mission for MeV - GeV gamma-ray astrophysics*. Journal of High Energy Astrophysics (2018) . http://www.sciencedirect.com/science/article/pii/S2214404818300168.

[4] J. Heisig, J. K. Kersten, I. Strümke, and J. Van den Abeele, *In progress: An open window for high reheating temperatures in gravitino dark matter scenarios*.

[5] A. Raklev, I. Strümke, and J. Van den Abeele, *In progress: Gamma rays from the annihilation of dark matter into heavy mesons*.

[6] D. Alvestad, N. Fomin, J. Kersten, and I. Strümke, *In progress: Enabling sneutrino detection in weak signal scenarios using machine learning methods*.

[7] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*. The Astrophysical Journal **707** (2009) no. 2, 916. http://stacks.iop.org/0004-637X/707/i=2/a=916.

[8] W. Porod, *SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e+ e- colliders*. Comput. Phys. Commun. **153** (2003) 275–315, arXiv:hep-ph/0301101 [hep-ph].

[9] B. C. Allanach, *SOFTSUSY: a program for calculating supersymmetric spectra*. Comput. Phys. Commun. **143** (2002) 305–331, arXiv:hep-ph/0104145 [hep-ph].

[10] A. Djouadi, J.-L. Kneur, and G. Moultaka, *SuSpect: A Fortran code for the supersymmetric and Higgs particle spectrum in the MSSM*. Comput. Phys. Commun. **176** (2007) 426–455, arXiv:hep-ph/0211331 [hep-ph].

[11] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*. JHEP **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[12] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*. Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[13] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: going beyond*. Journal of High Energy Physics **2011** (2011) no. 6, . https://doi.org/10.1007/JHEP06(2011)128.

[14] A. Zee, *Quantum Field Theory in a Nutshell*. Nutshell handbook. Princeton Univ. Press, Princeton, NJ, 2003. https://cds.cern.ch/record/706825.

[15] P. Dirac, *The quantum theory of the electron*. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **117** (1928) no. 778, 610–624. http://rspa.royalsocietypublishing.org/content/117/778/610.

[16] H. Weyl, *Electron and Gravitation*. Z. Phys. **56** (1929) 330–352.

[17] W. Pauli, *Dear radioactive ladies and gentlemen*. Phys. Today **31N9** (1978) 27. http://inspirehep.net/record/45177/files/meitner_0393.pdf.

[18] E. Majorana, *Teoria simmetrica dellelettrone e del positrone*. Nuovo Cim. **14** (1937) 171–184.

[19] G. 't Hooft, C. Itzykson, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora, *Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979*. NATO Sci. Ser. B **59** (1980) pp.1–438.

[20] **WMAP** Collaboration, C. L. Bennett *et al.*, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results*. Astrophys. J. Suppl. **148** (2003) 1–27, arXiv:astro-ph/0302207 [astro-ph].

[21] **Planck** Collaboration, Aghanim *et al.*, *Planck 2018 results. VI. Cosmological parameters*. ArXiv e-prints (2018) , arXiv:1807.06209.

[22] **Super-Kamiokande Collaboration** Collaboration, Y. Fukuda *et al.*, *Evidence for Oscillation of Atmospheric Neutrinos*. Phys. Rev. Lett. **81** (1998) 1562–1567. https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[23] J. Schechter and J. W. F. Valle, *Neutrinoless double-β decay in SU(2)×U(1) theories*. Phys. Rev. D **25** (1982) 2951–2954. https://link.aps.org/doi/10.1103/PhysRevD.25.2951.

[24] P. Minkowski, *$\mu \to e\gamma$ at a rate of one out of 109 muon decays?* Physics Letters B **67** (1977) no. 4, 421 – 428. http://www.sciencedirect.com/science/article/pii/037026937790435X.

[25] O. Sawada and A. Sugamoto, eds., *Proceedings: Workshop on the Unified Theories and the Baryon Number in the Universe*, Natl.Lab.High Energy Phys. Natl.Lab.High Energy Phys., Tsukuba, Japan, 1979.

[26] M. e. a. Lévy, ed., *Proceedings of the 1979 Cargése Summer Institute on Quarks and Leptons*. Plenum Press, New York, 1980.

[27] M. Gell-Mann, P. Ramond, and R. Slansky, *Complex Spinors and Unified Theories*. Conf. Proc. **C790927** (1979) 315–321, arXiv:1306.4669 [hep-th].

[28] R. N. Mohapatra and G. Senjanović, *Neutrino Mass and Spontaneous Parity Nonconservation*. Phys. Rev. Lett. **44** (1980) 912–915. https://link.aps.org/doi/10.1103/PhysRevLett.44.912.

[29] *Machines Teach Astronomers About Stars*, https://www.jpl.nasa.gov/news/news.php?feature=4433. Accessed: October 2018.

[30] O. Behnke, K. Krninger, T. Schrner-Sadenius, and G. Schott, eds., *Data analysis in high energy physics*. Wiley-VCH, Weinheim, Germany, 2013. http://www.wiley-vch.de/publish/dt/books/ISBN3-527-41058-9.

[31] K. Pearson, *III. Contributions to the mathematical theory of evolution*. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **185** (1894) 71–110. http://rsta.royalsocietypublishing.org/content/185/71.

[32] B. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1986. https://books.google.no/books?id=e-xsrjsL7WkC.

[33] R. Brun and F. Rademakers, "Root - an object oriented data analysis framework," in *Proceedings AIHENP'96 Workshop, Lausanne, Sep, 1996*. 1996. http://root.cern.ch/.

[34] *NnKde: Gaussian kernel density estimator for classifier outputs*, http://gitlab.com/BSML/sneutrinoML/nn_kde/. Published: January 2019.

[35] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*. Annals Math. Statist. **9** (1938) no. 1, 60–62.

[36] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **231** (1933) 289–337. http://rsta.royalsocietypublishing.org/content/231/694-706/289.

[37] J. Skilling, "Nested sampling," in *American Institute of Physics Conference Series*, vol. 735, pp. 395–405. 2004.

[38] F. Feroz, M. P. Hobson, and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*. Mon. Not. Roy. Astron. Soc. **398** (2009) 1601–1614, arXiv:0809.3437 [astro-ph].

[39] N. CHOPIN and C. P. ROBERT, *Properties of nested sampling*. Biometrika **97** (2010) no. 3, 741–755. http://www.jstor.org/stable/25734120.

[40] P. Mukherjee, D. Parkinson, and A. R. Liddle, *A nested sampling algorithm for cosmological model selection*. Astrophys. J. **638** (2006) L51–L54, arXiv:astro-ph/0508461 [astro-ph].

[41] R. Shaw, M. Bridges, and M. P. Hobson, *Clustered nested sampling: Efficient Bayesian inference for cosmology*. Mon. Not. Roy. Astron. Soc. **378** (2007) 1365–1370, arXiv:astro-ph/0701867 [ASTRO-PH].

[42] P. Berglund, T. Hübsch, and D. Minic, *Relating the cosmological constant and supersymmetry breaking in warped compactifications of IIB string theory*. Phys. Rev. D **67** (2003) 041901. https://link.aps.org/doi/10.1103/PhysRevD.67.041901.

[43] S. Coleman and J. Mandula, *All Possible Symmetries of the S Matrix*. Phys. Rev. **159** (1967) 1251–1256. https://link.aps.org/doi/10.1103/PhysRev.159.1251.

[44] S. P. Martin, *A Supersymmetry primer*. arXiv:hep-ph/9709356 [hep-ph]. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[45] H. Baer and X. Tata, *Weak scale supersymmetry: From superfields to scattering events*. Cambridge University Press, 2006. http://www.cambridge.org/9780521290319.

[46] C. F. Kolda, *Gauge mediated supersymmetry breaking: Introduction, review and update*. Nucl. Phys. Proc. Suppl. **62** (1998) 266–275, arXiv:hep-ph/9707450 [hep-ph].

[47] B. de Carlos and J. A. Casas, *One loop analysis of the electroweak breaking in supersymmetric models and the fine tuning problem*. Phys. Lett. **B309** (1993) 320–328, arXiv:hep-ph/9303291 [hep-ph].

[48] Z. Chacko, M. A. Luty, and E. Ponton, *Massive higher-dimensional gauge fields as messengers of supersymmetry breaking*. JHEP **07** (2000) 036, hep-ph/9909248.

[49] D. E. Kaplan, G. D. Kribs, and M. Schmaltz, *Supersymmetry breaking through transparent extra dimensions*. Phys. Rev. **D62** (2000) 035010, hep-ph/9911293.

[50] Z. Chacko, M. A. Luty, A. E. Nelson, and E. Ponton, *Gaugino mediated supersymmetry breaking*. JHEP **01** (2000) 003, hep-ph/9911323.

[51] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudk, eds., vol. 15, pp. 315–323. 2011.

[52] Z. C. Lipton, *A Critical Review of Recurrent Neural Networks for Sequence Learning*. CoRR **abs/1506.00019** (2015) , arXiv:1506.00019. http://arxiv.org/abs/1506.00019.

[53] F. Dyson, *A meeting with Enrico Fermi*. Nature **427** (2004) 297.

[54] L. Hrmander, *The Analysis of Linear Partial Differential Operators I : Distribution Theory and Fourier Analysis*, 1998.

[55] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*. arXiv:1506.02169 [stat.AP].

[56] J. Friedman, T. Hastie, and R. Tibshirani, *Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)*. Ann. Statist. **28** (2000) no. 2, 337–407.

[57] J. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*. Annals of Statistics **29** (2001) 1189–1232.

[58] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. 2016. http://doi.acm.org/10.1145/2939672.2939785.

[59] *CatBoost by Yandex*, https://tech.yandex.com/catboost. Accessed: 2018-12-01.

[60] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*. Helv. Phys. Acta **6** (1933) 110–127. [Gen. Rel. Grav.41,207(2009)].

[61] F. Zwicky, *Republication of: The redshift of extragalactic nebulae*. General Relativity and Gravitation **41** (2009) no. 1, 207–224. https://doi.org/10.1007/s10714-008-0707-4.

[62] A. Refregier, *Weak gravitational lensing by large scale structure*. Ann. Rev. Astron. Astrophys. **41** (2003) 645–668, arXiv:astro-ph/0307212 [astro-ph].

[63] J. A. Tyson, G. P. Kochanski, and I. P. Dell'Antonio, *Detailed mass map of CL0024+1654 from strong lensing*. Astrophys. J. **498** (1998) L107, arXiv:astro-ph/9801193 [astro-ph].

[64] J. H. Oort, *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*. Bull. Astron. Inst. Netherlands **6** (1932) 249–287. http://cds.cern.ch/record/436532.

[65] T. D. Kinman, *Globular Clusters, III. An Analysis of the Cluster Radial Velocities*. Monthly Notices of the Royal Astronomical Society **119** (1959) no. 5, 559–575. http://dx.doi.org/10.1093/mnras/119.5.559.

[66] H. W. Babcock, *The rotation of the Andromeda Nebula*. Lick Observatory Bulletin **19** (1939) 41–51.

[67] F. D. Kahn and L. Woltjer, *Intergalactic Matter and the Galaxy.*

[68] Y. Sofue and V. Rubin, *Rotation Curves of Spiral Galaxies*. Annual Review of Astronomy and Astrophysics **39** (2001) no. 1, 137–174, https://doi.org/10.1146/annurev.astro.39.1.137. https://doi.org/10.1146/annurev.astro.39.1.137.

[69] V. Trimble, *Existence and Nature of Dark Matter in the Universe*. Annual Review of Astronomy and Astrophysics **25** (1987) no. 1, 425–472, https://doi.org/10.1146/annurev.aa.25.090187.002233. https://doi.org/10.1146/annurev.aa.25.090187.002233.

[70] M. Milgrom, *A Modification of the Newtonian Dynamics - Implications for Galaxy Systems*.

[71] D. Clowe, A. Gonzalez, and M. Markevitch, *Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657558: Direct Evidence for the Existence of Dark Matter*. The Astrophysical Journal **604** (2004) no. 2, 596. http://stacks.iop.org/0004-637X/604/i=2/a=596.

[72] A. Einstein, *The Foundation of the General Theory of Relativity*. Annalen Phys. **49** (1916) no. 7, 769–822. [Annalen Phys.354,no.7,769(1916)].

[73] A. Friedmann, *Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes*. Zeitschrift für Physik **21** (1924) no. 1, 326–332. https://doi.org/10.1007/BF01328280.

[74] G. Lemaître, *Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques*. Annales de la Société Scientifique de Bruxelles **47** (1927) 49–59.

[75] H. P. Robertson, *Kinematics and World-Structure III.* The Astrophysical Journal **83** (1936) 257.

[76] **Particle Data Group** Collaboration, J. Beringer *et al.*, *Review of Particle Physics*. Phys. Rev. D **86** (2012) 010001. https://link.aps.org/doi/10.1103/PhysRevD.86.010001.

[77] L. Verde *et al.*, *The 2dF Galaxy Redshift Survey: The Bias of galaxies and the density of the Universe*. Mon. Not. Roy. Astron. Soc. **335** (2002) 432, arXiv:astro-ph/0112161 [astro-ph].

[78] P. Gondolo, *Non-baryonic dark matter*. NATO Sci. Ser. II **187** (2005) 279–333, arXiv:astro-ph/0403064 [astro-ph]. [,279(2003)].

[79] S. D. M. White, C. S. Frenk, and M. Davis, *Clustering in a neutrino-dominated universe*.

[80] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, *Formation of Galaxies and Large Scale Structure with Cold Dark Matter*. Nature **311** (1984) 517–525. [,96(1984)].

[81] K. N. Abazajian *et al.*, *Light Sterile Neutrinos: A White Paper*. arXiv:1204.5379 [hep-ph].

[82] L. J. Rosenberg and K. A. van Bibber, *Searches for invisible axions*. Phys. Rept. **325** (2000) 1–39.

[83] A. Boyarsky, O. Ruchayskiy, and M. Shaposhnikov, *The Role of sterile neutrinos in cosmology and astrophysics*. Ann. Rev. Nucl. Part. Sci. **59** (2009) 191–214, arXiv:0901.0011 [hep-ph].

[84] **Fermi-LAT** Collaboration, M. Ackermann *et al.*, *Updated search for spectral lines from Galactic dark matter interactions with pass 8 data from the Fermi Large Area Telescope*. Phys. Rev. **D91** (2015) no. 12, 122002, arXiv:1506.00013 [astro-ph.HE].

[85] **Fermi-LAT** Collaboration, A. Albert, G. A. Gomez-Vargas, M. Grefe, C. Munoz, C. Weniger, E. D. Bloom, E. Charles, M. N. Mazziotta, and A. Morselli, *Search for 100 MeV to 10 GeV γ-ray lines in the Fermi-LAT data and implications for gravitino dark matter in $\mu\nu$SSM*. JCAP **1410** (2014) no. 10, 023, arXiv:1406.3430 [astro-ph.HE].

[86] C. Gattringer and C. B. Lang, *Quantum chromodynamics on the lattice*. Lect. Notes Phys. **788** (2010) 1–343.

[87] **Fermi-LAT** Collaboration, A. W. Strong, "Interstellar gamma rays and cosmic rays: new insights from Fermi-LAT AND INTEGRAL," in *Cosmic rays for particle and astroparticle physics. Proceedings, 12th ICATPP Conference, Como, Italy, October 7-8, 2010*, pp. 473–481. 2011. arXiv:1101.1381 [astro-ph.HE].

[88] T. Bringmann, A. Galea, A. Hryczuk, and C. Weniger, *Novel spectral features in MeV gamma rays from dark matter*. Phys. Rev. D **95** (2017) 043002. https://link.aps.org/doi/10.1103/PhysRevD.95.043002.

[89] M. Srednicki, S. Theisen, and J. Silk, *Cosmic Quarkonium: A Probe of Dark Matter*. Phys. Rev. Lett. **56** (1986) 263. [Erratum: Phys. Rev. Lett.56,1883(1986)].

[90] S. Rudaz, *Cosmic Production of Quarkonium?* Phys. Rev. Lett. **56** (1986) 2128.

[91] L. Bergstrom, *Possible Structure in Cosmic gamma-rays From Dark Matter Particle Annihilation*. Nucl. Phys. **B325** (1989) 647–659.

[92] *The Imaging Compton Telescope*, https://heasarc.gsfc.nasa.gov/docs/cgro/comptel/. Accessed: October 2018.

[93] *AMEGO, the All-sky Medium Energy Gamma-ray Observatory*, https://asd.gsfc.nasa.gov/amego/. Accessed: October 2018.

[94] A. H. Guth, *Inflationary universe: A possible solution to the horizon and flatness problems*. Phys. Rev. D **23** (1981) 347–356. https://link.aps.org/doi/10.1103/PhysRevD.23.347.

[95] A. D. Sakharov, *Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe*. Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35. [Usp. Fiz. Nauk161,61(1991)].

[96] W. Buchmuller, P. Di Bari, and M. Plumacher, *Leptogenesis for pedestrians*. Annals Phys. **315** (2005) 305–351, arXiv:hep-ph/0401240 [hep-ph].

[97] W. Buchmuller, P. Di Bari, and M. Plumacher, *The Neutrino mass window for baryogenesis*. Nucl. Phys. **B665** (2003) 445–468, arXiv:hep-ph/0302092 [hep-ph].

[98] S. Davidson and A. Ibarra, *A Lower bound on the right-handed neutrino mass from leptogenesis*. Phys. Lett. **B535** (2002) 25–32, arXiv:hep-ph/0202239 [hep-ph].

[99] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel, *Lyman-alpha constraints on warm and on warm-plus-cold dark matter models*. JCAP **0905** (2009) 012, arXiv:0812.0010 [astro-ph].

[100] M. Bolz, A. Brandenburg, and W. Buchmuller, *Thermal production of gravitinos*. Nucl. Phys. **B606** (2001) 518–544, arXiv:hep-ph/0012052 [hep-ph]. [Erratum: Nucl. Phys.B790,336(2008)].

[101] I. V. Falomkin, G. B. Pontecorvo, M. G. Sapozhnikov, M. Y. Khlopov, F. Balestra, and G. Piragino, *Low-energy $\bar{p}$ ^{4}He Annihilation And Problems Of The Modern Cosmology, GUT And SUSY Models*. Nuovo Cim. **A79** (1984) 193–204. [Yad. Fiz. **39** (1984), 990].

[102] J. R. Ellis, J. E. Kim, and D. V. Nanopoulos, *Cosmological Gravitino Regeneration and Decay*. Phys. Lett. **B145** (1984) 181.

[103] M. Kawasaki, K. Kohri, T. Moroi, and Y. Takaesu, *Revisiting Big-Bang Nucleosynthesis Constraints on Long-Lived Decaying Particles*. Phys. Rev. **D97** (2018) no. 2, 023502, arXiv:1709.01211 [hep-ph].

[104] J. Pradler and F. D. Steffen, *Thermal gravitino production and collider tests of leptogenesis*. Phys. Rev. D **75** (2007) 023509. https://link.aps.org/doi/10.1103/PhysRevD.75.023509.

[105] J. Pradler and F. D. Steffen, *Thermal gravitino production and collider tests of leptogenesis*. Phys. Rev. **D75** (2007) 023509, arXiv:hep-ph/0608344.

[106] M. Bolz, A. Brandenburg, and W. Buchmller, *Thermal production of gravitinos*. Nuclear Physics B **606** (2001) no. 1, 518 – 544. http://www.sciencedirect.com/science/article/pii/S0550321301001328.

[107] J. L. Feng, A. Rajaraman, and F. Takayama, *SuperWIMP dark matter signals from the early universe*. Phys. Rev. **D68** (2003) 063504, arXiv:hep-ph/0306024 [hep-ph].

[108] L. Covi, J. E. Kim, and L. Roszkowski, *Axinos as cold dark matter*. Phys. Rev. Lett. **82** (1999) 4180–4183, arXiv:hep-ph/9905212 [hep-ph].

[109] M. Olechowski, S. Pokorski, K. Turzynski, and J. D. Wells, *Reheating Temperature and Gauge Mediation Models of Supersymmetry Breaking*. JHEP **12** (2009) 026, arXiv:0908.2502 [hep-ph].

[110] J. Heisig, *Gravitino LSP and leptogenesis after the first LHC results*. JCAP **1404** (2014) 023, arXiv:1310.6352 [hep-ph].

[111] M. Kawasaki, K. Kohri, T. Moroi, and A. Yotsuyanagi, *Big-Bang Nucleosynthesis and Gravitino*. Phys. Rev. **D78** (2008) 065011, arXiv:0804.3745 [hep-ph].

[112] L. Covi, M. Olechowski, S. Pokorski, K. Turzynski, and J. D. Wells, *Supersymmetric mass spectra for gravitino dark matter with a high reheating temperature*. JHEP **01** (2011) 033, arXiv:1009.3801 [hep-ph].

[113] **ATLAS** Collaboration, M. Aaboud *et al.*, *Search for squarks and gluinos in final states with jets and missing transverse momentum using 36fb^{-1} of $\sqrt{s}$ = 13TeV pp collision data with the ATLAS detector*. Phys. Rev. **D97** (2018) no. 11, 112001, arXiv:1712.02332 [hep-ex].

[114] **CMS** Collaboration, A. M. Sirunyan *et al.*, *Search for disappearing tracks as a signature of new long-lived particles in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. JHEP **08** (2018) 016, arXiv:1804.07321 [hep-ex].

[115] F. Ambrogi, S. Kraml, S. Kulkarni, U. Laa, A. Lessa, V. Magerl, J. Sonneveld, M. Traub, and W. Waltenberger, *SModelS v1.1 user manual: Improving simplified model constraints with efficiency maps*. Comput. Phys. Commun. **227** (2018) 72–98, arXiv:1701.06586 [hep-ph].

[116] **ATLAS** Collaboration, H. Okawa, "Interpretations of SUSY Searches in ATLAS with Simplified Models," in *Particles and fields. Proceedings, Meeting of the Division of the American Physical Society, DPF 2011, Providence, USA, August 9-13, 2011*. 2011. arXiv:1110.0282 [hep-ex].

[117] **CMS** Collaboration, S. Chatrchyan *et al.*, *Interpretation of Searches for Supersymmetry with simplified Models*. Phys. Rev. **D88** (2013) no. 5, 052017, arXiv:1301.2175 [hep-ex].

[118] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall, and J. S. Kim, *CheckMATE: Confronting your Favourite New Physics Model with LHC Data*. Comput. Phys. Commun. **187** (2015) 227–265, arXiv:1312.2591 [hep-ph].

[119] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*. JHEP **02** (2014) 057, arXiv:1307.6346 [hep-ex].

[120] W. Buchmüller, J. Kersten, and K. Schmidt-Hoberg, *Squarks and sleptons between branes and bulk*. JHEP **02** (2006) 069, hep-ph/0512152.

[121] F. Brümmer, S. Kraml, and S. Kulkarni, *Anatomy of maximal stop mixing in the MSSM*. JHEP **08** (2012) 089, arXiv:1204.5977 [hep-ph].

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 978-82-308-3511-1