Doubly Charged Higgs Signal with Same-Sign τ-inclusive Dilepton Final States

Simon Arnling Bååth

Master's thesis

Supervised by Else Lytken and co-supervised by Oxana Smirnova

Division of Particle Physics at the Department of Physics

Lund University

April, 2018

Abstract

The current Standard Model of particle physics contains few interactions resulting in a final state with two leptons with the same charge. Any observation of an excess of such same-sign dileptons would be an indication of new physics processes, making it a useful channel when searching for exotic physics. This thesis presents three projects conducted within the ATLAS Same-Sign Dilepton group for extending their previous same-sign dilepton analysis to include τ-leptons. Including the heaviest lepton in the analysis gives access to more statistics from the available data and allows probing of theories with mass-dependent lepton coupling to, for example, the doubly charged Higgs particle. To this end, Monte Carlo signal samples with pair production of doubly charged Higgs decaying into two same-sign dileptons was requested, and the resulting signal samples and the method for generating them will be presented. A filter able to generate a specific τ-lepton final state in event generation was created, and validation of its function is shown. Finally, a small proof-of-concept study of τ-lepton charge flip estimation using a data-driven likelihood method will be presented.

In memory of Caspar Håkansson

Contents

1 Introduction **1**

1.1 Purpose and motivation . . . 1

1.1.1 τ-inclusive $H^{\pm\pm}$ signal sample generation . . . 1

1.1.2 `TauFilter` event generation filter update . . . 1

1.1.3 τ-lepton charge flip rate estimation . . . 2

1.2 Thesis structure . . . 2

1.3 Author's contribution . . . 3

2 The ATLAS experiment **4**

2.1 The ATLAS detector . . . 4

2.2 ATLAS data flow . . . 5

3 Theoretical background **8**

3.1 The Standard Model of particle physics . . . 8

3.2 The τ-lepton . . . 10

3.2.1 Historical background . . . 10

3.2.2 Decay modes and standardized notation . . . 10

3.2.3 Reconstruction and identification . . . 12

3.3 Same-sign dilepton signal . . . 13

3.3.1 Same-sign dilepton Standard Model background . . . 13

4 τ-inclusive $H^{\pm\pm}$ signal sample generation **16**

4.1 Introduction . . . 16

4.2 Theory . . . 16

4.2.1 The Left-Right Symmetric Model . . . 16

4.2.2 Doubly Charged Higgs in the LRSM . . . 17

4.2.3 $H^{\pm\pm}$ signal . . . 18

4.2.4 τ-leptons in $H^{\pm\pm}$ decay . . . 19

4.3 Method . . . 19

4.3.1 Pythia . . . 20

4.3.2 ATLAS detector simulation software . . . 20

4.4 Data sets . . . 21

4.4.1 Event generation samples . . . 21
4.4.2 Signal samples . . . 22
4.4.3 Filter and final state statistics . . . 23
4.5 Result . . . 24
4.5.1 Event generation validation histograms . . . 24
4.5.2 Signal sample histograms . . . 30
4.6 Discussion . . . 36

5 Update of event generation filter `TauFilter` 39
5.1 Purpose and motivation . . . 39
5.2 Method . . . 39
5.3 Data sets . . . 40
5.4 Result . . . 41
5.5 Discussion . . . 46

6 Summary and outlook 48

Acknowledgments 49

A Definition of particle physics terms 50

B Tools and software used by ATLAS 53

C The collinear approximation for τ-lepton mass reconstruction 54

D Signal sample job options file 56

E Combinatorics of the τ-inclusive signal sample 57

F τ-lepton charge flip probability 59
F.1 Introduction . . . 59
F.2 Theory . . . 59
F.2.1 Methods for estimation of charge flip rates . . . 60
F.3 Likelihood charge flip estimation method . . . 61
F.4 Method . . . 63
F.4.1 Data set . . . 63

F.4.2 Background removal 64
F.4.3 Log-likelihood maximization 65
F.5 Result 66
F.6 Discussion 67

References **69**

Acronyms **74**

1 Introduction

1.1 Purpose and motivation

The ATLAS experiment is an international scientific collaboration based around the ATLAS particle detector at CERN, with one of the aims being the search for physics beyond the current Standard Model of particle physics. The collaboration is divided up among many institutions and groups, all of which get access to the data produced by the detector. This thesis presents work done within the ATLAS Same-Sign Dilepton (SSDiLep) group with the aim of including τ-leptons in a future analysis with a final state consisting of one or more leptons pairs with the same charge, so-called same-sign dileptons. In the current Standard Model of particle physics interactions that give rise to same-sign dileptons are rare. However, there exist several models extending the Standard Model where same-sign dileptons are expected to be more common. Observation of an excess of same-sign dileptons would thus be a powerful indication of new physics. The physics model considered in this thesis to give rise to a same-sign dilepton signal is the existence of a doubly charged Higgs boson ($H^{\pm\pm}$ or DCH).

The thesis has mainly been split into three separate projects: Generation of signal samples with pair production of $H^{\pm\pm}$, the update of the event generation filter `TauFilter` for use in future signal sample generation, and a study of a data-driven charge-flip rate estimation method for τ-leptons. Each which will be given a brief introduction below. This thesis also aims to be a general introduction to ATLAS data flow and particle physics terminology so as to be useful as a reference work for new members of the collaboration.

1.1.1 τ-inclusive $H^{\pm\pm}$ signal sample generation

Simulating how a new type of particle would be detected in a particle detector, referred to as signal sample generation, is an important part of any particle physics analysis. Signal samples give information on how a new physics model would appear in the detector data, and are used in comparison with the observed data to be able to exclude physics models for the accessible energy scale of a particle detector. Due to the computational resources required to simulate particle collisions all signal samples are generated centrally by request from the different ATLAS subgroups. The request process involves deciding which process within which physics model that should be simulated and then validating that the correct particle interactions have indeed been simulated before requesting a detector simulation.

As part of this thesis work a signal sample request was made for pair production of $H^{\pm\pm}$ decaying into same-sign dileptons. The purpose for this was to extend the previous signal samples of the group which included the same process but without any τ-leptons, thus providing viable signal samples for a future τ-inclusive same-sign dilepton analysis.

1.1.2 `TauFilter` event generation filter update

To better control the simulation of particle collision events, so-called event generation, filters can be used that selectively chooses only certain of the generated events and rejects the others. This

is useful when signal samples for a specific event type are needed, as use of a filter can prevent generation of unnecessary events.

This project entailed extending and validation the previously existing event generation filter `TauFilter` to have more flexible functionalities. The purpose was to enable more specific τ-lepton final state for future signal samples, as the current filter lacked the ability to filter in regards to both the number of τ-leptons and whether they are hadronic or leptonic. Updating the ATLAS codebase also enables other teams and analyses to benefit from work done within one of the ATLAS sub-groups.

1.1.3 τ-lepton charge flip rate estimation

The charge flip rate is a measure for how often a particle detector assigns an incorrect charge to a detected particle, and an example would be for the detector to mistake an electron for a positron. For a study of same-signed dilepton signal charge flip is an important background to model correctly. As two leptons with the same charge are produced rarely in the Standard Model it must be estimated how often a fake same-signed pair is observed due to charge misidentification. Previously electron charge flip has been the most important factor to calculate, as muon charge is a well-reconstructed variable in the ATLAS detector compared and τ-lepton charge flip has been assumed to be negligible compared to other background. However, when including τ-leptons in the same-sign dilepton analysis there is value in a more rigorous examination of τ charge flip and its kinematic dependencies.

Presented in this thesis is a small overview of charge flip estimation methods, and an example of how a data-driven likelihood method used to estimate electron charge flip probability could be extended to τ-leptons.

1.2 Thesis structure

Due to the often small overlap between the subprojects, they will be presented in separate sections with the aim of keeping the text cohesive and enabling easier navigation of the thesis material. The generation of signal samples will be presented in section 4. Due to the work done being primarily administrative in nature the part will be written as a method report, with focus on describing the different parts of the process rather than on the theoretical aspects of event generation or analysis of the resulting samples. The event filter update will be presented in 5. While sharing much of the method with the signal sample generation, the purpose and result of the subprojects differ enough to warrant a separate section. Due to its status as a proof-of-concept study, the τ-lepton charge flip rate study will be presented in section F so as to establish it as a side project and not part of the main results of the thesis.

A general introduction to the ATLAS detector and the theoretical background of same-sign dilepton signals and τ-leptons, useful as context for all parts, will be given in section 2 and 3. A summary and outlook for the future of the same-sign dilepton analysis will be presented in section 6. Lists of definitions, supporting pieces of code and mathematical derivations will be part of the appendix to be easily referenced if the need arises without taking up space in the main thesis.

Specific terms that are defined in the scope of this thesis are written in *italics*. Software names are written in Small Caps and are referenced in appendix B. Names referring to a specific piece of code are written in `Courier`.

Many parts of this thesis reference code or information available only for members of the ATLAS project, most notably links to the internal ATLAS twiki. While these links thus does not serve as strict references it is the hope that members of the ATLAS project will find their addition useful.

1.3 Author's contribution

This thesis work has been conducted under the ATLAS SSDiLep group. This thesis work aims to enable an extension of the groups' previous publication [18], a general search for same-sign dilepton signals using only light leptons and $H^{\pm\pm}$ as benchmark model, with the inclusion of τ-leptons. All data sets used in this thesis has been produced by and for the group.

Little code used in this thesis has been developed from scratch, with most code being based on existing code used by the SSDiLep group. The aim has been to not only develop the code further for use in this thesis but also for general use of the group for future analyses. Code that has been developed in the context of this thesis include code for event sample validation histogram creation, for data ntuple histogram and plot creation, and for data-driven charge flip estimation. My constribution has primarily been based on work done by Miha Muškinja and Katja Mankinen.

Decisions regarding the approach of several of the problems encounter have been a process often including parts of the group. As such I contributed to the discussion and with proposing possible solutions, with the final decision being taken by the group. This led to some projects being abandoned, and some work done has as such been omitted from this thesis or added to the appendix.

2 The ATLAS experiment

2.1 The ATLAS detector

The following summary of the ATLAS detector is based on the similar treatment of the topic in Ref. [11]. A reference list of common particle physics terms and definitions as used in this thesis is presented in appendix A.

The ATLAS detector is a general-purpose particle physics detector located at the LHC at CERN. A simplified overview of the detector design is presented in figure 1. It is designed to identify particles arising from proton-proton collisions at high center-of-mass energies with the general aim of finding new particles and expanding the current Standard Model. The ATLAS detector is currently used for, among other things, precisions measurements of the Standard Model parameters, search for dark matter candidates, and the testing of several different kinds of beyond the Standard Model physics models.

ATLAS consists of four different detector systems in a shell-like structure. When working in tandem they can measure the kinetic parameters of and identify the final state particles of the proton-proton collisions. Innermost is a tracking detector, consisting of both tracking chambers and semiconductor pixel detectors. It enables the path of the particles created at the collision point to be observed, and by applying a large magnetic field over the tracking system the curvature of the track can also be measured. From that information the charge and transverse momentum of the particles can be inferred. The particles then encounter the electromagnetic calorimeter, whose primary function is to absorb and measure the energy of particles interacting electromagnetically, such as electrons and photons. Energetic jets and hadrons able to pass through the EM calorimeters without being absorbed instead deposit their energy in the hadronic calorimeters, designed to fully measure the energy of jets and also able to absorb neutral hadrons such as neutrons or π^0. The design of the hadronic calorimeter ensures that few particles are able to escape without depositing all of their energy within the system. The calorimeter systems primarily consist of combinations of Liquid Argon calorimeters and iron tile calorimeters, and used together with the tracking detector the kinetic parameters of each detected particle can be measured. An outer detector layer is also present to detect any muons present in an event, as muons generally are able to traverse the other parts of the detector without being absorbed.

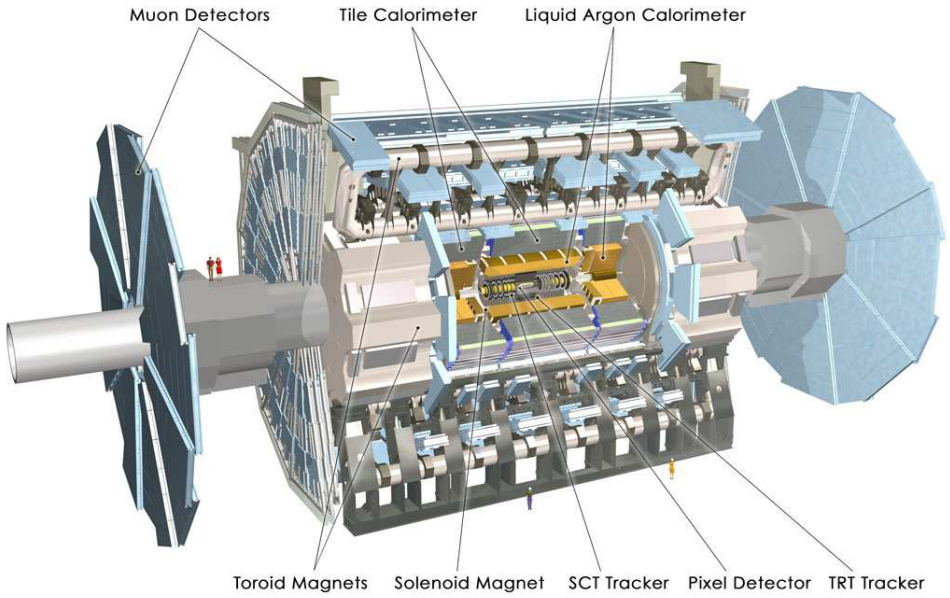


Figure 1: Simplified schematics of the ATLAS detector [1].

2.2 ATLAS data flow

An overview of ATLAS data flow from raw detector output to data files that can be used for a physics analysis is found in Ref. [22]. Due to its relevance to this thesis it will be summarized here. A reference list of the software and tools used in ATLAS and during this thesis work to handle data transfer and analysis is presented in appendix B.

The ATLAS trigger system

When operational the number of proton collisions that can be detected by the ATLAS detector exceeds its ability to record and store the resulting data: There are up to 40 million bunch crossings per second, and only about one thousand events can be recorded per second. One of the principal challenges of particle physics is thus to make sure that as many as possible of the relevant events are recorded while discarding those of little interest. This is handled by the ATLAS trigger system, which uses both hardware- and software-based components to rapidly identify interesting collision events to trigger that those events are recorded permanently.

For a first selection the ATLAS detector uses a hardware base trigger system, called the *Level 1* trigger. Each event signature will result in a different electronic readout from the detector components, and interesting events can therefore be identified in real time based on certain specified criteria, such as a signal from one of the calorimeters. As speed is essential not all of the detector data is used for the final decision of keeping or discarding an event, and the output of the Level 1 trigger is only regions of interest in the detector. If an event passes the Level 1 trigger it is passed on to the software based *High Level Trigger* (HLT), which uses more thorough software based algorithms. The HLT takes into account the full detector data and from that information

tries to reconstruct the actual particles traversing the detector systems to make the decision if the event passes the specified triggers or not.

Only if passing both trigger systems will the event be recorded permanently, and the specific trigger specifications used are tuned to allow the maximum number of events to be recorded. In general a large number of different triggers for interesting physics signatures is used together in a so-called trigger menu, making sure that any recorded event has at least one interesting signature. One problem that must always be taken into account in a particle physics analysis is that no trigger is fully efficient in catching a particular type of event. Trigger efficiency can be estimated by looking at how well the triggers perform for a simulated dataset, and such information must be taken into account when estimating the effective luminosity of a data set.

Particle reconstruction

The ATLAS detectors primary function is to use the gathered detector data, which includes hits in the tracking detector or calorimeter readouts, to correctly identify particles and reconstruct their kinematic properties. By using the detector output the particle that caused a particular readout can be identified and its properties such as charge, momentum and flight path can be so-called *reconstructed*. It is important to note that all particle identification is uncertain in that it is based on inference of the available data, and that it is possible that a certain detector signature can be attributed with similar probability to different kinds of particles. As such, one of the challenges of particle physics is to correctly reconstruct as many of the particles of an event as possible while ensuring the accuracy of the reconstruction. For some particles this process is comparatively straightforward: In the case of the muon the ATLAS detector has dedicated muon detectors encircling the calorimeters. If a track is seen in both the inner tracking detector, the calorimeters and the outer muons detectors the track can with high probability be identified as belonging to a true muon traversing the detector. The situation for τ-leptons is however more complex and will be described in more detail in section 3.2.3.

ATLAS data formats

An *Analysis Data Object* (xAOD[1]) is the data format as constructed from the raw detector data, and contains all the reconstructed particles and their associated variables. As an xAOD contains all reconstructed particles with full information it is a comparatively dense and large data format. Most ATLAS data is available in the form of *Derived Analysis Data Objects* (DxAOD), which is a lighter data format created for a specific purpose or analysis. Depending on derivation a DxAOD has gone through three processes reducing the amount of data, thus making it easier to handle than an xAOD. *Skimming* removes whole events from the xAOD with some criteria, for example leaving only events with at least one τ-lepton. *Slimming* removes object within events, such as removing all electron objects if they are not of interest for the analysis. *Thinning* finally removes object variables that are not needed. The resulting DxAOD is thus significantly lighter and easier to handle than the full xAOD. A simplified flowchart of the ATLAS data flow is presented in figure 2.

[1]For the LHC run 1 this was called an AOD, with the *x* being added to differentiate the new Data Analysis Objects created for run 2. As only run 2 has been relevant for this thesis xAOD will be used for consistency.

While DxAODs in principle can be used directly in a physics analysis the data set is often reduced further into a more basic *ntuple* format. A ntuple is defined as an ordered list with n elements. In the context of particle physics the ntuple is used as a term for the most common data structure used for final analysis, where each element of the list represents an event. Each element contains information concerning the number of particles, their kinematics and other event variables. A common way to use the data ntuple is to define a loop algorithm that steps through each entry of the ntuple, performs some operation such as calculating the invariant mass of a certain set of particles if present, and then outputs the result of the operations as histograms. It is often these histograms that then form the basis of the actual physics analysis.

Simulated data sets used for comparison with the actual detector data is important for any particle physics analysis. They go through several steps before ending up as DxAODs: A particle interaction event is generated, its interaction with the detector is simulated and that interaction is then converted to actual digital electric signals similar to the actual output of the detector before being reconstructed to particles in the same way as detector data. Formats containing simulated data contain so-called *truth* level information, which is a record of the particles that were generated and part of the simulated events. These are expected to be different from the particles that are reconstructed by the detector simulation, and by keeping the truth information in the final DxAOD files comparison can be made on for example how efficiently the reconstruction algorithms work.

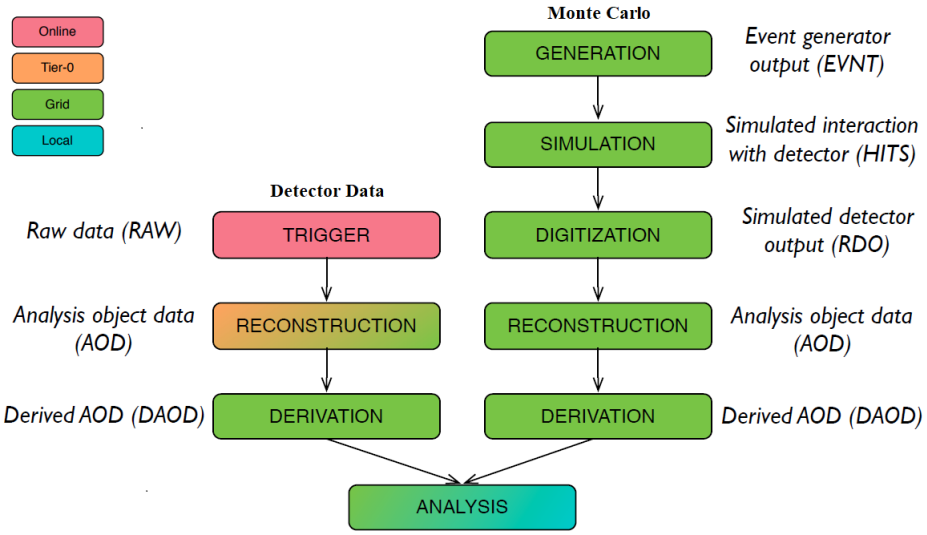


Figure 2: Simplified diagram of the ATLAS data flow for both detector data (left) and Monte Carlo generated data (right) [22].

3 Theoretical background

3.1 The Standard Model of particle physics

The following summary of the Standard Model is based on the similar treatment of the topic in Ref. [11].

The Standard Model of particle physics is the current framework for describing the fundamental particles and how they interact with each other. It emerged during the latter half of the twentieth century accompanied by experimental success, with one of its last major predictions, the existence of the Higgs boson, being confirmed at the LHC in 2012. The Standard Model describes a number of elementary particles (Fig. 3), divided into two main families of leptons and quarks. These families are further divided into three generations, where the particles of each generation behave similarly but have different mass. The standard model explains how these particles interact via the weak, the strong and the electromagnetic force. These three forces are in turn represented by the gauge bosons which mediate all interactions.

Each of the Standard Model particles is associated with a set of quantum numbers which are generally conserved in all interactions. Antiparticles are defined by having the same magnitude but opposite sign of their charges (such as electromagnetic or color charge) than their particle counterparts. Quantum numbers include the spin and charge of the particles, but each fermion is also assigned a specific flavor quantum number denoting its identity. The leptons are assigned a positive electron, muon or tauon number depending on the generation. This corresponds to the observed fact that an electron is always created together with either a positron or an anti-neutrino to conserve the electron lepton number. The quarks behave similarly with the higher generations being assigned quantum numbers like strangeness and charm, while the first generation uses the similar system of either positive or negative isospin.

The strong force is mediated by the massless gluon, affects the six quarks, and is described in terms of color charges. A quark can be either green, blue or red (or the corresponding anti-colors), and the colors attract each other to form neutral, or *colorless*, objects. This together with the fact that the gluon itself has a color gives rise to the concept of color confinement, stating that quarks can only exist if they are bound with other quarks so they together are colorless. Quarks are always found in either triplets with one of each color or in pairs with a color and its corresponding anti-color. For example, triplets of quarks make up the nucleons of the atomic nucleus: The proton consists of two up quarks and one down quark, while the neutron consists of one up quark and two down quarks. The extreme strength of this confinement gives the strong interaction its name, but the self-interaction of the gluon makes it only relevant on scales comparable to the small radius of a nucleon.

While the strong force gives rise to the atomic nucleus, it is the electromagnetic force that is responsible for the atom and most macroscopic phenomena such as chemistry and light. Electromagnetism affects both quarks and the charged leptons, which are the common electron and the heavier muon and tauon. The electromagnetic force couples to electric charge of which the leptons have $-\mathrm{e}$ and the quarks have either $-\frac{1}{3}$ e or $\frac{2}{3}$ e, where e is the unit charge. As quarks never appear alone the electric charge always appears in multiples of the unit charge. Opposite electric charges attract each other which explains the basic feature of the neutral atom: The nucleus contains positive protons that attract a cloud of negative electrons. The electromagnetic

force is mediated by the massless photon, which in contrast to the gluon has no charge of its own implying that it travels freely until it encounters charged matter.

The weak force is, in contrast to the above mentioned forces, more subtle. It affects all the fermions of the Standard Model including the otherwise chargeless neutrinos. It is mediated by the $W^{\pm}$ and Z^0 bosons, which in contrast to the other mediator particles have large masses. The result of this is a very small probability for two particles to interact weakly. In the case of the neutrino, the only elementary particle which exclusively interacts weakly, billions pass through us at any given moment without leaving any trace. The weak force is important for explaining radioactivity and is the only force capable of changing the type of elementary particles, turning a electron into a neutrino or a up quark into a down quark.

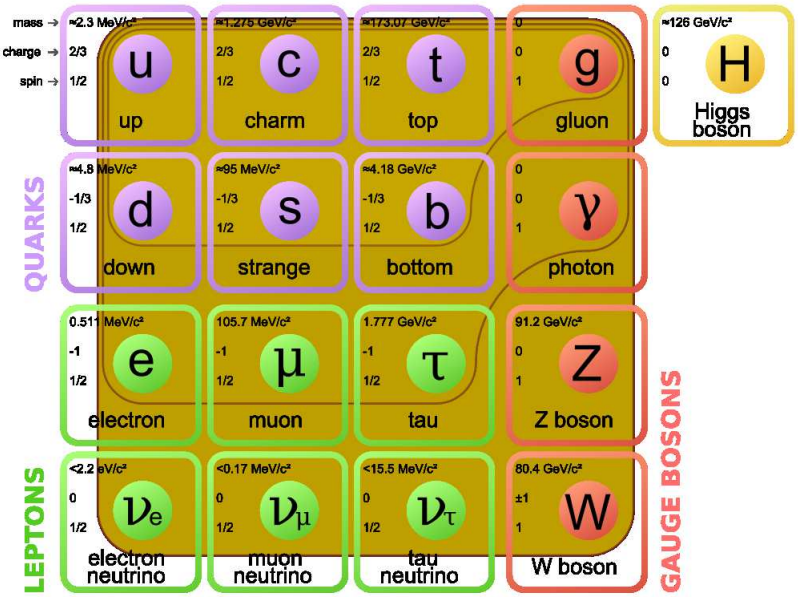


Figure 3: The elementary particles of the Standard Model [2].

While the Standard Model is able to account for almost all observable phenomena it still leaves several problems unsolved. A common criticism is that it is inelegant: While it obviously works well it makes no attempt to explain why precisely these elementary particles and forces exist. Its large number of such so-called free parameters is often a reason to try to replace the Standard Model with something more elegant. Concrete theoretical issues include the problem of gravity: At low energies and small distances where gravity is negligible the Standard Model is fine, but attempts to unify all the forces into one theory has failed. The problem can be exemplified with black holes as they combine an extreme mass with quantum behavior, something which currently cannot be modeled by either general relativity nor the Standard Model separately. Other problems include the mysterious existence of dark matter and dark energy and the problem of the mass and particle nature of the neutrino.

The latter problem is of relevance of this thesis as many models proposing answers to the neutrino question also predict lepton number violation and the existence of doubly charged Higgs bosons. Searching for same-sign dilepton signal not predicted by the Standard Model would thus be a direct test of such models.

3.2 The τ-lepton

The particle data presented in the following section is taken from the Particle Data Group (PDG) [39]. For ease of readability the uncertainty for any cited value has not been included.

The τ-lepton is a charged particle with the basic properties as presented in table 1. Being the third generation of the charged lepton family it behaves similarly to the electron and the muon with the exception of its large mass, which implies shorter lifetime and more possible decay modes.

Table 1: Basic properties of τ^-.

Property	Value
Mass	$M_\tau = 1776.86$ MeV
Charge	$Q_\tau = -e$
Mean lifetime	$\tau_\tau = 290.3 \cdot 10^{-15}$ s

3.2.1 Historical background

The τ-lepton was the last of the charged leptons to be discovered, with the first indication of its existence found at the SLAC e^+e^- collider in 1975. An unexpected number of events of the type

$$e^+ + e^- \rightarrow e^\pm + \mu^\mp + \geq 2 \text{ unobservable particles}$$

was detected, with no conventional explanation to be found. The lack of any jets in the final state put heavy constraints on this signature arising from any hadronic intermediate state. However, pair production of new heavy lepton decaying weakly into the lighter leptons would give rise to exactly this type of final state being observed in events of the type

$$e^+ + e^- \rightarrow \tau^\pm + \tau^\mp \rightarrow e^\pm + \nu_e + \mu^\mp + \nu_\mu + 2\nu_\tau.$$

The latter scenario was quickly found to be the most plausible, and the new particle was eventually given the name τ from the Greek word $\tau\rho\iota\tau o\nu$ meaning *third*, due to it being the third lepton discovered [41, 40].

3.2.2 Decay modes and standardized notation

The τ-lepton is the heaviest particle of its family and will in time, similarly to the muon, decay weakly. However, while the mass of the μ only allows decays into electrons no such constraint exists for the τ-leptons: Its mass allows for a large number of both leptonic and hadronic decay modes. The branching ratios for the most common decay modes are presented in table 2. Hadronic decay is classified depending on the number of charged particles in the final state. Hadronic modes with one charged particle in the final state, and thus one observed track, are called *1-pronged*, while those with three charged particles in the final states are called *3-pronged*.

While there exist possible decay modes with five or more charged hadrons in the final state their branching ration is exceedingly small. The leading order diagram for τ-decay is presented in figure 4.

Table 2: Common decay modes and their branching ratio.

Decay mode	Branching ratio [%]
All leptonic modes	35.2
$\tau \rightarrow \nu_\tau \mu \nu_\mu$	17.4
$\tau \rightarrow \nu_\tau e \nu_e$	17.8
All hadronic modes	64.8
All 1-pronged modes	50.0
$\tau \rightarrow \nu_\tau \pi^\pm$	10.8
$\tau \rightarrow \nu_\tau \pi^\pm \pi^0$	25.5
$\tau \rightarrow \nu_\tau \pi^\pm 2\pi^0$	9.26
All 3-pronged modes	14.6
$\tau \rightarrow \nu_\tau \pi^\pm \pi^\pm \pi^\mp$	8.99
$\tau \rightarrow \nu_\tau \pi^\pm \pi^\pm \pi^\mp \pi^0$	2.79

The standard shorthand notation for τ-lepton decays will be used throughout this thesis as laid out by the ATLAS Tau Working Group [19]. This notation is summarized in table 3.

Table 3: Standardized notation for τ-leptons.

Notation	Meaning
τ	Undecayed τ-lepton
τ_e	$\tau \rightarrow \nu_\tau e \nu_e$
τ_μ	$\tau \rightarrow \nu_\tau \mu \nu_\mu$
τ_{lep}	τ_e or τ_μ
τ_{had}	$\tau \rightarrow \nu_\tau q \bar{q} \rightarrow \nu_\tau +$ hadrons
$\tau_{\text{had-vis}}$	τ_{had} without the ν_τ
$\tau_{1-\text{pronged}}$	τ_{had}, 1 charged hadron
$\tau_{3-\text{pronged}}$	τ_{had}, 3 charged hadrons

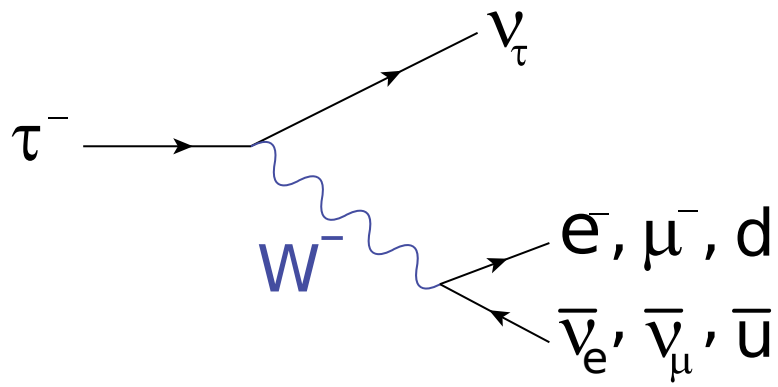


Figure 4: Leading order Feynman diagram of τ-lepton decay [8].

3.2.3 Reconstruction and identification

The full technical report for ATLAS run 2 τ-lepton identification and reconstruction is given by Ref. [15].

The large mass of the τ-lepton gives it a short lifetime and many possible decay modes, and both of these properties makes it more difficult to identify compared to the lighter leptons. A τ-lepton track cannot be observed directly as it decays close to the primary $p-p$ collision vertex before entering the detector tracking system. Instead τ-leptons can only be detected by inference by observation of the τ decay products. However, in the case of the τ-lepton the decay products are either indistinguishable from prompt light lepton or resemble hadronic jets, and the ATLAS detector is currently unable to distinguish a prompt light lepton and a light lepton arising from a τ-lepton decay. The challenge of identifying and using τ-leptons for analysis is then of correctly separating jets from a τ-lepton decay from the background jets, which is an extremely common background in hadron colliders.

$\tau_{had-vis}$ identification consist of two main steps. First, $\tau_{had-vis}$ candidates are found using jets formed with the anti-k_t algorithm with certain constraints as seeds to a $\tau_{had-vis}$ reconstruction algorithm. The production vertex (not necessarily the same as the primary vertex of an event) and the association of tracks to the jet candidate is similarly associated with a candidate jet.

However, the candidate selection does not reliably separate hadronic τ-leptons from other kinds of jets, and to decide whether a reconstructed $\tau_{had-vis}$ should be identified as a true τ-lepton or not a Boosted Decision Tree (BDT) is used. A BDT is an algorithm that takes a number of kinematic variables from a jet τ-candidate as input and produces a score as output, giving a measure on how τ-like the candidate behaved. Hadronic τ decay has several characteristics that can be used to distinguish it from other jets: For example, a jet from a τ-lepton decay always have either one or three tracks in the tracking detector, and the τ jet is often relatively collimated. A BDT is *trained* by using simulation data as input where the true particles are known. By then comparing the output BDT score to the truth information the BDT can be optimized to give the true $\tau_{had-vis}$ a higher score. After having been trained on simulated data the reconstructed $\tau_{had-vis}$ candidates can thus be run through the same algorithm, making it possible to reject

the low-scoring jet candidates a background and identifying the high-scoring candidates as true τ-leptons.

The efficiency for τ-lepton identification is dependent on both prongedness and p_t, and different working points with different identification efficiency are defined to simplify analysis. As reference, the tight working point for 1-pronged τ_{had} is 45% and for 3-pronged τ_{had} 30%, showing that for the more stringent chosen working point the identification efficiency for τ_{had} is currently far from perfect at the ATLAS detector.

3.3 Same-sign dilepton signal

A more extensive introduction to same-sign dilepton signal can be found in Ref. [32].

Processes giving rise to opposite sign dileptons are present in the Standard Model giving final states with two prompt leptons with opposite charge, with pair production from a Z-boson and γ being the most notable example. This can commonly occur at the LHC in a quark anti-quark annihilation process

$$q\bar{q} \xrightarrow{Z/\gamma^*} l^{\pm}l^{\mp}.$$

Processes giving rise to same-sign dileptons are however rare: As both charge and lepton number are conserved quantities in the Standard model there is no direct process giving rise to an isolated prompt same-sign dilepton. There are however many possible Standard Model extensions that postulate same-sign dilepton signal with different proposed mechanisms for how lepton number violation can be introduced into the standard model. This includes exotic models such as microscopic black holes or the more commonly encountered Left-Right Symmetric Model. The latter includes both doubly charged resonances that decay into same-sign leptons and lepton number violation due to neutrinos being their own anti-particle and thus not conserving any specific lepton number. Observation of a same-sign dilepton signal is thus a direct indication of new physics regardless of chosen model, making it a useful channel for an inclusive search for any new physics.

One of the principal advantages of the channel is the low expected Standard Model background, especially at center-of-mass energies far away from the weak boson resonances. This would make a potential signal easier to differentiate from the noise, thus making signals of interest visible at lower center-of-mass energy than other new physics signals with a more notable standard model background. The same-sign dilepton channel also uses the fact that reconstruction and identification of the light leptons in the ATLAS detector is accurate and less prone to be confused with the hadronic background present in any $p-p$ collider.

3.3.1 Same-sign dilepton Standard Model background

A commonly discussed experimental signature in the context of doubly charged scalar resonances, such as the $H^{\pm\pm}$, is the 4-lepton final state. The scalar is pair produced by some process and each of the produced particles decays into a same-signed dilepton, resulting in two dileptons

and conserving overall charge of the event. Being the principal channel discussed in the context of this thesis the Standard model background for the 4-lepton final state will be briefly presented.

Standard Model processes that give rise to final states that cannot be distinguished from the signal channel is called *irreducible* or *prompt background.* The primary prompt background process for a 4-lepton final state with two same-sign dileptons is the diboson process

$$q\bar{q} \rightarrow 2(Z/\gamma^*) \rightarrow l^{\pm}l^{\mp}l^{\pm}l^{\mp}.$$

It is in many cases indistinguishable from signal same-sign lepton final states as it shares an identical set of prompt particles in the final state. The background can be identified based on the enhanced cross-section for opposite-sign dilepton invariant mass to be found within the Z resonance mass peak. This process will also never give rise to any lepton number violating final state.

A selection of leading order processes that give rise to same-sign dilepton signal is presented in figure 5 giving a possible identical final state of four charged leptons that would be background to a 4-lepton channel.

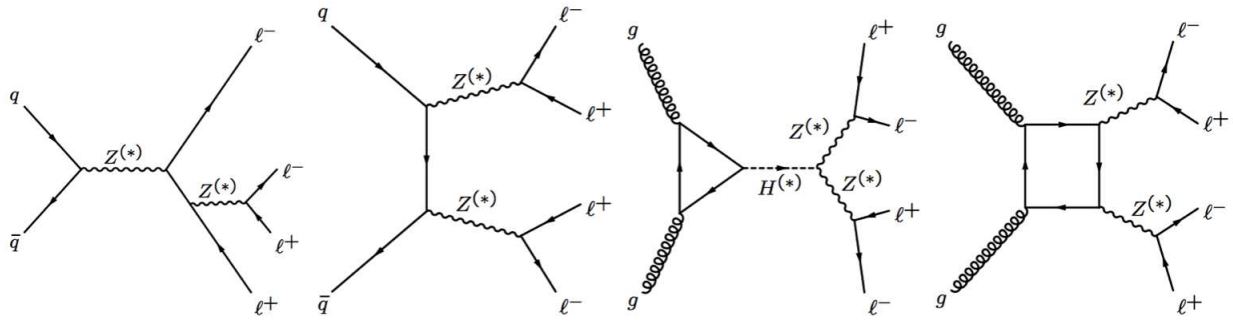


Figure 5: Leading order Feynman diagrams for a selection of $ZZ \rightarrow l^{\pm}l^{\mp}l^{\pm}l^{\mp}$ background processes [18].

Other diboson processes can give rise to single dilepton final states, such as

$$q\bar{q} \rightarrow ZW^{\pm} \rightarrow l^{\pm}l^{\mp}l^{\pm}\nu_l.$$

Such processes can be identified by the increase of p_T^{miss} in the final state due to the neutrino, and that one of the opposite sign lepton pairings should have an invariant mass in the Z resonance mass window. Several prompt same-sign backgrounds are also associated with jets, such as the diboson

$$q\bar{q} \rightarrow q\bar{q} + W^{\pm}W^{\pm} \rightarrow q\bar{q} + l^{\pm}\nu_l l^{\pm}\nu_l,$$

but can be identified due to presence of jets and p_T^{miss} from the neutrinos. To note is that the diboson backgrounds are higher order weak interactions and thus have a small intrinsic production cross-section. Prompt dilepton background can also come from $t\bar{t}$ processes with associated weak boson radiation such as

$$q\bar{q} \rightarrow t\bar{t}Z \rightarrow W^{\pm}W^{\mp}b\bar{b} \rightarrow l^{\pm}\nu_l l^{\mp}\nu_l l^{\pm}l^{\mp} + b - jets,$$

$$q\bar{q} \rightarrow t\bar{t}W \rightarrow W^{\pm}W^{\pm}W^{\mp}b\bar{b} \rightarrow l^{\pm}\nu_l l^{\pm}\nu_l l^{\mp}\nu_l + b - jets,$$

which can be identified by the presence of jets arising from b-quarks and missing momentum from the neutrinos present.

As the standard model background is expected to be small, detector background such as *fakes* and *charge flip* will be the more important background to take into account. Fakes are particles that are incorrectly reconstructed and assigned the wrong particle type by the detector. Examples often include jets or charged π-mesons being reconstructed as electrons and vice versa. The number of fakes is in general hard to estimate from simulation due to the intricate dependency on specific detector parameters. Often different kind of data-driven methods are favored, and the correct estimation of fakes is often one of the most complex parts of an analysis.

Charge flip, or *charge misidentification*, is when the detector correctly reconstructs the particle type but misidentifies its electromagnetic charge. Charge flip will be discussed in more detail in appendix F.2.

4 τ-inclusive $H^{\pm\pm}$ signal sample generation

4.1 Introduction

While any excess of observed same-sign dilepton events would indicate new physics, the benchmark model chosen by the SSDiLep group is the doubly charged Higgs particle $H^{\pm\pm}$. It is present in certain models extending the Standard Model to explain phenomena such as neutrino mass. The $H^{\pm\pm}$ could be created through a number of different processes in $p-p$ collisions and is expected to be heavy as it has not yet been observed at the current LHC center of mass energies. Being massive it would be observed as a short-lived resonance state, and would in the model considered in this thesis decay into two same-charged leptons of any generation, thus giving it a unique experimental signal of a same-sign dilepton final state that might not conserve lepton number.

A search for new physics involves the use of data sets containing the simulated detector signal of the expected new particles, and the use of these signal samples in physics analyses are twofold. They are necessary for comparison with the data to reject different kinds of physics models: While any difference in the data compared to the expected background would indicate that *something* is wrong with the Standard Model, signals samples can be used for comparison with the observed signal to pinpoint what kind of addition to the Standard Model might complete it. Signal samples also make it possible to analyze what kind of method is needed to be able to plausibly be able to observe a new physics signal and separate it from the background noise.

The purpose of this project was to request a new signal sample consisting of simulated $p-p$ collision events with pair production of $H^{\pm\pm}$ decaying into two same-sign dileptons. A previous signal sample containing the same events but with only light lepton decay modes for the $H^{\pm\pm}$ already existed. To not simulate more of those kinds of events the new simulated events were requested to include at least one τ-lepton in the final state. This project entailed generating event samples, validating that the event generation process worked as intended and requesting the production of several signal samples with different mass for the $H^{\pm\pm}$.

4.2 Theory

4.2.1 The Left-Right Symmetric Model

The following short description of the Left-Right Symmetric Model has been based on Ref. [34].

The Standard Model as it is currently formulated makes a distinction between particles with different so-called chiral state: Depending on whether the spin of a particle is aligned with its momentum or not it interacts differently. The weak interaction has been observed to couple only to particles with their spin anti-parallel to their direction of motion, a so-called *left-handed* chiral state. Particles with spin in parallel with their momentum, so-called *right-handed* particles, cannot interact weakly. While an experimental fact, a common extension of the Standard Model is the so-called Left-Right Symmetric Model (LRSM) which restores symmetry between left- and right-handed particles. It proposes a new set of right-handed weak vector bosons that couples to the right-handed set of elementary particles.

The Standard Model electroweak sector consists of two symmetry groups, $SU(2)_L \times U(1)_Y$. Requiring that the Standard Model Lagrangian is invariant to gauge transformations of those groups naturally give rise to the weak and electromagnetic interactions and their force carriers bosons $W_L^{\pm}$, Z_L, and γ, with the subscript L signifying that only left-handed particles couple to the weak interaction bosons. The LRSM proposes the addition of another symmetry group $SU(2)_R$ giving rise to right-handed weak bosons $W_R^{\pm}$ and Z_R, with the full electroweak symmetry groups commonly being notated as

$$SU(2)_L \times SU(2)_R \times U(1)_{B-L}.$$

These symmetry groups would reduce to the familiar $SU(2)_L \times U(1)_Y$ due to spontaneous symmetry breaking due to a new set of Higgs bosons. Depending on the model this puts lower limits on the masses of the right-handed weak bosons at several hundreds of GeV, as no such interactions have yet been observed.

The LRSM solves several of the current problems of particle physics, most famously the problem of neutrino mass. While the observation of neutrino oscillation proves they are massive particles, no mechanism in the Standard Model can account for this. However, in the LRSM the neutrino can be its own anti-particle, also-called a Majorana particle, which enables new mass neutrino generation terms to be allowed when coupling to the new Higgs sector. Even if there currently is no experimental indication that any LRSM is correct, searches for physics predicted by the LRSM is one of many paths for solving the remaining discrepancies in the Standard Model.

4.2.2 Doubly Charged Higgs in the LRSM

The following short description of the Doubly Charged Higgs has been based on Ref. [34].

A new set of scalars bosons to account for the spontaneous symmetry breaking is one possible experimental signature predicted by the LRSM. These consist of a Higgs bi-doublet and a Higgs triplet Δ, with the latter having the components

$$\Delta_{L/R} = \begin{pmatrix} H^{+}_{L/R} & \sqrt{2}H^{++}_{L/R} \\ \sqrt{2}H^{0}_{L/R} & -H^{+}_{L/R} \end{pmatrix}$$

with corresponding charge conjugated triplet giving negatively charged components. The symmetry of the LRSM model predicts both a left- and a right-handed Δ triplet, with the right-handed one being responsible for the symmetry breaking of the $SU(2)_R$ symmetry group. As dictated by their handedness these two Higgs triplets will interact differently with the Standard Model particles and will thus have different production cross section depending on the chosen model. Both should however give rise to the same final states.[2]

The vacuum expectation value (*vev*) of the new Higgs triplet is given by

$$< \Delta_{L/R} >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ v_{L/R} & 0 \end{pmatrix}$$

[2]When referenced to in this experimental sense the handedness subscript will be dropped.

with v being the potentially non-zero *vev* of the neutral component of the triplet. The coupling of the Higgs triplet to $W^{\pm}$ depend directly on v, and its value thus decides the possible decay modes of the $H^{\pm\pm}$. All possible decay modes are found in Ref. [34], with a smaller *vev* implying a preference for the leptonic decay modes. In the case of a small or identically zero *vev* the leptonic decay modes will dominate exclusively, and in the context of this thesis only the leptonic decay modes will be considered.

4.2.3 $H^{\pm\pm}$ signal

There exist several different production channels for $H^{\pm\pm}$ in $p-p$ collisions. In the context of this thesis the most common one encountered, described in for example Ref. [42], is pair production of $H^{\pm\pm}$ by the annihilation of a quark anti-quark pair with an intermediate state of an off-shell Z/γ, commonly referred to as a Drell-Yan process:

$$q\bar{q} \xrightarrow{Z/\gamma^*} H^{\pm\pm}H^{\mp\mp}.$$

The $H^{\pm\pm}$ are assumed to be on-shell resonances and decay rapidly into two same-sign dileptons, not necessarily of the same generation. The leading order Feynman diagram for the process is presented in figure 6.

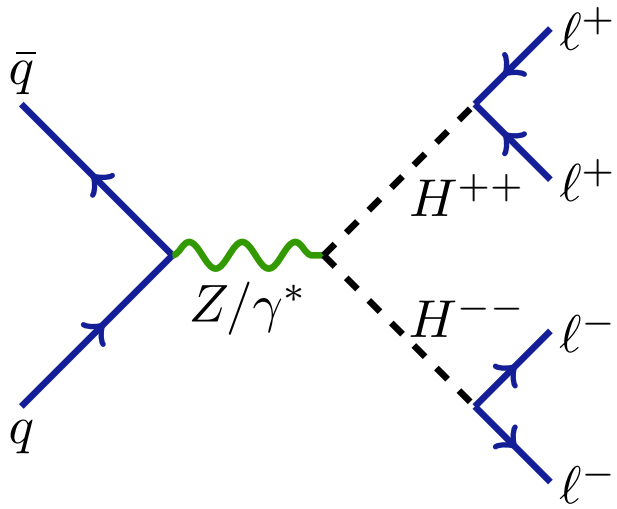


Figure 6: Leading order Feynman diagram of the Drell-Yan process $q\bar{q} \to H^{++}H^{--} \to l^+l^+l^-l^-$ [16].

This kind of basic $H^{\pm\pm}$ model implies a clear detector signal that is not predicted by the standard model: For high enough collision center-of-mass energies there should be a significant excess of prompt same-sign dileptons, and most events should be lepton number violating. Reconstructing the invariant mass of the dileptons would result in a resonance peak in the invariant mass spectrum at the $H^{\pm\pm}$ mass. All leptons should have high p_T due to the assumed high mass of the $H^{\pm\pm}$. The final state should contain low p_T^{miss} due to the absence of neutrinos expect for ν_τ from

potential τ-leptons. As discussed in section 3.3 the most significant prompt background for this final state is from Standard Model diboson processes.

In the case of a non-zero *vev* a single $H^{\pm\pm}$ could also be produced in association with a $W^{\pm}W^{\pm}$-fusion. That production channel is of interest as the high center-of-mass energy required to pair-produce two heavy resonances rapidly decrease the Drell-Yan cross section for increasing $H^{\pm\pm}$mass. Such a channel would thus be able to probe significantly higher $H^{\pm\pm}$ masses than any pair-production channel. However, it is highly dependent on both the *vev* and the $W^{\pm}_{R/L}$ masses, and while a same-sign dilepton would still be a possible final state the inclusion of $W^{\pm}$ coupling increases the number of possible final states. No $W^{\pm}$ dependent channel has been considered in the context of this thesis.

4.2.4 τ-leptons in $H^{\pm\pm}$ decay

There exist various models for the different coupling schemes between the $H^{\pm\pm}$ and the different lepton generations. For the simple case of equal coupling to all leptons there exist a strong case for the inclusion of τ-leptons, as it would give access to more statistics than only light lepton channels can provide. However, due to the τ decay modes the light lepton final states are still expected to show the clearest signal if one exists. Some of the τ statistics is also already present in a light lepton analysis due to the light lepton decay modes, or will lost due to the reconstruction algorithms failing to separate τ-jets from the background.

Theories with mass-dependent coupling would however strongly prefer the $H^{\pm\pm}$ to decay into τ-leptons, and such signals would thus be suppressed in same-sign analyses only using light lepton final states. An interesting new possibility when including τ-leptons into the same-signed analysis is then to probe $H^{\pm\pm}$models with a mass-dependent coupling, as described in for example Ref. [43]. $H^{\pm\pm}$ lepton coupling can also depend on the neutrino mass hierarchy, which again requires τ-lepton-inclusive final states to be fully examined [29, 23].

4.3 Method

Producing signal samples is computationally expensive, and all signal samples are produced centrally for all ATLAS groups by request from the subgroups. To prevent the generation of unnecessary or erroneous data sets every signal sample generation request needs to be properly validated and discussed before being produced. The required request procedure for the Exotics ATLAS subgroup was followed and can be found at Ref. [7].

Event generation data sets needed to be generated for validation before the signal samples could be produced. Event generation validation was done by producing validation histogram from Pythia event generation output of the final state particles and their kinematics. The histograms were used to show that the expected behavior of the process was correctly generated. The event generation samples needed to produce these validation histograms was generated on the grid using Athena interfaced with the grid submission tool PanDA.

Validation histograms of the event generation samples were produced using Athena with a C++ based validation script. The full validation code is available at Ref. [14], and includes basic

instruction on how to both generate $H^{\pm\pm}$ event samples locally and using the grid, and how to produce validation plots.

ROOT-readable ntuples created from the signal sample DxAODs using the DxAODAnahelpers framework. Kinematics histograms of these ntuples were created using the Python-based `MergedFramework` scripts as developed for the SSDiLep light lepton analysis, and extended as part of this thesis to include τ-lepton variables.

The plots of these histograms as presented in section 4.5 were produced using a ROOT script.

4.3.1 Pythia

Pythia will refer to the C++ based Pythia8 which has been used exclusively for this thesis. Pythia is a programming framework for simulating particle collisions and their outcome using the Monte Carlo method. Theoretically it is based on the so-called Lund String Model, which is a phenomenological model of the hadronization process of the strong interaction. The full introduction to the Pythia software and its theoretical framework can be found in Ref. [46].

Pythia is implemented in the Athena framework, and take as input a JobOptions Python file which specifies the details of the events to be generated. The basic setup for any Pythia event generation consists of choosing a *tune* and a *parton density function* (PDF). The tune decides a wide array of different parameters of the event generation, which most notably include the strong coupling constant α_s of the Standard Model needed for the simulation of parton showers. The tune parameters are dependent on detector data, and there exist several different tunes based on different detector data sets and processes. The most common one used by ATLAS event generation, including the ATLAS A14 central tune used in this thesis, is based on the *Monash 2013* tune [47]. The PDF describes the distribution of quarks and gluons of the proton, a needed part to correctly simulate the result of a $p-p$ collisions. As with the tune the PDF is modeled from data, as there currently exists no theoretical framework that can calculate a PDF with the same precision as experimental data can.

4.3.2 ATLAS detector simulation software

A full description of ATLAS detector simulation is given by Ref. [27].

The software framework used to simulate particle collisions in the ATLAS detector is called GEANT. It takes as input event generation samples and outputs xAOD data sets. It simulates both the interaction of the event particles with the detector material and the effect reading the kinematics variables as digital information has on the data (*digitization*).

Running the full GEANT simulation is computationally intensive, and to save computing time several methods to approximate a full simulation has been developed. These so-called *fast simulation* methods use different methods to estimate the effects of the detector without having to run the full simulation. The goal is to reproduce the result as with full simulation while using less computational power, and while the quality of the simulation by definition is worse fast simulation is in many cases similar to full simulation. A common use for fast simulation is for samples simulating a new physics signal. Often many sample need to be generated over a range

of masses for the particles of the new model, and in general many new physics models exist that need signal samples. Signal samples as educated guesses on how a new physics signal might look also have a lesser need of being perfectly modeled. Conversely, for precision measurements of Standard Model processes full simulation is often chosen, as simulation accuracy is needed for comparison to the large amount of data available for common Standard Model processes.

4.4 Data sets

4.4.1 Event generation samples

The τ-inclusive $H^{\pm\pm}$ event samples used as a foundation for the signal sample request were generated with Pythia using the ATLAS central tune A14 with the parton density function NNPDF23LO [20]. These settings were unchanged from the previous $H^{\pm\pm}$ signal samples with only light leptons [38].

The job options file used to define the parameters for the event generation is presented in appendix D. It specifies the event simulation to be the Drell-Yan pair-production scenario of $H^{\pm\pm}$ as described in Ref. [34] and discussed in section 4.2.3. The *vev* of the neutral part of the Higgs triplet of which $H^{\pm\pm}$ is a part of was set to zero: A non-zero *vev* enables decay into W-bosons, a decay mode that was not to be simulated. The branching ratio of the Yukawa coupling of $H_L^{\pm\pm}$ to all combinations of leptons[3] was set to the same value, giving an equal branching ratio of $BR(l^{\pm}l^{\pm}) = \frac{1}{6}$ to all leptonic decay modes disregarding phase space considerations. No $H_R^{\pm\pm}$ was generated as part of the signal sample as the model would result in the same kinematic and decay modes as $H_L^{\pm\pm}$. Therefore the Pythia cross-section of the signal sample could be recalculated to account for the presence of $H_R^{\pm\pm}$ if needed. The theoretical cross section of Drell-Yan pair production of $H^{\pm\pm}$ as calculated to NLO at $p-p$ collision with a center-of-mass energy of $\sqrt{s} = 13$ TeV is presented in table 4.

[3]The six possible decay modes in this scheme are $H^{\pm\pm} \to e^{\pm}e^{\pm}$, $H^{\pm\pm} \to \mu^{\pm}\mu^{\pm}$, $H^{\pm\pm} \to \tau^{\pm}\tau^{\pm}$, $H^{\pm\pm} \to e^{\pm}\mu^{\pm}$, $H^{\pm\pm} \to e^{\pm}\tau^{\pm}$ and $H^{\pm\pm} \to \mu^{\pm}\tau^{\pm}$.

Table 4: Theoretical NLO cross section for Drell-Yan pair production of DCH with $p-p$ collisions at $\sqrt{s} = 13$ GeV. LO cross sections shown with the correction factor k, with $\sigma_{LO} \cdot k = \sigma_{NLO}$. Reproduced from table 2 in reference [18].

$m(H^{\pm\pm})$ [GeV]	$\sigma_{LO}(H_L^{\pm\pm})$ [fb]	k-factor$(H_L^{\pm\pm})$	$\sigma_{LO}(H_R^{\pm\pm})$ [fb]	k-factor$(H_R^{\pm\pm})$
300	13.294	1.2565	5.6299	1.2536
400	3.9255	1.2483	1.6732	1.2483
500	1.4200	1.2414	0.60807	1.2441
600	0.58311	1.2355	0.25072	1.2401
700	0.26132	1.2304	0.11281	1.2370
800	0.12463	1.2267	0.05402	1.2354
900	0.06228	1.2267	0.02712	1.2375
1000	0.03229	1.2281	0.01412	1.2409
1100	0.01722	1.2315	0.00757	1.2463
1200	0.00939	1.2376	0.00415	1.2545
1300	0.00522	1.2465	0.00232	1.2652

10 event generation samples with difference $H^{\pm\pm}$ mass $m(H^{\pm\pm})$ was generated in the range $[300, 1300]$ GeV in steps of 100 GeV. A selection of truth-level validation histogram of the PYTHIA event generation output is presented in section 4.5.1.

4.4.2 Signal samples

The signal sample request was based on the event generation samples generated. All mass points were requested to be simulated using Atlfast II (*fast*) simulation. One mass point ($m(H^{\pm\pm}) = 600$ GeV) was also requested to be simulated using full simulation to have a data set for comparison: Due to the inclusion of τ-leptons the effect on fast simulation on the signal samples is uncertain, and by comparing fast and full simulation enables any systematical errors to be identified and the validity of using fast simulation can be examined. The specific mass point was chosen in the lower intermediate range where there is more likely for a signal to be observed in data while still being a general mass point than those at the extreme ends of the mass point range.

The samples are summarized in table 5. The samples were requested as DxAODs with the SUSY3 derivation [6] but without any skimming. As such all events were preserved during the derivation steps but the variable collection matched that of a SUSY3 DxAOD. This was done in anticipation of using SUSY3 as the main derivation for τ-inclusive data, as the derivation requires each event to contain at least one τ_{had} with $p_T > 15$GeV making it useful for an inclusive analysis with at least one τ-lepton present in an event.

A selection of event variable histograms such as different invariant mass distributions of the simulated $m(H^{\pm\pm}) = 600$GeV signal samples is shown in section 4.5.2.

Table 5: PYTHIA pair production of $H^{\pm\pm}$ MC signal samples. DSID is the dataset ID, used for identifying the JobOptions used for a particular simulation. Cross section and filter efficiency taken from the PYTHIA output logs for use when normalizing the data sets to a particular integrated luminosity. Filter efficiency as calculated by taking the ration of events of the data set $n_{dataset}$ with the total generated events $n_{generated}$. The cross-sections can be compared to the theoretical LO cross sections in table 4.

$m(H^{\pm\pm})$ [GeV]	DSID	$n_{dataset}$ [10^3]	$\sigma_{Pythia}(H_L^{\pm\pm})$ [fb]	$\varepsilon_{filter} = \frac{n_{dataset}}{n_{generated}}$
300	309651	50	13.896	0.80250
400	309652	50	4.1186	0.80559
500	309653	50	1.4844	0.80087
600	309654	50	0.61400	0.80337
700	309655	50	0.27600	0.80265
800	309656	50	0.13337	0.80075
900	309657	50	0.067400	0.80226
1000	309658	50	0.035400	0.80499
1100	309659	50	0.019200	0.80317
1200	309660	50	0.010600	0.80405
1300	309661	50	0.0060186	0.80131

4.4.3 Filter and final state statistics

As pair production of $H^{\pm\pm}$ decaying into light leptons existed, the event generation filter *ParentChildFilter [31]* was used to require that the generated event contains at least one τ-lepton with a $H^{\pm\pm}$ as the parent. This would filter out any generated pure light lepton event, and thus increase the number and statistics of the τ-inclusive events of interest. It should be noted that the event sample still contained final states with four apparently prompt light leptons, but of which at least one would have originated from a τ_{lep}.

In a 4-lepton final state in which each lepton can be either an e, μ or τ with equal probability there exist a total of $3^4 = 81$ lepton combinations. In a similar manner, the possible combinations of 4-lepton final states with only e or μ is $2^4 = 16$. Due to filtering away all pure light lepton final states the ratio of generated events that are removed by the filter and the total amount of generated events, referred to as theoretical filter efficiency ε_{filter}, is

$$\varepsilon_{filter} = \frac{(81-16)}{81} \approx 80.2\%.$$

4.5 Result

4.5.1 Event generation validation histograms

Details of the event generation for the signal sample request including full sets of validation plots are found at Refs. [12, 13]. A selection of the event generation validations histograms is presented here. To increase the histogram readability the bin errors have not been plotted in favor for plotting all mass points, as this allows comparison of final state distribution and kinematics between different mass points. The kinematics validation plots are divided up between negative and positive dileptons to validate charge invariance in the generated dataset, and no difference in the distributions can be seen

The observable final state of the four daughter leptons of the pair-produced $H^{\pm\pm}$ of the signal sample is presented in figure 7. The method to find the expected statistics of each 4-lepton final state channel is presented in appendix 8, and the observed final state distribution matches the expected from the final state combinatorics.

The decay modes of the $H^{\pm\pm} \to l^{\pm}l^{\pm}$ is presented in figure 8. The invariant mass spectrum of the dileptons showing the clear resonance peaks of the $H^{\pm\pm}$ is presented in figure 9. p_T of the leading and sub-leading lepton is presented in figures 10 and 11. The leading and sub-leading lepton η is presented in figures 12 and 13. All histograms show the expected distributions based on the event generation settings. The same lepton kinematics distributions are observed as for the previous pure light lepton event sample.

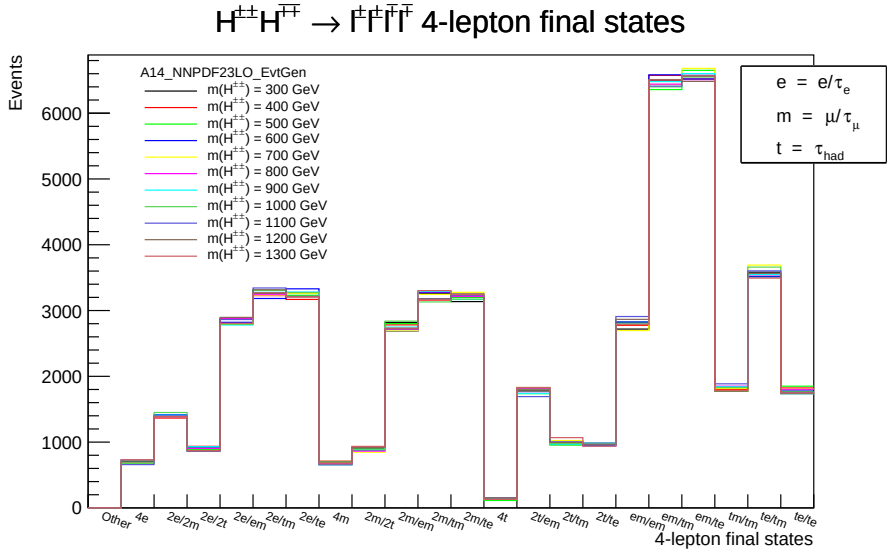


Figure 7: 4-lepton final state distribution for the τ-inclusive $H^{\pm\pm}$ signal samples. Each bin on the x-axis represents a final state of the type $l^{\pm}l^{\pm}/l^{\mp}l^{\mp}$, with e =electron/τ_e, m =muon/τ_μ and $t = \tau_{had}$.

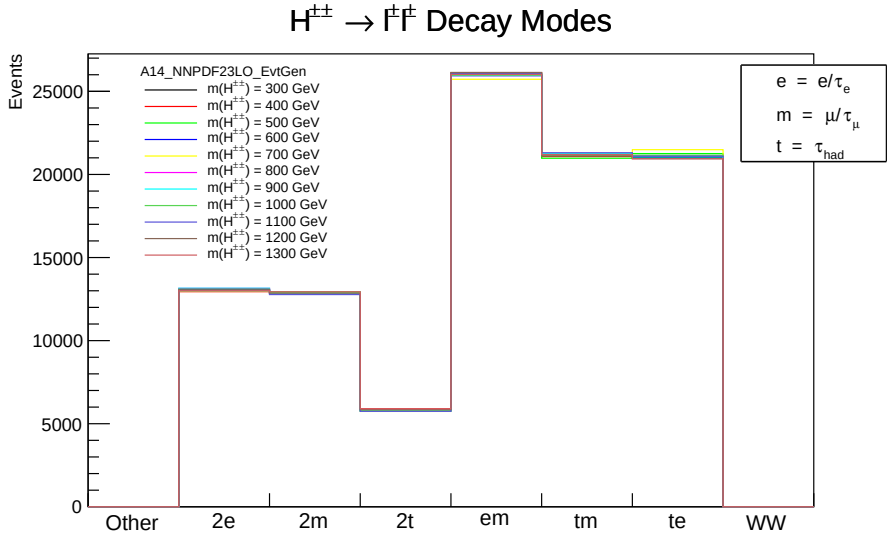
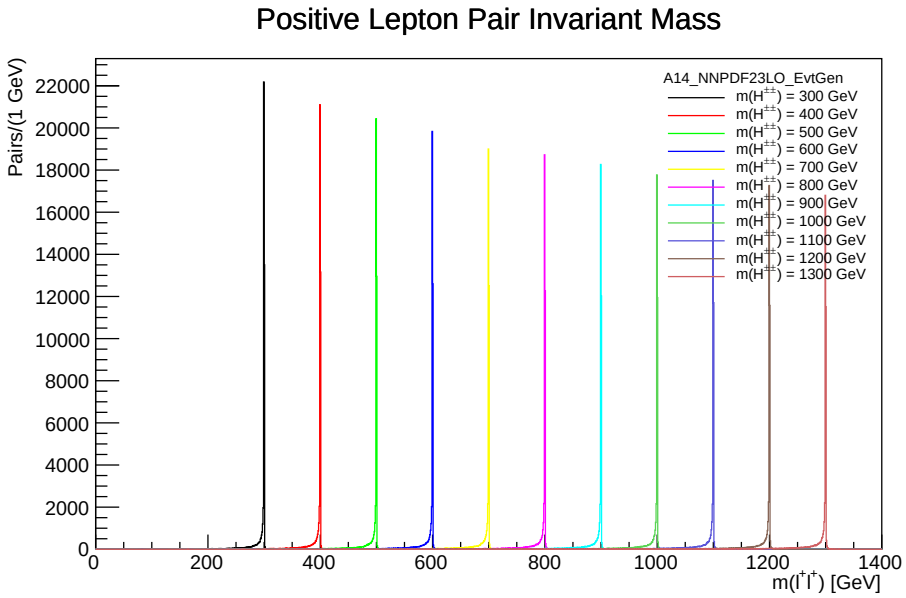
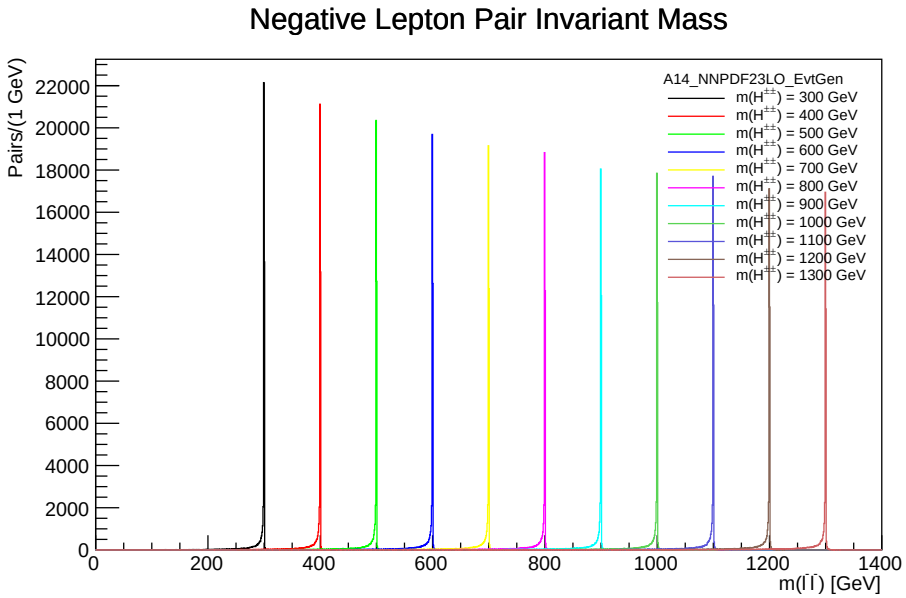


Figure 8: $H^{\pm\pm}$ decay modes for the τ-inclusive event samples. The favour of pure light lepton final states due to τ_{lep} can be clearly seen. WW and Other decay channels displayed to validate nothing else was generated.

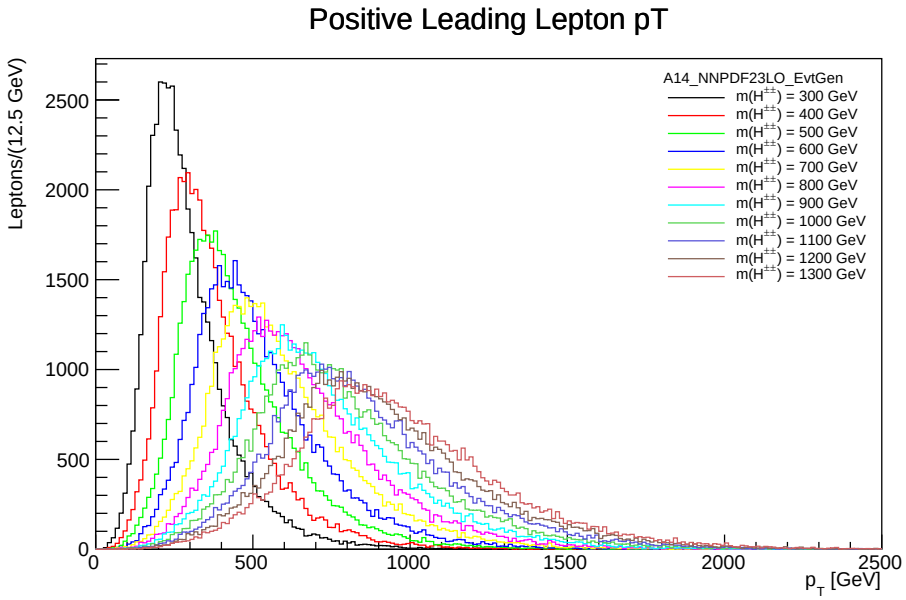


(a) Positively charged dilepton.

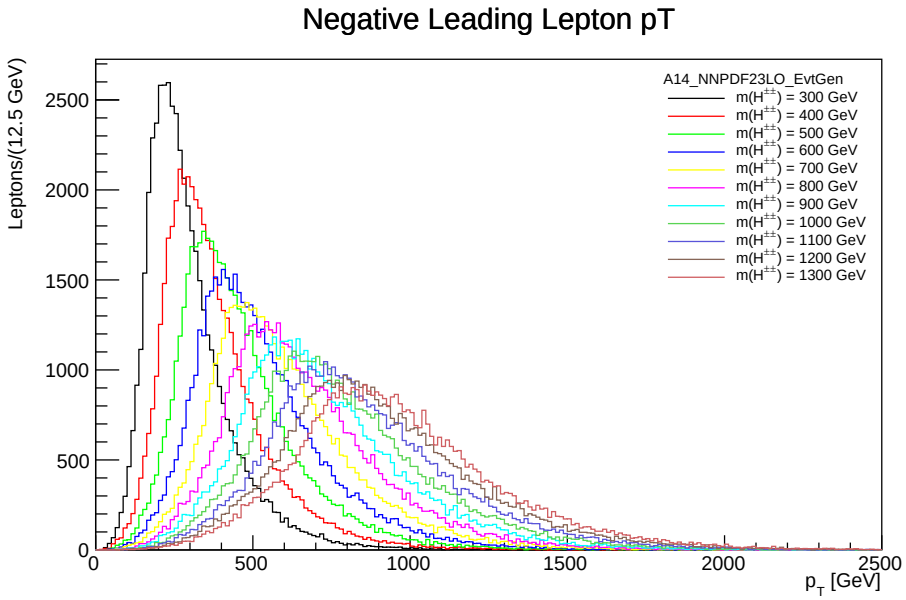


(b) Negatively charged dilepton.

Figure 9: Invariant mass of same-sign dileptons for the τ-inclusive $H^{\pm\pm}$ signal samples. Close match with the input mass value can be observed, with minor smearing of the resonance peak to lower invariant mass caused by photon radiation of the leptons.

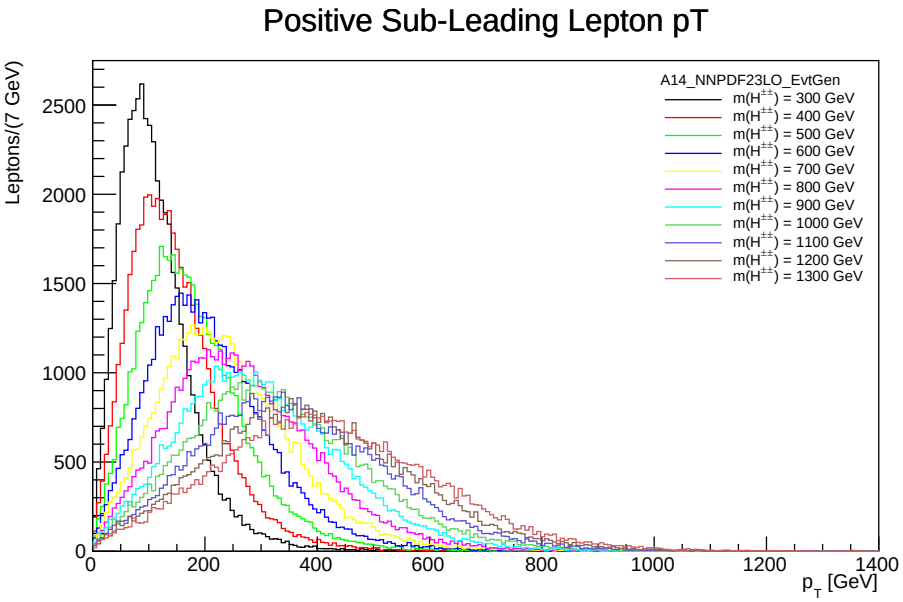


(a) Positively charged dilepton.

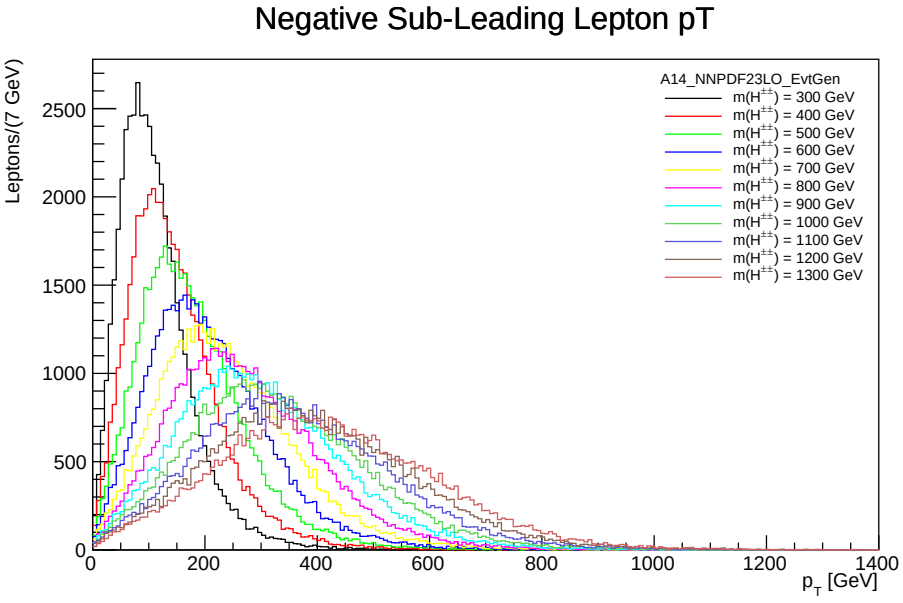


(b) Negatively charged dilepton.

Figure 10: Leading lepton p_T for the τ-inclusive $H^{\pm\pm}$ signal samples.

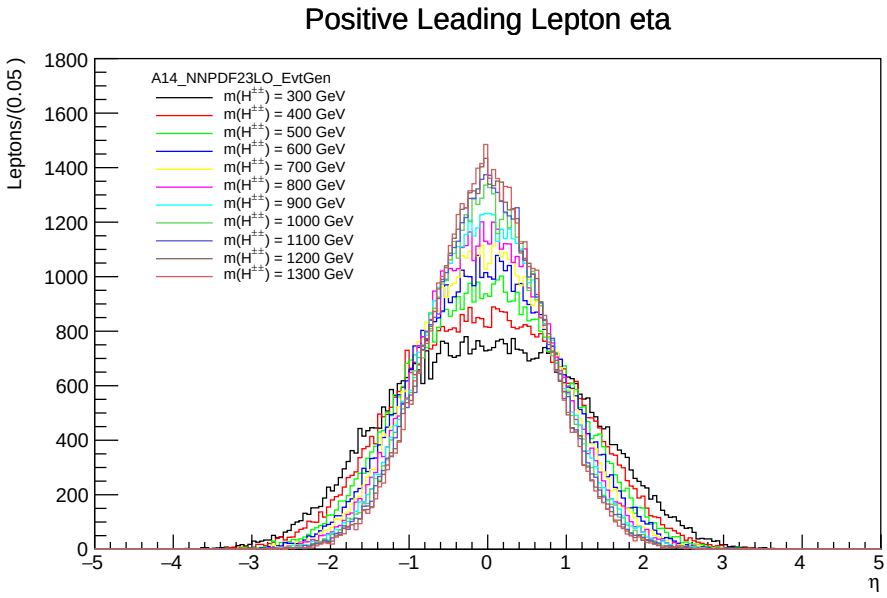


(a) Positively charged dilepton.

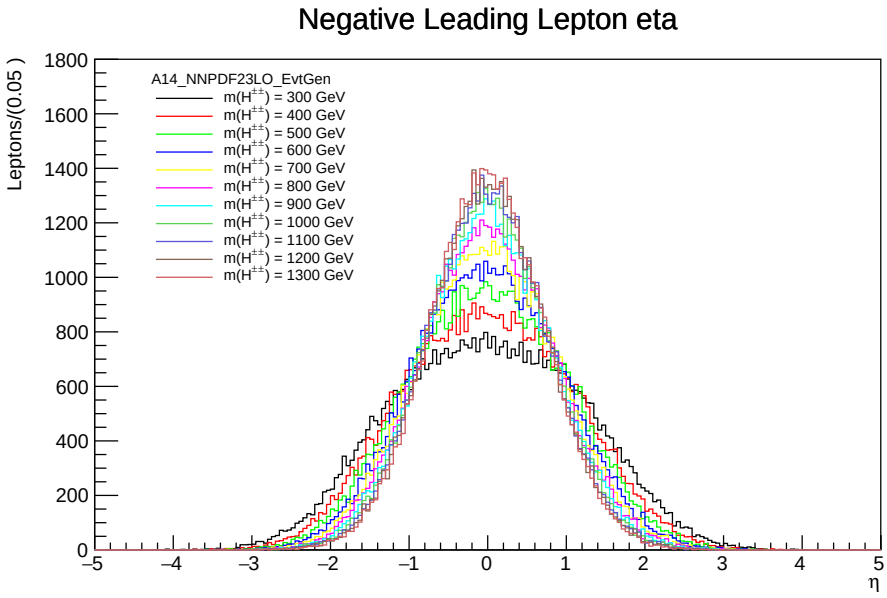


(b) Negatively charged dilepton.

Figure 11: sub-leading lepton p_T for the τ-inclusive $H^{\pm\pm}$ signal samples.

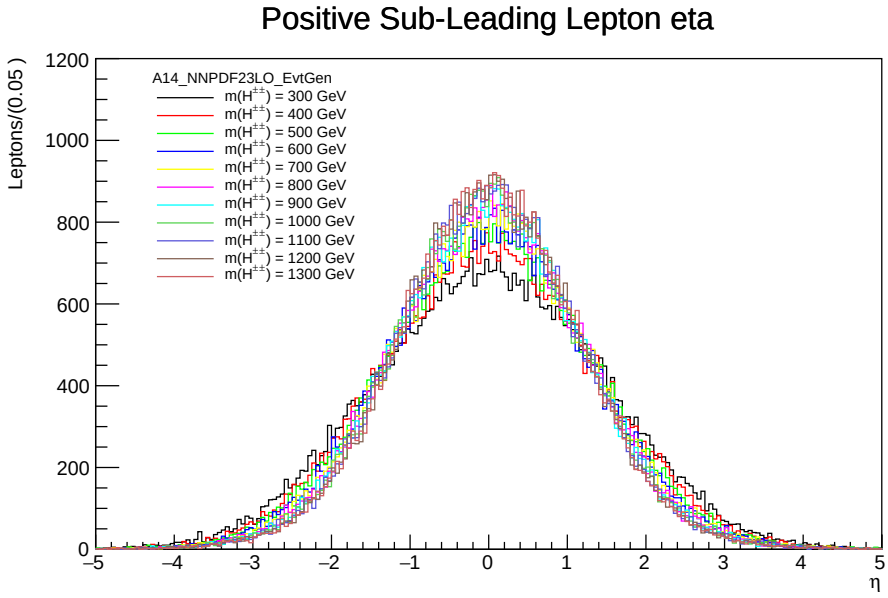


(a) Positively charged lepton.

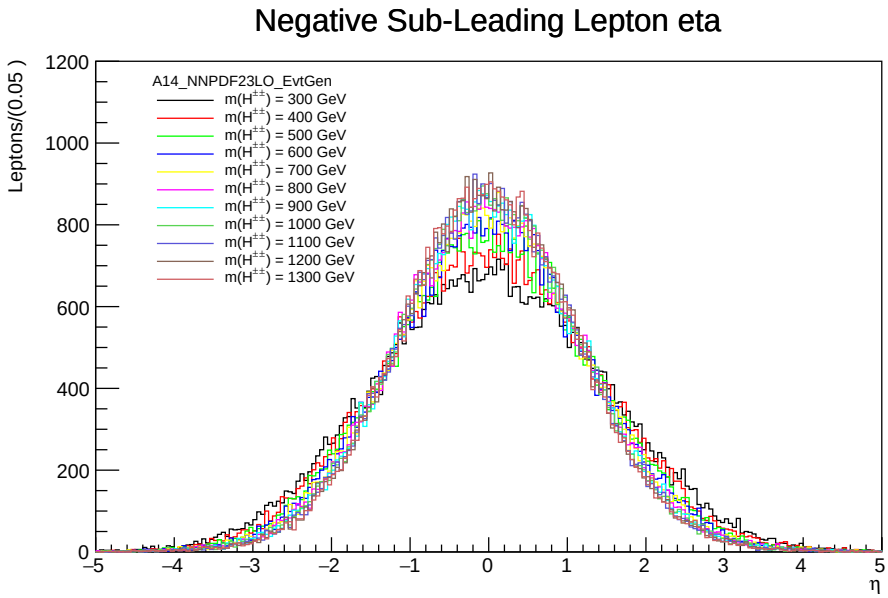


(b) Negatively charged lepton.

Figure 12: Leading lepton η for the τ-inclusive $H^{\pm\pm}$ signal samples.



(a) Positively charged lepton.



(b) Negatively charged lepton.

Figure 13: Sub-leading lepton η for the τ-inclusive $H^{\pm\pm}$ signal samples.

4.5.2 Signal sample histograms

A selection of events variables of the finished signal samples is presented in the following sections. The purpose is to get a qualitative overview of the effect of the detector simulation on the event generation samples and confirm that the signal samples give expected results. It is important to

note that no quantitative conclusions should be drawn based on the following histograms.

One same-sign dilepton where at least one of the leptons was a τ_{had} was required unless otherwise noted. In the event of multiple same-sign dileptons the one having the highest sum of p_T (the *leading* same-sign dilepton) of its constituting leptons was chosen with the assumption that a lepton daughter of a heavy resonance should have a large p_T. Several leptons of different generation but with the same charge, p_T, η, and ϕ indicated a problem with fakes. As no overlap removal was implemented a small cut of opening angle in the transversal plane, $|\Delta\phi| < 0.05\pi$, was used to remove the most obvious overlap. This should remove few true events as the large mass of the resonance compared to the daughter particles give rise to few collinear opening angles.

The number of τ_{had} per event for the $m(H^{\pm\pm}) = 600$GeV signal samples is presented in Figure 14. As identical event generation was used for both samples, any difference in the number of reconstructed τ_{had} must arise due to the detector simulation step. To note is that there are 28 events with $4\tau_{had}$ for the fast simulation sample, and the full simulation sample has 34 $4\tau_{had}$ events. A cutflow overview detailing the number of events present after different cuts are presented in table 6.

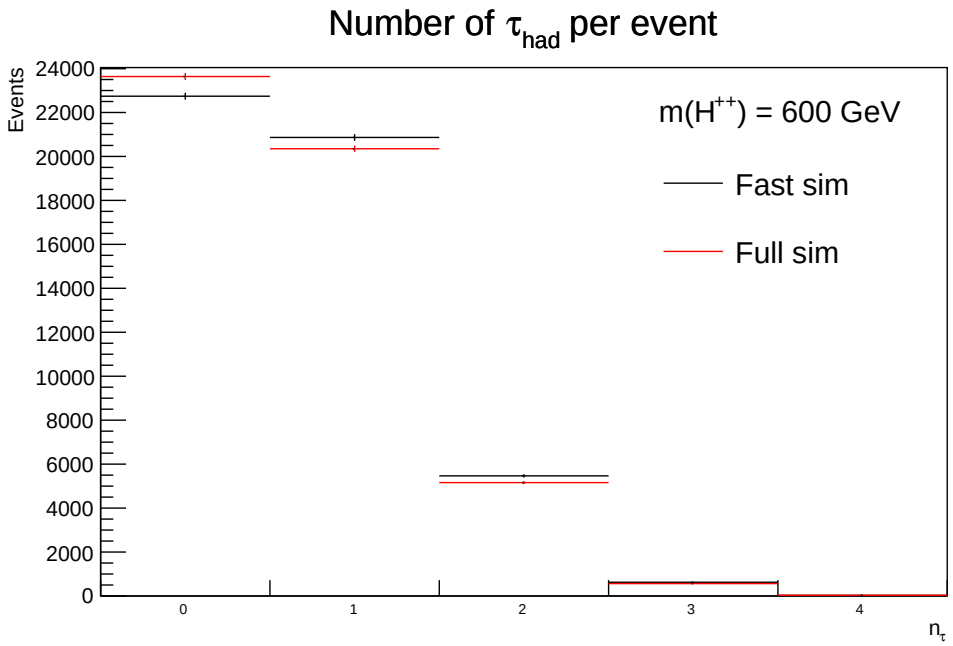


Figure 14: Number of τ_{had} per event of the $m(H^{\pm\pm}) = 600$GeV signal samples.

Table 6: Cutflow overview.

	Events		Normalized events		
	Fast sim	Full sim	Fast sim	Full sim	$\frac{N_{fast}}{N_{full}}$
Signal sample	50000	50000	1	1	1
$(l^{\pm}l^{\pm}) \geq 1$, $\Delta\phi \geq 0.05\pi$	38196	37938	0.76392	0.75876	1.0068
$\tau_{had} \geq 1$	18893	18192	0.37786	0.36384	1.0385

Histograms using the $m(H^{\pm\pm}) = 600$GeV signal sample are showed as to allow comparison between full and fast simulation. Both simulation types have been plotted in all histograms to show the expected similarity of their distributions, and in the case of a discrepancy it will be noted.

Leading same-sign dilepton type

The leading same-sign dilepton type is presented in figure 15, giving which combination of leptons the chosen dilepton consists of. As expected from requiring at least one τ_{had} as part of the dilepton, no pure light lepton combinations are seen.

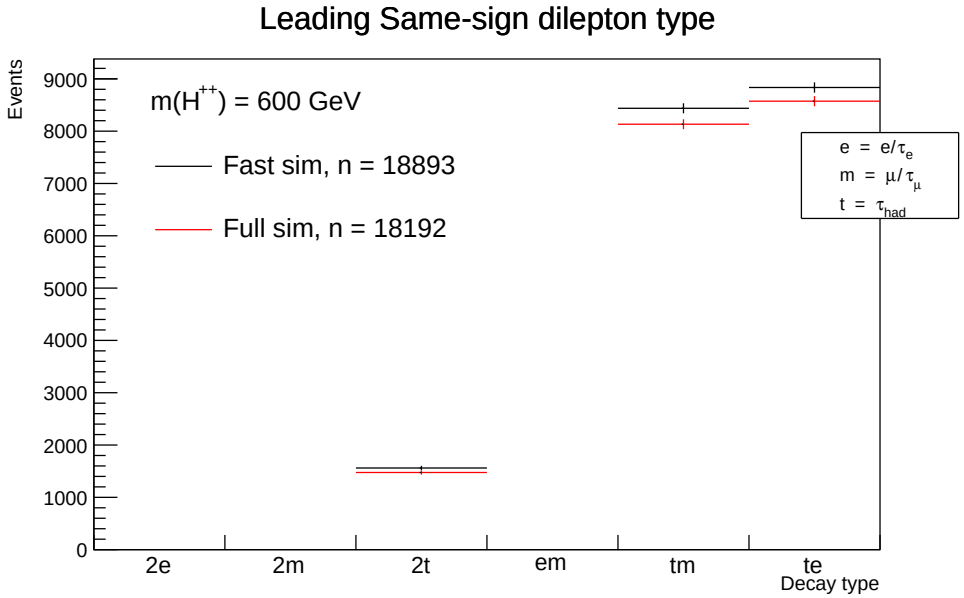


Figure 15: Number of τ_{had} per event of the $m(H^{\pm\pm}) = 600$GeV signal samples. Fast simulation reconstructs slightly more τ_{had} overall.

For a more direct comparison with the decay modes of the event generation (Fig. 8) the leading same-sign dilepton type without any requirement of the presence of at least one τ_{had} is presented in figure 16.

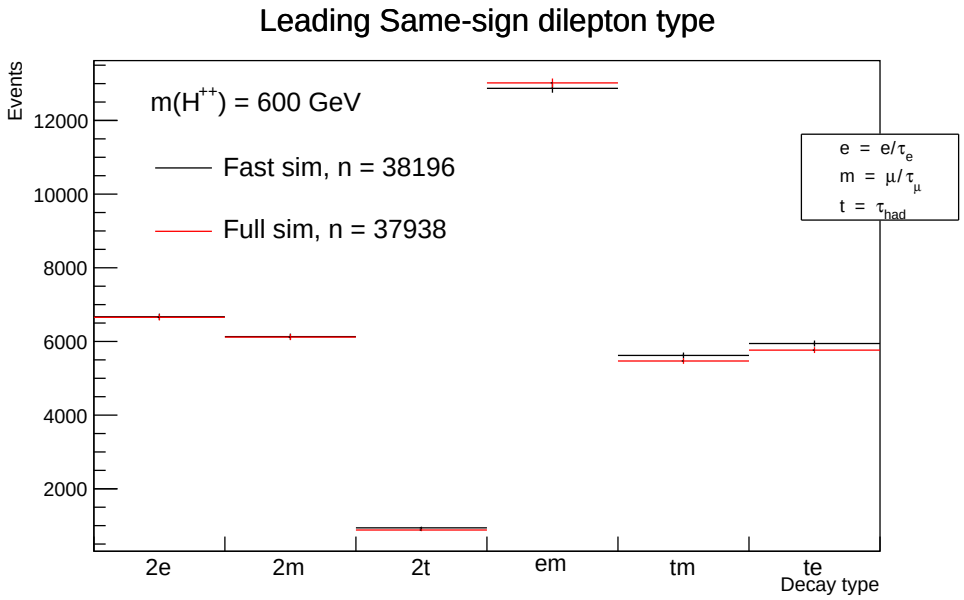


Figure 16: Number of τ_{had} per event of the $m(H^{\pm\pm}) = 600$GeV signal samples. No $\tau_{had} \geq 1$ requirement.

$\Delta\eta$ and $\Delta\phi$ distributions

The transversal opening angle $\Delta\phi$, and the pseudorapidity difference $\Delta\eta$ (corresponding to opening angle in the beam-line direction) between the visible decay products of a $H^{\pm\pm}$ are shown in figure 17 and figure 18 respectively.

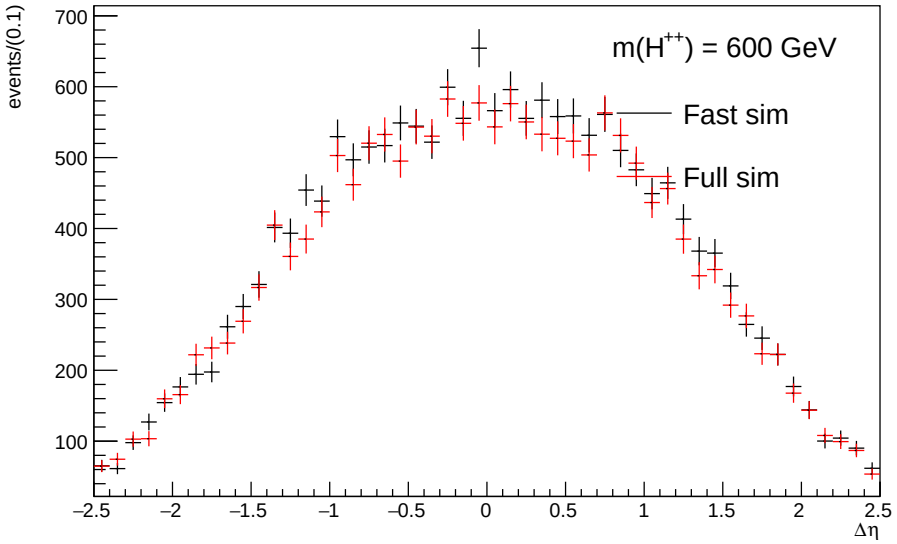


Figure 17: Pseudorapidity difference $\Delta\eta$ of the leading same-sign dilepton.

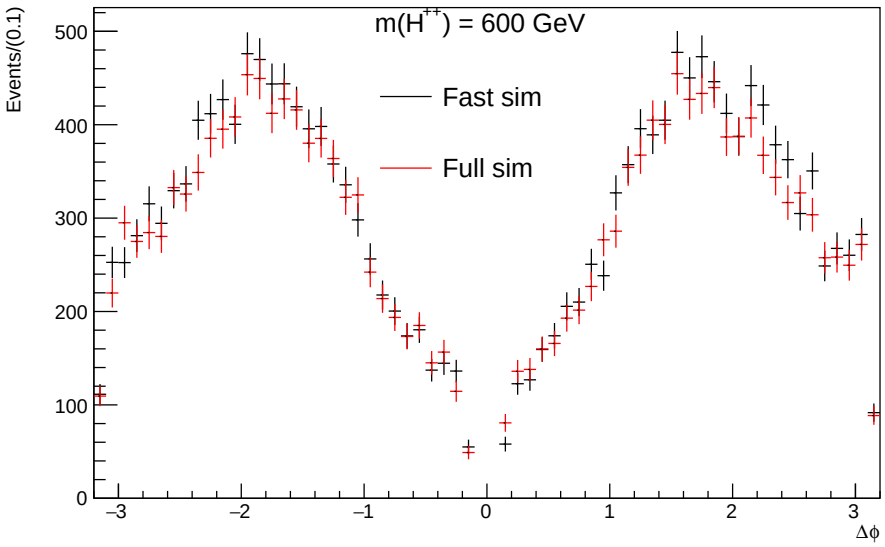


Figure 18: Transversal opening angle $\Delta\phi$ of the leading same-sign dilepton. The lack of events around the origin is due to the $|\Delta\phi| < 0.05\pi$ cut.

Invariant mass

The visible invariant mass $m((l/\tau_{had})\tau_{had})$ is presented in figure 19. It exhibits the expected shape of a resonance peak with missing energy present in the form of the neutrinos from the τ-lepton decay: A broad shape with a cutoff at the resonance mass.

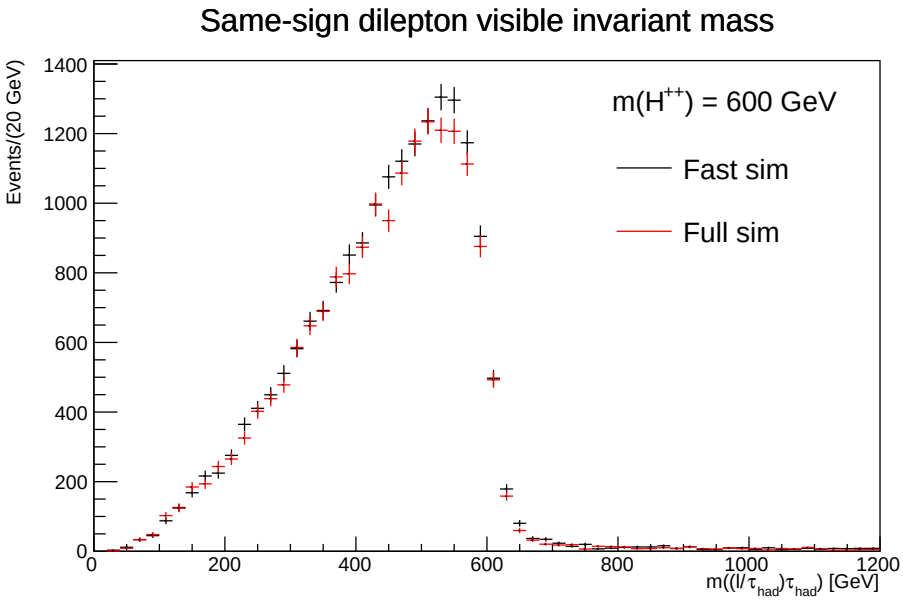


Figure 19: Visible invariant mass of the leading same-sign dilepton of the $m(H^{\pm\pm}) = 600$GeV signal samples.

The total invariant mass $m((l/\tau_{had})\tau_{had})p_T^{miss})$ where the missing energy of the event is part of the mass reconstruction is presented in figure 20. It correctly reconstructs the resonance mass at around 600GeV but is broadened by the fact that there are multiple possible sources of missing energy of the event.

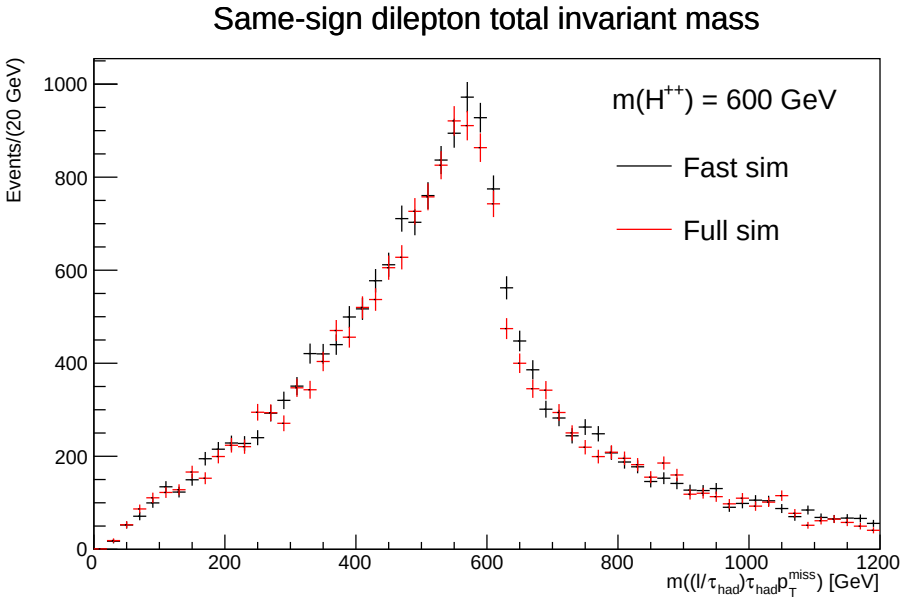


Figure 20: Total invariant mass of the leading same-sign dilepton of the $m(H^{\pm\pm}) = 600$GeV signal samples.

The invariant mass as calculated using *the collinear approximation* for invariant mass reconstruction of partly invisible resonance decays, $m_{col}((l/\tau_{had})\tau_{had}))$, is presented in figure 21. It correctly reconstructs the resonance mass at around 600GeV without the broadening of the distribution at lower invariant mass. The collinear approximation is discussed more in depth in appendix C. It is characterized by enabling the full reconstruction of the invariant mass of two resonance particles that might have invisible parts of their decay. It assumes that all invisible decay products (such as ν_τ in this case) are collinear with the visible decay products (the τ_{had}). This approximation, however, breaks down in decays where the resonance particles are created back-to-back, resulting in a large tail of the invariant mass distribution.

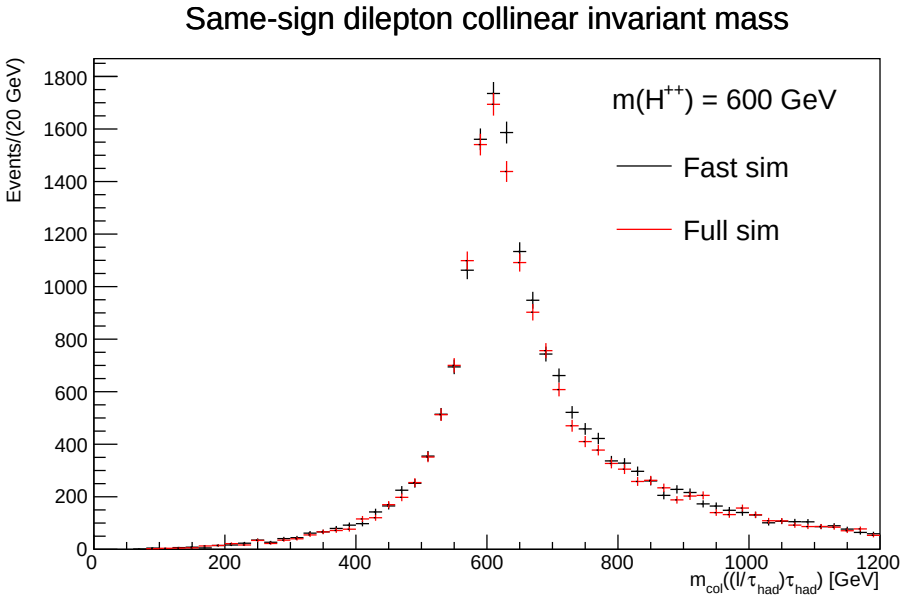


Figure 21: Collinear invariant mass of the leading same-sign dilepton of the $m(H^{\pm\pm}) = 600$GeV signal samples.

4.6 Discussion

Event generation

As mentioned in section 4.2.3 there are theories with $H^{\pm\pm}$ coupling to mass. In that case the golden channel for observing a signal would be the 4τ final state, which exemplifies the problem with different statistics between the decay channels: In the context of the τ-inclusive $H^{\pm\pm}$ signal sample comparatively few events 4τ will be generated with the event generation settings.

There exist several possible filter schemes to solve the problem of low event count of certain rarer final states, as in the case of the $4\tau_{had}$ final state. One method would be to introduce a weighting for each final state in the generator filter based on the relative abundance of each final state. This could result in every $4\tau_{had}$ final state that is generated passes the filter, while the more common channels have a probability of not passing the filter, with the most common ones having the largest probability of being rejected. The end result would be a event sample where the expected

number of events in every final state would be the same, ensuring all channels have reasonable statistics for any physics analysis. The weights used to generate the signal sample would then have to be used in the analysis stage to rescale all channels to their appropriate relative scale. This approach requires the combinatorics of each final state to be calculated, which while tedious is not complex. It would also give more power to rescale the signal sample to test for different kinds of signal models, for example those with higher coupling to τ-leptons, while not having to generate a completely new signal sample. The primary reasons to why this approach was not chosen for the final request was that the scheme would be harder to implement and validate, and more difficult to argue for in the context of requesting signal sample production. However, from a pure physics analysis perspective the probabilistic filter approach has certain advantages and could be considered for future generation of signal samples.

Signal samples

It is expected that fast simulation of τ-inclusive signal samples should give results similar to full simulation, as shown in for example in Ref. [44] comparing fast and full simulation of τ-leptons using $Z \to 2\tau$ signal samples. The high mass of the signal $H^{\pm\pm}$ will however significantly affect the kinematics of the τ-lepton decay products, making it uncertain on how similar much of the Z-decay comparison study is applicable to the simulation of high mass resonances. The full and fast simulation signal samples, while not identical, show very similar kinematics distributions as expected by being based on the same event generation. However, one notable difference is how the fast sample have slightly more reconstructed τ_{had} than the full simulation as can be clearly seen in figure 14. This effect seems to be mostly regarding τ-leptons: As can be seen in figure 16 the biggest discrepancies are seen in the channels involving a τ_{had}, while the pure channels $2e$ and 2μ are identical across both samples. Random effects are introduced in the detector simulation which might give rise to these effects, but a more thorough investigation if fast simulation overestimates the detectors ability to reconstruct $\tau_{had-vis}$ at the $H^{\pm\pm}$ masses is warranted based on these results. While any result of a comparison study of the fully simulated $m(H^{\pm\pm}) = 600$ GeV signal sample might be unsurprising, it can be argued that fast simulation still should be validated for τ-lepton signal samples.

The choice of selecting only for one dilepton in the final state for the signal sample histograms was made due to the lack of statistics in the 4-lepton final state. This was anticipated with for example the last SSDiLep same-sign analysis using the single dilepton channel as the primary signal channel. The 4-lepton final state requires all daughter leptons to be reconstructed correctly with their correct charge, which especially when including τ_{had} is not a certainty. The channel is also extra susceptible to noise from fake leptons who can replace the true lepton in one of the same-sign pairs. A similar study showing if the 4-lepton final state is a worse signal channel than a single same-sign dilepton must however be studied. To note is also that a reconstructed $4\tau_{had}$ final state is incredibly rare in both signal samples, a problem that will be discussed more in section 5.

The invariant mass distributions for the leading same-sign dilepton are expected, but the well defined resonance as seen in the collinear approximation invariant mass spectrum can be noted. One drawback of the collinear approximation is the loss of useful statistics as if the dilepton is created back to back ($\Delta\phi \approx \pm\pi$) the approximation breaks down and returns an overestimated

invariant mass. However, as can be seen in figure 18 few events are back-to-back making the distribution tail for the collinear approximation similar to the one seen for the total invariant mass, but with significantly less events at lower invariant mass. As a *discriminant*, a variable that separates signal from background, the collinear approximation might be thus be an interesting alternative to other variables. These signal samples are useful as a stepping stone in a study to find a good discriminant to use when including τ_{had} in a resonance search: While invariant mass is the classic choice other more advanced techniques such as the Missing Mass Calculator (MMC) for invisible resonance decays might be more appropriate [28].

Other discriminants for τ-inclusive same-sign dilepton searches that could be considered are trained networks such as a BDTs that can be trained to recognize same-sign dileptons from resonance decays using more variables than invariant mass, which might be appropriate when using τ_{had} due to the intrinsic increase of background from jets. One example would be Ref. [10] that looks at $Z \to l\tau_{had}$ events and uses a neutral network to output a score as discriminant. Input variables include p_T of the event particles, missing energy of the event, and also the collinear mass $m(l\tau_{had})$. Using trained networks allows for several discriminants to be used in tandem to improve the result, and also avoids the pitfall of of choosing a suboptimal discriminant by chance.

5 Update of event generation filter TauFilter

5.1 Purpose and motivation

When requesting and generating signal samples it is important to be precise in what events and particles should be included: Due to the computational resources required for signal sample generation no redundant events should have to be simulated. One way of doing this is to interface an event filter with the event generation process. Depending on its function, a filter will only allow the generation of certain events that fulfill the filter criteria, thus discarding unwanted events and increasing the number of events in the relevant channels.

ATLAS software already contains many different filters that can be used for event generation. However, due to the changing nature of the signals of interest these filters must sometimes be created or updated. As described in section 4 $H^{\pm\pm}$ signal samples with decays of the type $H^{\pm\pm} \rightarrow (e/\mu)^{\pm}(e/\mu)^{\pm}$ existed, but had to be extended to include $H^{\pm\pm}$ that decay into τ-leptons. To this end a filter was needed to only generate τ-inclusive events, as pure $H^{\pm\pm}H^{\mp\mp} \rightarrow 2(e/\mu)^{\pm}2(e/\mu)^{\pm}$ events from the previous signal sample could be reused.

One possible approach was to use a filter only allowing the generation of events with a certain number of τ_{had} or τ_{lep}. Of the current available ATLAS event generation filters there was none that could provide this. The existing TauFilter had a basic implementation of requiring a minimum number of both τ-leptons in total and τ_{lep}/τ_{had} separately. However, this could not be used to require a maximum number of either τ-lepton type, thus preventing a filtering scheme requiring for example at least one τ_{lep} but a veto on τ_{had}.

The purpose of this subproject was to update the TauFilter to be able to precisely choose the number and type of τ-leptons generated in an event sample. The updated TauFilter will be useful in future event generation both concerning the inclusion of τ-leptons in the SSDiLep group's new analysis and for new signal sample generation in general. One example of this would be to use it to generate signal samples of certain rare final states, such as $H^{\pm\pm}H^{\mp\mp} \rightarrow 2\tau^{\pm}_{had}2\tau^{\mp}_{had}$, where we see that the statistics of those channels are too low to be useful in an inclusive τ-lepton $H^{\pm\pm}$signal sample.

5.2 Method

To update the TauFilter the copy of the Athena filter package, containing all current ATLAS event generation filters, was created locally. The TauFilter was then changed and used in the *JobOption* files for the standard Athena job transform to produce events using Pythia. The effects of the filter settings were checked by generating truth level validation histograms in the same manner as the validation histograms of the τ-inclusive $H^{\pm\pm}$ signal sample were generated as described in section 4.3.

The logic of the TauFilter was changed to allow the user to specify both a maximum and a minimum number of and type of τ-leptons. The updated code can be found at Ref. [33], and the JIRA ticket for the requested filter update can be found at Ref. [5]. The filter was validated using the Athena 19.2.5.27 release together with Pythia version 8.212 and was included as part of the Athena 19.2.5.29 release. The validation histograms, as presented in this thesis, were created using the Athena 19.2.5.29 release.

It should be noted that the coding of the filter logic was a straightforward task. Most of the project time was spent in discussion on the appropriate approach, creating different sets of validation histograms and with the actual administrative process in getting the filter update accepted into the official ATLAS code.

5.3 Data sets

The event samples used to validate the updated `TauFilter` were generated the same way as the τ-inclusive $H^{\pm\pm}$ event samples as described in section 4.4. Three event samples with a final event number of $n = 5 \cdot 10^4$ with pair-production of $H^{\pm\pm}$ were generated, all with $m(H^{\pm\pm}) = 300$ GeV but with different filter options. The filter schemes were chosen to allow the generated samples to be merged in order to form a complete event sample with all possible τ-lepton final states from $H^{\pm\pm}$ pair production. An additional sample with $n_{4\tau} = 10^4$, events meant to contain events with only the decay $H^{\pm\pm} \to 2\tau^{\pm}_{had}$, was also created with the purpose of examining if this filter could efficiently be used to isolate such a specific final state. The event samples are summarized in table 7.

Some of the generated events contained τ-leptons from other sources than $H^{\pm\pm}$ decay and such events could pass the `TauFilter` even if the prompt final state leptons only included light leptons. Such so-called *filter leaks* imply that some events that should have been filtered might still be present in the generated event sample. The filter leak percentage ε_{leaks} as presented in table 7 is the ratio of all events with a leak to the total number of events. Some of these events might still contain valid final states and the presented ε_{leaks} is thus an upper estimate of the number of events lost due to filter leaks.

Table 7: PYTHIA pair production of $H^{\pm\pm}$ MC event samples with different `TauFilter` settings. ε_{leaks} is the percentage of final events that passed the filter while containing at least one τ-lepton from a source other than a $H^{\pm\pm}$.

`TauFilter` settings	ε_{filter}[%]	$n_{event}[10^3]$	ε_{leaks}[%]
$n_{\tau_{had}} = 0,\ n_{\tau_{lep}} \geq 1$	17.86 ± 0.04	50	0.76 ± 0.04
$n_{\tau_{had}} \geq 1,\ n_{\tau_{lep}} = 0$	41.27 ± 0.12	50	1.07 ± 0.05
$n_{\tau_{had}} \geq 1,\ n_{\tau_{lep}} \geq 1$	21.43 ± 0.04	50	2.24 ± 0.07
	$\sum_{filter} = 80.56 \pm 0.13$		
$n_{\tau_{had}} \geq 4,\ n_{\tau_{lep}} = 0$	0.2392 ± 0.0001	10	9.75 ± 0.31

The numbers presented for filter efficiencies, $\varepsilon_{filter} = \frac{n_{event}}{n_{generated}}$, and filter leaks, $\varepsilon_{leaks} = \frac{n_{leaks}}{n_{event}}$, are taken from the PYTHIA output. Their associated errors σ are estimated assuming that the event generation events and filter leaks are generated from a Poissonian distribution, given that the

standard deviation of an observed count of n independent events is given by $\sigma = \sqrt{n}$. The errors were therefore calculated according to

$$\sigma_{filter} = \varepsilon_{filter} \cdot \frac{\sigma_{generated}}{n_{generated}} = \frac{\varepsilon_{filter}}{\sqrt{n_{generated}}},$$

$$\sigma_{leaks} = \frac{\varepsilon_{leaks}}{\sqrt{n_{leaks}}}$$

and is presented in the form $\varepsilon \pm \sigma_\varepsilon$.

5.4 Result

Validation histograms of the distribution of $H^{\pm\pm}H^{\mp\mp} \rightarrow l^{\pm}l^{\pm}l^{\mp}l^{\mp}$ 4-lepton final state and the $H^{\pm\pm} \rightarrow l^{\pm}l^{\pm}$ decay modes are presented for three different `TauFilter` configurations. The decay modes include $H^{\pm\pm} \rightarrow W^{\pm}W^{\pm}$ for validation purposes, as any such decay would indicate that the JobOptions file was wrongly defined.

The general trend of more events in the final states favoring e over μ in the filter settings requiring at least one τ_{lep} can be attributed to the slight favor of $BR(\tau_e)$ over $BR(\tau_\mu)$ as seen in table 2. τ_{lep} is τ_e with a probability of $P(\tau_{lep} = \tau_e) \approx 50.6\%$, and the required presence of at least one τ_{lep} in every event will thus imply a global excess of e over μ in the final state.

Events with filter leaks have been excluded from the validation plots to show that the correct final state distribution has been generated. The exception is for the $n_{\tau_{had}} \geq 4$, $n_{\tau_{lep}} = 0$ case where all generated events will be included as the number of true $4\tau_{had}$ final states that were generated with that particular filter setting is of interest.

At least one τ with no τ_{had}

The 4-lepton final state histogram for the filter setting $n_{\tau_{had}} = 0$, $n_{\tau_{lep}} \geq 1$ is presented in figure 22 and the $H^{\pm\pm}$ decay modes is presented in figure 23. As expected from a τ_{had} veto only the light lepton states are present, and the distributions are similar to what is expected of a pure light lepton event sample.

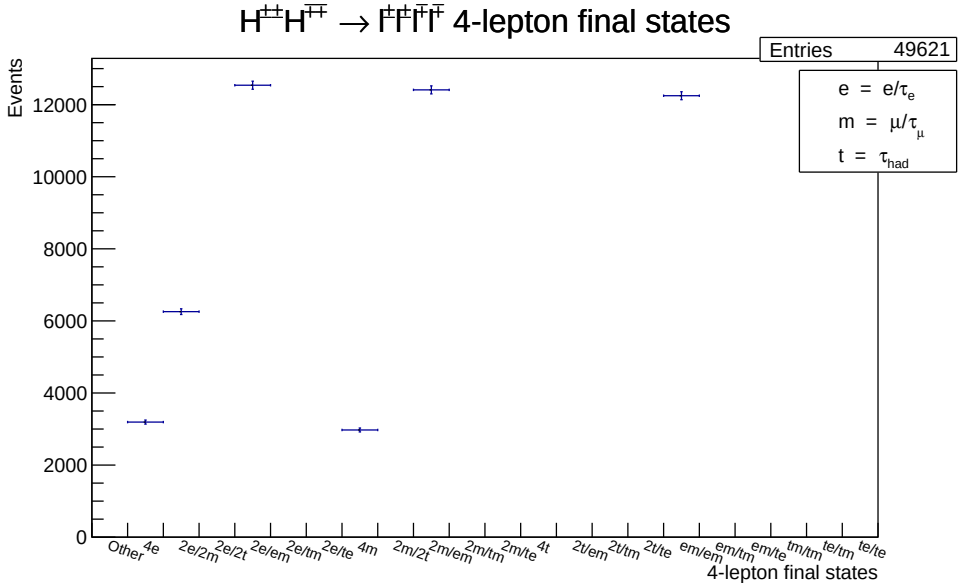


Figure 22: 4-lepton final state distribution for the filter requirements $n_{\tau_{had}} = 0$, $n_{\tau_{lep}} \geq 1$.

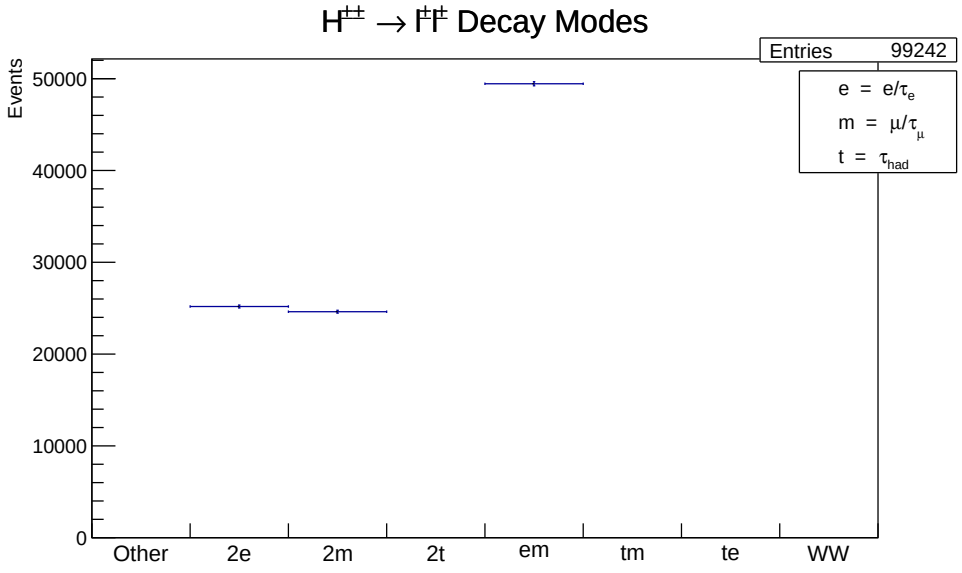


Figure 23: $H^{\pm\pm}$ decay modes for the filter requirements $n_{\tau_{had}} = 0$, $n_{\tau_{lep}} \geq 1$. WW decay channel displayed to validate that is is not generated.

At least one τ with no τ_{lep}

The 4-lepton final state histogram for the filter setting $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} = 0$ are presented in figure 24 and the $H^{\pm\pm}$ decay modes are presented in figure 25. As expected, no 4-light lepton final state is present.

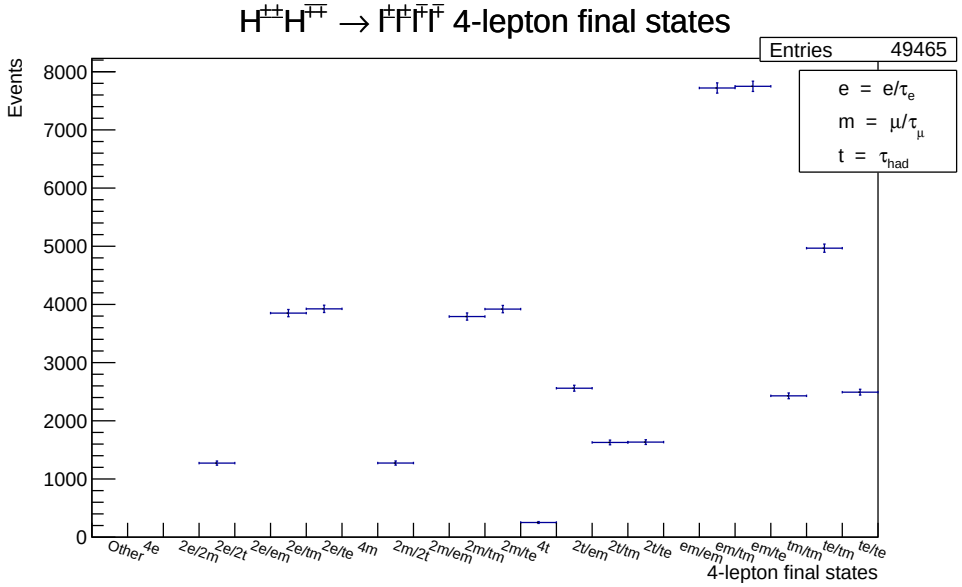


Figure 24: 4-lepton final state distribution for the filter requirements $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} = 0$.

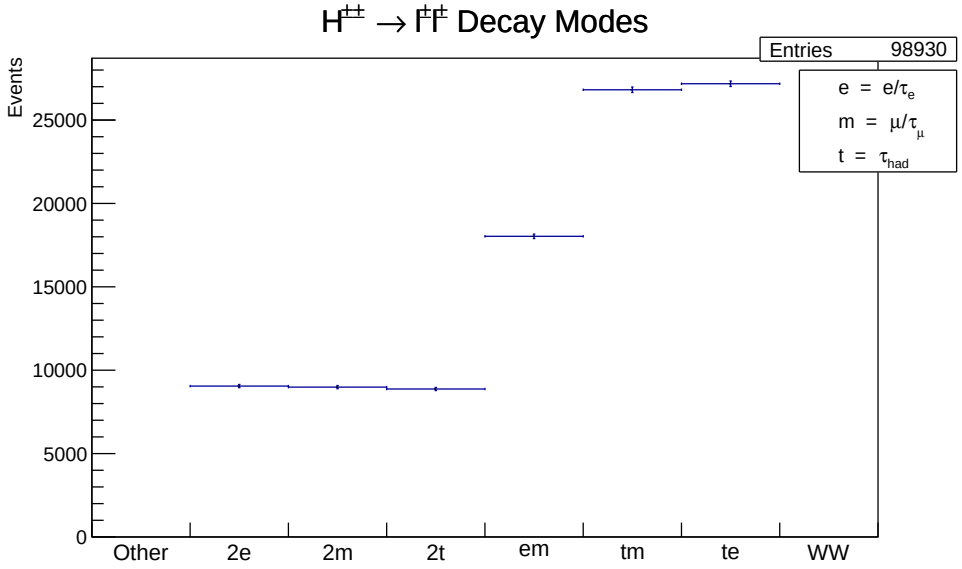


Figure 25: $H^{\pm\pm}$ decay modes for the filter requirements $n_{\tau_{had}} \geq 0$, $n_{\tau_{lep}} = 1$.

At least one τ_{had} and one τ_{lep}

The 4-lepton final state histogram for the filter setting $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} \geq 1$ is presented in figure 26 and the corresponding $H^{\pm\pm}$ decay modes are presented in figure 27. No pure 4-light lepton final state is present, and neither is the $4\tau_{had}$ channel.

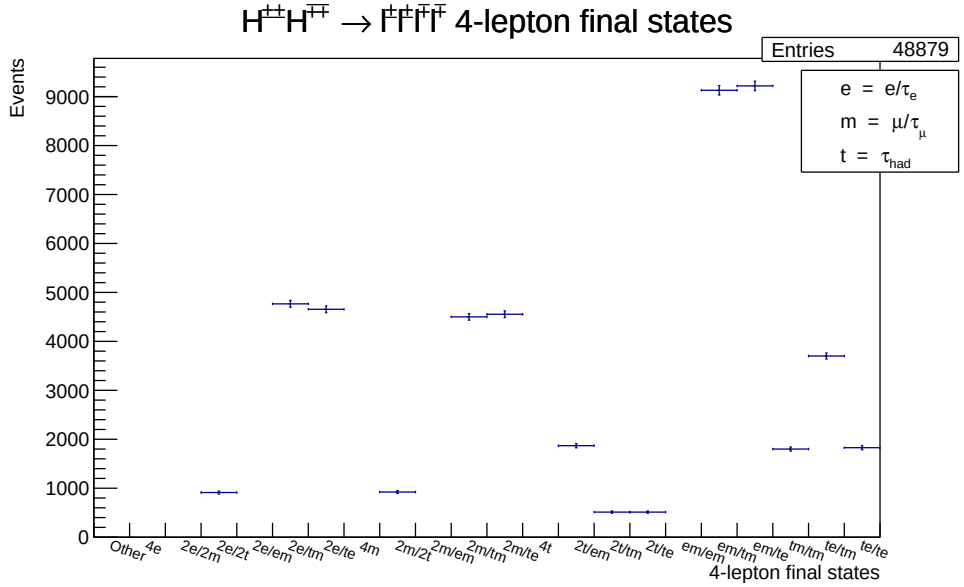


Figure 26: 4-lepton final state distribution for the filter requirements $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} \geq 1$.

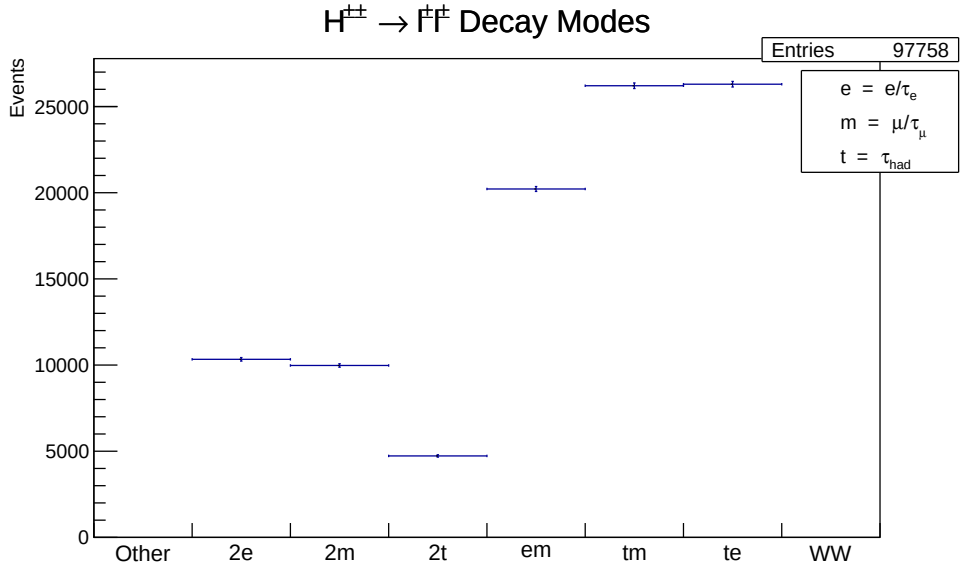


Figure 27: $H^{\pm\pm}$ decay modes for the filter requirements $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} \geq 1$.

At least four τ_{had} with no τ_{lep}

The 4-lepton final state histogram for the filter setting $n_{\tau_{had}} \geq 4$, $n_{\tau_{lep}} = 0$ is presented in figure 28 and the corresponding $H^{\pm\pm}$ decay modes are presented in figure 29. All events generated are included in the validation histograms, including the ones with filter leaks, to show explicitly the effects of the leaks on the final state distribution. The number of generated events with a $4\tau_{had}$

final state is $n_{4\tau_{had}} = 9090$, meaning that only $\frac{n_{4\tau_{had}}}{n_{events}} = 90.9\%$ of the generated events resulted in the intended final state.

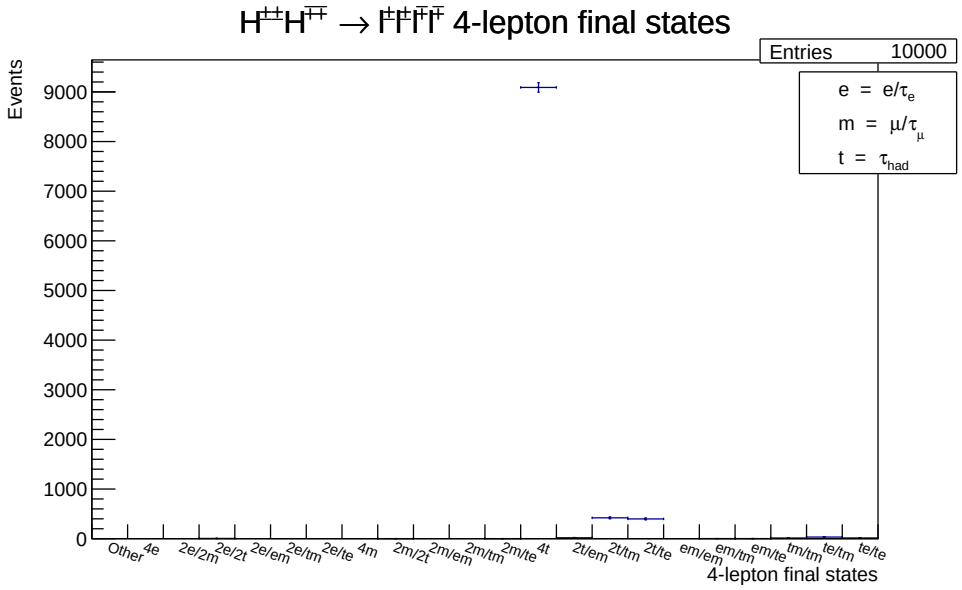


Figure 28: 4-lepton final state distribution for the filter requirements $n_{\tau_{had}} \geq 4$, $n_{\tau_{lep}} = 0$.

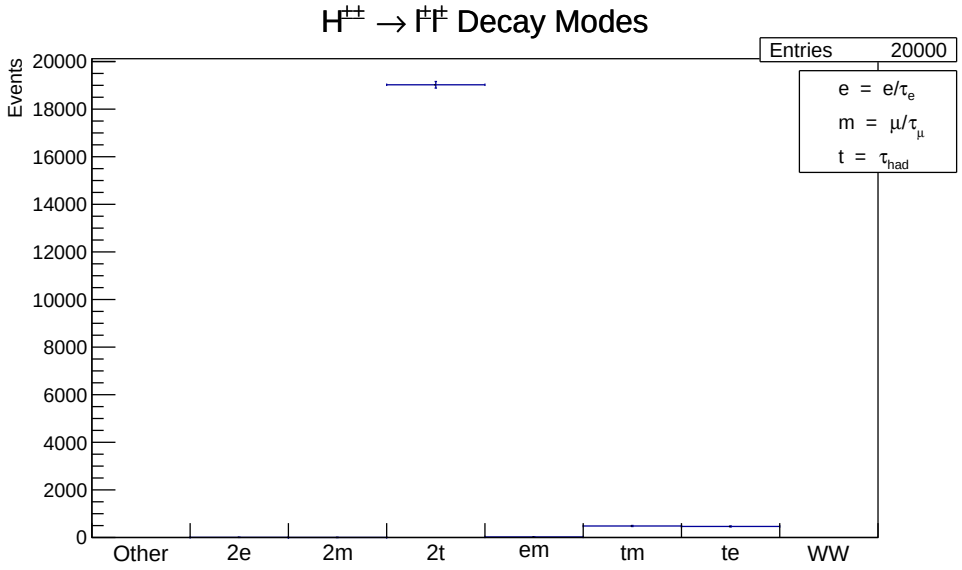


Figure 29: $H^{\pm\pm}$ decay modes for the filter requirements $n_{\tau_{had}} \geq 4$, $n_{\tau_{lep}} = 0$.

Filter leaks parent particles

To find the source of the event leaks the non-$H^{\pm\pm}$ parent particles to τ-leptons was plotted. Most of the parent particles were mesons created in the PYTHIA hadron shower and the particle type distribution was similar regardless of filter settings used. The non-$H^{\pm\pm}$ parent particles for $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} = 0$ is presented in figure 30, with the distributions for the other filter settings being similar.

While the $D^{\pm}$ is the lightest meson that can decay into a τ-lepton the branching ration of weak decay into a τ-lepton is small. The distribution is instead heavily dominated by the $D_s^{\pm}$ whose only relevant leptonic decay mode is by the τ. Due to the B meson mass τ-leptons can be both produced by pair production with the decay of the neutral B mesons or weak decay of $B^{\pm}$. Other modes are γ pair production and decay from several types of rarer hadrons.

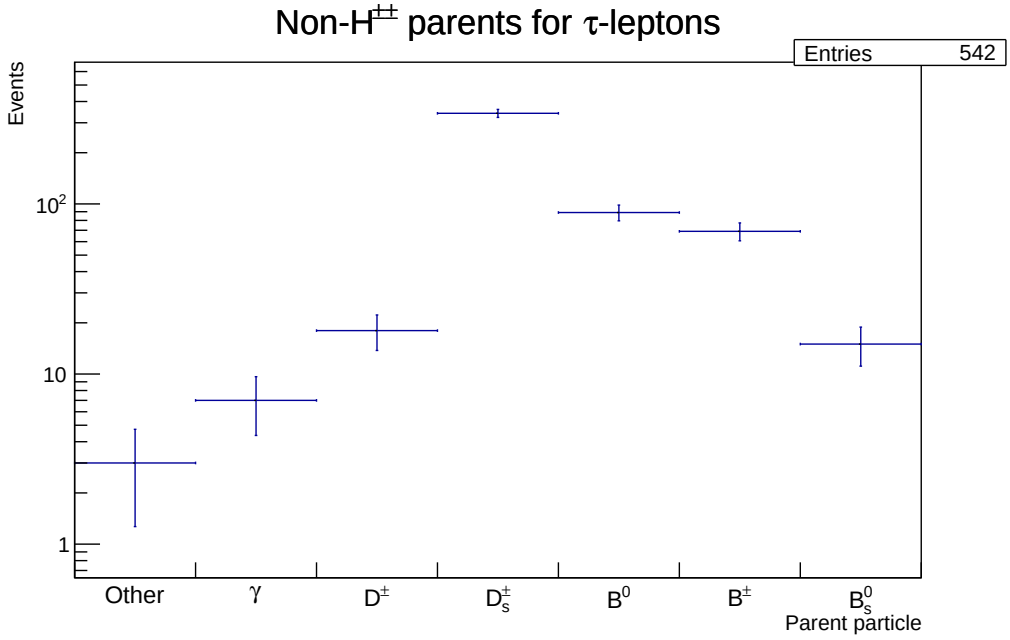


Figure 30: Number of non-$H^{\pm\pm}$ parent particles for τ-leptons in the event sample for the filter requirements $n_{\tau_{had}} \geq 1$, $n_{\tau_{lep}} = 0$. *Other* include more exotic hadrons such as the Λ_B^0 baryon.

5.5 Discussion

Based on the validation plots of the event samples with different `TauFilter` settings the `TauFilter` update seems to have succeeded in its goal to allow more specific τ-lepton final states to be generated within the ATLAS project.

The updated `TauFilter` does not contain any way to impose parentage on the τ-leptons. While for general use such a function might not be needed, adding it would be straightforward: Several other filters already implement parentage requirement, and the functions can with few changes be added to the `TauFilter` to prevent the loss of statistics due to the generation of non-prompt τ-leptons. This could be relevant when generating for example specialized τ-inclusive $H^{\pm\pm}$ signal

samples with certain requirements on the number of τ-leptons. A direct example would be a future need for a pure $2H^{\pm\pm} \to 4\tau_{had}$ signal sample: Using the filter as it is a single non-$H^{\pm\pm}$ τ_{had} would count towards the filter requirements, indicating that many unwanted $3\tau_{had} + l$ final states might be generated.

This last effect can be clearly seen in figure 28 for the $4\tau_{had}$ event sample with the $3\tau_{had} + l$ final states being the principal channels for the filter leak events. The effect of having no parentage condition results in 9.1% of the total generated events to be of the wrong kind due to filter leaks. While in principle this does not prevent the filter from being used, the addition of a possible parentage condition of the τ-lepton would immediately add valuable statistics to the final signal sample. Alternative event generation approaches, for example where the coupling to other leptons is set to 0, would also mitigate this problem considerably.

The three filter settings with $(n_{\tau_{had}} = 0,\ n_{\tau_{lep}} \geq 1)$, $(n_{\tau_{had}} \geq 0,\ n_{\tau_{lep}} = 0)$ and $(n_{\tau_{had}} = 1,\ n_{\tau_{lep}} \geq 1)$ cover all possible final states with at least one τ-lepton present. A sanity check can be performed to see if the filters behave as they should. Together with the pure light lepton final state the filter efficiencies should add up to 100%, as a pure $2H^{\pm\pm} \to 4l$ event generation would generate all possible states and sort out none. There are 16 possible light lepton final states and 81 possible combinations of all lepton final states: The theoretical filter efficiency of a filter only selecting light lepton final states would thus be $\varepsilon_{\mu/e}^{theory} = \frac{16}{81} \approx 19.8\%$. This check gives $\sum_{filter} + \varepsilon_{theory}^{\mu/e} \approx 100.31 \pm 0.13\%$, which implies an overestimation in the filter efficiencies as output by PYTHIA. This is expected as the background τ-leptons will make the filter accept events that should be rejected, thus inflating the filter efficiency. If the filter leak events are not counted when calculating the filter efficiencies the same check gives $\sum_{filter,no\,leaks} + \varepsilon_{theory}^{\mu/e} \approx 99.26 \pm 0.13\%$, which shows an analogous but reversed effect: As the filter will reject an event if it contains the wrong combination of τ-leptons regardless of their source the effective filter efficiencies will be slightly deflated due to τ-leptons from the generated particle shower.

Several additions and changes to the `TauFilter` were done with the purpose of increasing the clarity and efficiency of the filter. This includes extensive commenting, renaming of variables to clearly reflect their purpose and a cleanup of inefficient or redundant logic. The rewriting could have been taken further but was abandoned due to both time constraint and the risk of inadvertently breaking previous filter functionality. In the context of this subproject there is a strong case of rewriting the filter code to untangle the logic and simplify future development of the `TauFilter` code. Apart from adding a parentage condition for each τ-leptons another functionality that might be useful could be a selection of prongedness.

6 Summary and outlook

This thesis work has entailed creating new τ-inclusive $H^{\pm\pm}$ signal samples and updating a τ-lepton event generation filter, and the resulting signal sample xAODs and the update of the `TauFilter` will hopefully be useful for the current and future same-sign dilepton analyses. The methods used have been described with the purpose of enabling someone new to the ATLAS collaboration to get an insight into some of the parts behind a particle physics analysis. In terms of analysis signal samples the created sample only serve as a first step, with a minimum of more mass points needed if the theoretical $H^{\pm\pm}$ model chosen for the signal samples is kept for the next analysis. The many different charge scalar models in existence do prompt the question of whether the simple $H^{\pm\pm}$model used in the context of this thesis is the most interesting one.

During the course of this thesis work the SSDiLep group has transitioned into a more general same-sign cluster, with the aim of unifying the same-sign lepton analyses in the ATLAS Exotics subgroup in terms of method such as lepton selection, fake leptons and charge flip estimation. In the regard of furthering the same-sign clusters work this thesis has opened up the possibility for several follow-up projects ranging in scope from small studies to topics suitable as a thesis topic. These include:

- Validation of the full range of the generated signal samples compared to the previously existing light lepton $H^{\pm\pm}$ signal samples.
- A comparison study between the full and fast simulation signal samples to examining the difference, if any, in τ-lepton reconstruction and kinematic.
- Study of τ-lepton mass reconstruction methods to identify a useful discriminant variable when searching for resonances resulting in semi-invisible daughter particles.
- Developing the `TauFilter` to reduce or remove event generation leaks when generating τ-inclusive samples.

Projects that also should be of interest in the wider scope of same-sign dilepton analysis and which has been touched upon in this thesis work include conducting a basic τ-lepton charge flip estimation using truth analysis, and continuation of evaluating the use of a data-driven charge flip estimation method, either of a tag-and-probe variant or the more complex likelihood estimation method as explored in appendix F.3. Another interesting possibility would be investigating trained networks as discriminant for future τ-inclusive searches to better handle the hadronic background in all τ_{had} channels.

Much work remains to be done in developing the methods for including τ-leptons into the SSDiLep analysis, and this thesis has only provided a small step along the way. However, access to improved statistics and the rich set of models that can be probed promise an interesting future for analysis of same-sign τ-inclusive dileptons at the ATLAS detector.

Acknowledgments

I would like to thank my supervisor Else Lytken for guiding me through this maze of a masters project, and for infinite patience and understanding of all the wrong turns I have taken during the way.

In the same spirit I would like to thank Katja Mankinen without whose help this thesis would never have been written. It has been an honor to have worked with the best PhD student in Lund (possibly the world), and you are and will continue to be a source of inspiration!

Also a big shout out to the ALTAS masters room, and especially Dimitris Sidiropoulos, Yosse Andrean and Zhiying Li, for a singular office experience.

A Definition of particle physics terms

Common particle physics terms and how they will be used in this thesis will be presented below. This thesis uses natural units with $c = \hbar = 1$, meaning mass, momentum, and energy will be denoted by the energy unit eV.

Primary and secondary interaction vertices The *primary interaction vertex* is considered the origin of a detector event and corresponds to the collision point of two protons of the LHC. A *secondary vertex* is the result of a particle from the primary vertex decaying somewhere in the tracking detector, thus creating a new origin point for its decay products. The standard example of a secondary vertex would be from mesons containing bottom quarks, as their relative longevity allows them to travel an observable distance from the primary vertex before decaying.

Prompt particle A *prompt particle* will refer to a particle that originates in the primary interaction vertex of the detector. The term is often used to differentiate between particles of interest in the analysis and so-called non-prompt particles such as particles created in the hadron shower, particles created within the detector material, cosmic muons and so on. All particles that are observed to come from the primary interaction vertex will be considered prompt. This includes the decay products from τ-leptons, as the ATLAS detector currently cannot assign a secondary vertex to a τ-lepton decay.

Final state A *final state* will refer to the collection of prompt particles that can be expected to be observed in the detector after a certain process has taken place. A specific final state that is used for an analysis is often interchangeably referred to as a *channel*. An example relevant in the context of this thesis is a *4-lepton final state*, meaning a process that would give rise to exactly four prompt leptons that would be observable by the detector.

Lepton and light lepton The term *lepton*, denoted l or lep, will refer to one of the charged leptons e, μ or τ. The term *light lepton* will be reserved for e and μ, whose mass does not allow any hadronic decay modes.

Jet A hadronic *jet* will refer to an object observed in particle detector consisting of a number of hadrons traveling together in a narrow cone. As an object reconstructed by the detector what is considered a jet change depending on which jet reconstruction algorithm that is used, with the most common one being the anti-k_t algorithm. They arise due to the color confinement of the strong force: If color charges are separated as part of a particle interaction, more color charged objects will be created in a process called *hadronization* to minimize the potential energy of the color field and keep each colored object colorless. The result is a shower of hadrons with a common origin and similar path, and is the signature expected from all interactions involving quarks in the final state. Due to the hadronic nature of $p - p$ collisions jets arising from the initial collision is a common background in the ATLAS detector.

Transverse momentum (p_T) The plane perpendicular to the particle beam direction of a particle detector is called the transverse plane, and the momentum of a particle in that plane is called the *transverse momentum*, denoted p_T. As the colliding particles are traveling perpendicularly to the transverse plane any momentum detected in the transverse plane must have been created during the particle collision, and must therefore sum up to zero due to conservation of momentum. If the transverse momentum of an event does not sum up to zero the missing part is called the *missing transverse momentum*, denoted p_T^{miss}. While detector limitations always give rise to some missing momentum, a large amount of missing transverse momentum is often indicative of undetectable particles such as neutrinos being part of an event.

Pseudorapidity (η) The *pseudorapidity*, denoted by the Greek letter η, is a measure of the angle of the observed particles relative to the beam direction. The observable η of particles is limited by the detector to$|\ \eta\ | \leq 2.5$ due to the coverage of the inner tracking detector. It is further limited in the so-called crack region $|\ \eta\ | \sim 1.5$ where the barrel calorimeter and the end-cap calorimeter meet. It is defined as $\eta = -\ln\left(\tan(\frac{\phi}{2})\right)$, where ϕ is the angle between the particle and the positive beam direction. [25]

Invariant mass *Invariant mass* is a constant of a physical system regardless of frame of reference and can be calculated if the energy and trajectory of the relevant particles of the system is known. The invariant mass of particle p_1 and p_2 will be denoted as $m(p_1 p_2)$, and is a common quantity to calculate as the invariant mass of the decay products from a resonance is equal to the resonance mass.

Monte Carlo simulation *Monte Carlo simulation* will refer to the use of a stochastic method to simulate a physical system. Monte Carlo simulation assigns probability distributions to all processes of a system and then uses these distributions to generate one outcome, which for example could be the detector readout of a $p-p$ collision in ATLAS. By repeating this procedure many times the behavior of a physical system is simulated. In particle physics the most important use of Monte Carlo simulation is to simulate particle collisions using the interaction probabilities of the Standard Model, and then compare the simulation output with the observed data.[4]

Event generation sample A dataset containing the output of a Monte Carlo simulation of particle interaction will be referred to as an *event generation sample*. In the context of ATLAS events generation samples contain simulations of $p-p$ collision events and contain event information such as the generated particles and their trajectories. There exist several specialized software packages called event generators (which the most common ones being Pythia, Sherpa and Herwig) that produces the event generation samples.

Signal sample A *signal sample* will refer to an event generation sample having been run through a detector simulation, giving as output the expected signal of an event if it would have happened

[4]It can be noted that as quantum mechanics, and thus by extension the Standard Model, is a stochastic physics model any simulation of a quantum mechanical systems is by this definition a Monte Carlo simulation.

in a particle detector. It contains information about for example reconstructed particle paths and calorimeter readouts. Signal samples also contain so-called truth information detailing the event generation information of the reconstructed particles of the signal sample.

Leading order (LO) and next-to-leading order (NLO) *Leading order* will refer to the interaction contributing the most to a particular particle physics process. The terms are often used in tandem with Feynman diagrams, where a LO diagram of a process contains no propagator loops. *Next-to-leading order* diagrams, in contrast, include loop terms, and thus modifies the total process probability if they are chosen to be included. Leading order interactions are also referred to as being on *tree level*.

B Tools and software used by ATLAS

The ATLAS experiment makes use of several software and tools for data management and physics analysis. The ones most commonly used during this thesis work will be summarized below.

ROOT A data analysis framework based on C++ commonly used for analysis and visualization of particle physics data. It both provides an extensive package of different statistics and histogram tools and a compiler and interpreter for C++ code. The end result of the ATLAS data flow is generally a ROOT-readable ntuple [21].

Athena A general purpose framework used by ATLAS for simulations, particle reconstruction, and physics analyses. It uses a common architecture to call and use several different physics tools and software. Generally, ATHENA takes a so-called *JobOptions* Python file specifying the relevant algorithms and tools needed for a task and executes it. ATHENA can be used to read most ATLAS data formats and overlaps with ROOT in this regard [36].

xAODAnaHelper A Python framework meant to process DxAOD files into analysis-ready ROOT ntuples. It implements several tools to correctly select and decorate the wanted particles, and can be used to perform an initial preselection of the analysis objects [17].

JIRA A project management tool used by ATLAS to keep track of software update requests, bug fixes, and other software-related issues.

Grid software

The grid refers to the network of computing centers accessible by the ATLAS experiment. A more detailed summary of grid jobs and software can be found in Ref. [45].

PanDA A software client that contains the different tools for submitting jobs to the grid, which for example includes the `pathena` command which submits ATHENA jobs to the grid instead of running locally. Grid jobs submitted through panda can be observed through the bigPanDA website [9, 37].

Rucio A software client that enables data sets located on the grid to be located and downloaded, and is used in tandem with PANDA to retrieve the output of jobs submitted to the grid [30].

C The collinear approximation for τ-lepton mass reconstruction

The following treatment is based on Ref. [35] but extended to show the explicit solution steps.

Due to τ-lepton decay giving rise to at least one neutrino, an event with pair production of τ-leptons will contain many sources of missing momentum $\boldsymbol{p}_{miss}$. This makes it impossible to calculate the invariant energy of the whole event, a useful quantity for finder for example a resonance (such as a $H^{\pm\pm}$) that decays into two τ-leptons. A common way to estimate the true invariant mass of a 2τ system (or any system where one or both of the daughters of a resonance are semi-invisible) is the so-called *collinear approximation*. It is based on two assumptions: All created neutrinos are collinear with the visibleτ decay products, and that the missing momentum of the event only comes from the neutrinos and no other source.

The momenta of the visible decay products arising from the 2τ-lepton decays will acquire some fraction of the total τ-lepton momenta $\boldsymbol{p}_{vis,i}$ for $i = 1, 2$. The missing momenta from the neutrinos in each decay, $\boldsymbol{p}_{miss,i}$, can be expressed in terms of the original τ-lepton momenta $\boldsymbol{p}_{\tau_i}$ and $\boldsymbol{p}_{vis,i}$ as

$$\boldsymbol{p}_{miss,i} = \boldsymbol{p}_{\tau_i} - \boldsymbol{p}_{vis,i} \approx F_i \boldsymbol{p}_{vis,i} \tag{1}$$

where the last step uses the assumptions that $\boldsymbol{p}_{miss,i}$ is collinear with $\boldsymbol{p}_{\tau_i}$ and thus can be expressed in terms of $\boldsymbol{p}_{vis,i}$ scaled with some factor F_i. In the transverse plane the total missing transverse momentum $\boldsymbol{p}_{T,miss}$ can be expressed as a sum of the missing momentum arising from each separate τ-lepton decay according to

$$\boldsymbol{p}_{T,miss} = \boldsymbol{p}_{T,miss,1} + \boldsymbol{p}_{T,miss,2} = F_1 \boldsymbol{p}_{T,vis,1} + F_2 \boldsymbol{p}_{T,vis,2}.$$

This gives the system of equation

$$\begin{cases} p_{Tx,miss} = F_1 p_{Tx,1} + F_2 p_{Tx,2} & (a) \\ p_{Ty,miss} = F_1 p_{Ty,1} + F_2 p_{Ty,2} & (b) \end{cases}.$$

This system has two unknowns F_1 and F_2 and can thus be solved for these as follows:

$$(a): \; p_{Tx,miss} = F_1 p_{Tx,1} + F_2 p_{Tx,2} \Rightarrow F_1 = \frac{p_{Tx,miss} - F_2 p_{Tx,2}}{p_{Tx,1}} \; (c)$$

$$(c) \rightarrow (b): \; p_{Ty,miss} = \left(\frac{p_{Tx,miss} - F_2 p_{Tx,2}}{p_{Tx,1}}\right) \cdot p_{Ty,1} + F_2 p_{Ty,2}$$

$$\Leftrightarrow p_{Ty,miss} p_{Tx,1} = p_{Tx,miss} p_{Ty,1} - F_2 p_{Tx,2} p_{Ty,1} + F_2 p_{Ty,2} p_{Tx,1}$$

$$\Leftrightarrow p_{Ty,miss} p_{Tx,1} - p_{Tx,miss} p_{Ty,1} = F_2 (p_{Ty,2} p_{Tx,1} - p_{Tx,2} p_{Ty,1})$$

$$\Leftrightarrow F_2 = \frac{p_{Ty,miss} p_{Tx,1} - p_{Tx,miss} p_{Ty,1}}{p_{Ty,2} p_{Tx,1} - p_{Tx,2} p_{Ty,1}} \; (d)$$

$$(d) \rightarrow (c): \; F_1 = \frac{p_{Tx,miss} - (\frac{p_{Ty,miss}p_{Tx,1} - p_{Tx,miss}p_{Ty,1}}{p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1}})p_{Tx,2}}{p_{Tx,1}}$$

$$\Leftrightarrow F_1 = \frac{p_{Tx,miss}p_{Ty,2}p_{Tx,1} - p_{Tx,miss}p_{Tx,2}p_{Ty,1} - p_{Tx,2}p_{Ty,miss}p_{Tx,1} + p_{Tx,2}p_{Tx,miss}p_{Ty,1}}{p_{Tx,1}(p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1})}$$

$$\Leftrightarrow F_1 = \frac{p_{Tx,miss}p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,miss}p_{Tx,1}}{p_{Tx,1}(p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1})}$$

$$\Leftrightarrow F_1 = \frac{p_{Tx,miss}p_{Ty,2} - p_{Tx,2}p_{Ty,miss}}{p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1}} \; (e)$$

As $F_{1,2}$ are by definition positive constants their final form is

$$\begin{cases} F_1 = |\frac{p_{Tx,miss}p_{Ty,2} - p_{Tx,2}p_{Ty,miss}}{p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1}}| \\ F_2 = |\frac{p_{Ty,miss}p_{Tx,1} - p_{Tx,miss}p_{Ty,1}}{p_{Ty,2}p_{Tx,1} - p_{Tx,2}p_{Ty,1}}| \end{cases} . \tag{2}$$

Disregarding rest mass of the particles, the invariant transverse mass of the 2τ system is

$$m_T(\tau_1\tau_2) = \sqrt{2(p_{T,\tau_1} \cdot p_{T,\tau_2} - \boldsymbol{p}_{T,\tau_1} \cdot \boldsymbol{p}_{T,\tau_2})}.$$

Using equation (1) the transverse momentum of the τ-leptons can be expressed in terms of $\boldsymbol{p}_{T,vis,i}$ and F_i according to

$$\boldsymbol{p}_{T,\tau_i} = \boldsymbol{p}_{T,vis,i}(F_i + 1).$$

The collinear transverse invariant mass of the 2τ system $m_{T,col}(\tau_1\tau_2)$ can thus be expressed as

$$m_{T,col}(\tau_1\tau_2) = \sqrt{2(p_{T,\tau_1} \cdot p_{T,\tau_2} - \boldsymbol{p}_{T,\tau_1} \cdot \boldsymbol{p}_{T,\tau_2})} = \sqrt{2((F_1+1)p_{T,vis,1} \cdot (F_2+1)p_{T,vis,2} - (F_1+1)\boldsymbol{p}_{T,vis,1} \cdot (F_2+1)\boldsymbol{p}_T}$$

$$\Leftrightarrow m_{T,col}(\tau_1\tau_2) = \sqrt{(F_1+1)(F_2+1)}m_{T,vis}(\tau_1\tau_2)$$

where in the last step the definition of the invariant visible transverse mass $m_{T,vis}(\tau_1\tau_2) = \sqrt{2(p_{T,vis,2} \cdot p_{T,vis,2} - \boldsymbol{p}_{T,vis,2} \cdot \boldsymbol{p}_{T,vis,2})}$ was used. As $m_{T,vis}(\tau_1\tau_2)$ and all variables in equation (2) can be calculated from detector data the collinear transverse invariant mass makes it possible to calculate the true transverse invariant mass of the 2τ-system.

One constraint of the collinear approximation is that its calculated mass diverges when the visible τ-lepton decay products are back-to-back in the transverse plane, as can be seen in equation (2). This implies that the approximation is only valid for boosted systems, and will overestimate the invariant mass for non-boosted systems. The collinear approximation thus loses statistics from not being handle back-to-back events, which is one of the principal drawbacks of the method.

D Signal sample job options file

Listing 1: The Pythia8 job options Python file for generating the Tau-inclusive DCH signal sample with m(DCH)=300 GeV.

```
m_dch = 300.0

evgenConfig.description = "Doubly charged higgs ("+str(m_dch)+") in lepton mode."
evgenConfig.process = "DCH -> same sign 2lepton"
evgenConfig.keywords = ["BSM", "chargedHiggs" ,"2lepton"]
evgenConfig.contact = ["Simon Erland Arnling Baath <simon.erland.arnling.baath@cern.ch>
  "]
include("MC15JobOptions/Pythia8_A14_NNPDF23LO_EvtGen_Common.py")
genSeq.Pythia8.Commands += [
"9900041:m0 = " + str(m_dch), # H++_L mass [GeV]
"LeftRightSymmmetry:ffbar2HLHL=on", #HL pair production
#set the VEV (vacuum expectation value) value
"LeftRightSymmmetry:vL=0.0",
# set all couplings to leptons to 0.02
"LeftRightSymmmetry:coupHee=0.02",
"LeftRightSymmmetry:coupHmue=0.02",
"LeftRightSymmmetry:coupHmumu=0.02",
"LeftRightSymmmetry:coupHtaue=0.02",
"LeftRightSymmmetry:coupHtaumu=0.02",
"LeftRightSymmmetry:coupHtautau=0.02"]

from GeneratorFilters.GeneratorFiltersConf import ParentChildFilter #Removes generation
   of all pure light-lepton events
filtSeq += ParentChildFilter("ParentChildFilter")
filtSeq.ParentChildFilter.PDGParent = [9900041, -9900041]  #Set DCH as valid parents
filtSeq.ParentChildFilter.PDGChild = [15, -15] #Require at least one Tau-lepton as a
  child to a DCH in the event
```

E Combinatorics of the τ-inclusive signal sample

In the simple model used for the τ-inclusive signal samples a $H^{\pm\pm}$ couples only to leptons and equally strong to all three lepton generations. If its decay is lepton number violating the total number of combinations of final state leptons from a $H^{\pm\pm}$ decay then becomes $3^2 = 9$. By extension, pair production of $H^{\pm\pm}$ will result in a 4-lepton final state with $3^4 = 81$ different possible combinations, where each combination has a probability of $\frac{1}{81} \approx 1.2\%$ to occur. However, as many of these states are indistinguishable from each other the total amount of unique 4-lepton final states add up to 21, as can be confirmed by tedious if simple counting. These final states are shown in figure 31 which shows the distribution of events between the states with the additional requirement that the event contains at least one τ-lepton. The last requirement removes all combinations containing pure light leptons, which by the same method consists of $2^4 = 16$ combinations, resulting in a possible of $81 - 16 = 65$ possible 4-lepton final state combinations.

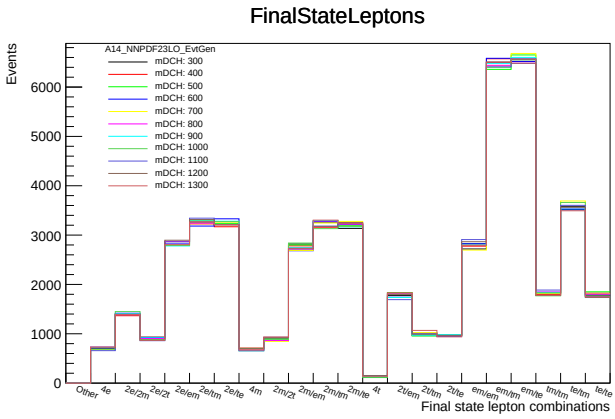


Figure 31: 4-lepton final state, with filter requirement $n_\tau \geq 1$. The four lepton daughters of the pair-produced $H^{\pm\pm}$, with e $= e/\tau_e$, m $= \mu/\tau_\mu$ and t $= \tau_{had}$.

The expected number of events in each of these final states can be easily calculated, and an example of the procedure will now be presented. The possible final states that can give rise to the 4-lepton final state $4e$ is presented in table 8. The procedure to calculated possible indistinguishable combinations is the following: In the case of the final state $2e/e\tau_e$ the combinations $e\tau_e/2e$, $2e/\tau_e e$ and $\tau_e e/2e$ all give the same observable final state. This gives it a multiplicity of 4, and the same check can be done for the other final states. The $4e$ final state corresponds to 15 out of 65 possible combinations, assuming the τ-leptons decay into electrons, which they do with a branching ratio $BR_{\tau_e} \approx 17.8\%$. To find the expected number of of $4e$ states you then need to weight every combination with the branching ratio of the τ to the particular final state.

Table 8: Possible lepton combinations giving rise to the final state $4e$ for the filter setting $n_\tau \geq 1$.

$4e$ final states	Multiplicity M	τ-leptons (n_{τ_e})	$BR_{\tau_e}^{n_{\tau_e}}$ [%]	Fraction of total events $(\frac{M}{65} \cdot BR_{\tau \to e}^{n_{\tau_e}})$[%]
$2e/e\tau_e$	4	1	17.8	1.10
$2e/2\tau_e$	2	2	3.28	0.098
$e\tau_e/e\tau_e$	4	2	3.28	0.195
$e\tau_e/2\tau_e$	4	3	0.57	0.035
$2\tau_e/2\tau_e$	1	4	0.10	0.002
	$\sum = 15$			$\sum \approx 1.43$

The result of this calculation can be found in table 8, giving a 1.43% probability of generating a $4e$ final state from $H^{\pm\pm}$ pair production. For a sample of $5 \cdot 10^4$ events it is expected that around $5 \cdot 10^4 \cdot 1.43\% \approx 715$ events would be generated. This indeed seems to be the case as can be seen in figure 31, both for the $4e$ and the analogous 4μ channels. The same method can be used for any of the 4-lepton final state channels.

F τ-lepton charge flip probability

F.1 Introduction

The charge flip rate is a measure for how often a particle detector assigns an incorrect charge to a detected particle, and an example would be for the detector to mistake an electron for a positron. For a study of same-signed dilepton signal charge flip is an important background to model correctly. As two leptons with the same charge are produced rarely in the Standard Model it must be estimated how often a fake same-signed pair is observed due to charge misidentification. Previously electron charge flip has been the most important factor to calculate, as muon charge is a well-reconstructed variable in the ATLAS detector compared and τ-lepton charge flip has been assumed to be negligible compared to other background. However, when including τ-leptons in the same-sign dilepton analysis there is value in a more rigorous examination of τ charge flip and its kinematic dependencies.

This section describes charge flip at the ATLAS detector and methods to estimate it, with focus on a data-driven likelihood charge flip estimation methods for electrons as used in the last SSDiLep same-sign dilepton analysis [18]. A proof-of-concept study using the same method for τ-lepton charge flip estimation is presented with the purpose of showing a possible method for charge-flip estimation in future τ-inclusive same-sign analyses.[5]

F.2 Theory

The accuracy of particle charge reconstruction in the ATLAS detector is essentially a question of tracking detector resolution. The charge of a reconstructed particle is calculated from the curvature of its track in the tracking detectors, as charged particles bend in magnetic fields. Charge flip rate ε is often expressed as a function of both p_T and η of the given particle, $\varepsilon(p_T, \eta)$. A high transverse momentum implies straighter tracks in the tracking detector, and result in a so-called *stiff track*. In the limit where the track is completely straight there is no possibility of definite charge reconstruction. The dependency of η arises as the detector is not spherically symmetric: For example, a high η implies the particle travels through less of the tracking detector and thus less track information in available for charge reconstruction.

For electrons another physical process increase the charge flip rate. In *trident events* an electron in the tracking detector radiates bremsstrahlung that produces an electron-positron pair. As the initial electron often has high p_T the created pair will be collinear with the original track due to conservation of momentum. When reconstructing the path of the initial electron there is then a chance that tracks of the created positron is chosen as the primary path, thus changing the curvature and the reconstructed charge of the initial electron. For both muons and τ_{had} trident events should however not be a factor for charge flip [32].

[5]The description of the method and the scripts used as a basis for this whole study was done Miha Muškinja as part of The SSDiLep group.

F.2.1 Methods for estimation of charge flip rates

There exist several methods for estimating the influence of charge flip in the Standard Model background. The most basic is to rely on Monte Carlo simulation where the simulated charge flip rate can be extracted from the truth information for each event. However, as charge flip is highly detector dependent this method depends on the detector simulation being accurate, which is hard to validate.

The alternative is to the actual data to estimate the charge flip rate, a so-called data-driven approach, as the estimate should then be independent of the detector simulation. One example of this to use the invariant mass spectrum of same-sign dileptons and look for a resonance peak at the Z-boson mass. No structure is expected in the same-sign invariant mass spectrum, and the assumption is that any observed same-sign resonance peak is due to charge flip of one of the pair-produced leptons of the Z decay. This enables a basic data-driven estimate of the charge flip rate as the ratio of same-sign dileptons whose reconstructed invariant mass is in a region centered around the Z resonance peak (the *Z mass window*) to the number of any-sign dileptons in the same region.

In general both Monte Carlo and data-driven methods are used in tandem in tandem for comparison and validation of each other. As an example of a more complex data-driven method the tag-and-probe method is presented below, while the theory behind the likelihood-based data-driven used for this proof-of-concept study will be explained in section F.3.

Tag-and-probe charge flip estimation for τ_{had} One estimation approach is based on the so-called *tag-and-probe* method. For pair-production of leptons from a Z-boson, one of the leptons whose charge is assumed to be reliably reconstructed is chosen as the *tag*. The tag is then used to estimate how often the other lepton, called the *probe*, is charge flipped by again using the Z mass window: If a resonance peak is seen in the same-sign dilepton spectrum it would then imply that the *probe* has been charge flipped.

To estimate the charge flip rate of τ_{had} the following decay

$$Z \to \tau\tau \to \tau_{had}\tau_{\mu}$$

can be used. The tag muon from τ_{μ} is assumed to have a correctly reconstructed charge as charge identification of muon in ATLAS is assumed to be insignificant compared to the charge flip rate of probe τ_{had}. With N_{SS} being all same-sign $\tau_{had}\tau_{\mu}$ events within the Z mass window and N_{AS} being all any-sign $\tau_{had}\tau_{\mu}$ events the probability of a τ_{had} to be charge flipped is calculated according to

$$\varepsilon_{\tau_{had}}(p_T, \eta) = \frac{N_{SS}(p_T, \eta)}{N_{AS}(p_T, \eta)},$$

where the charge flip probability dependency of η and p_T has been added.

The tag-and-probe method depends on both a well-defined Z resonance peak and a method for removing the other Standard Model background from the mass window. The former condition is not apparent in this case, as the event contain missing energy from neutrinos implying that

the visible transverse invariant mass $m_{T,vis}(\tau_{had}\tau_{\mu})$ resonance peak will be both broadened and shifted to lower invariant mass. Ways to alleviate this would be to use a more precise mass reconstruction algorithm for 2τ decay, with the commonly encountered *collinear approximation* method being presented in appendix C. It is also currently impossible to differentiate a τ_{μ} from a prompt muon, giving rise to more background as there is no guarantee that the invariant mass $m_T(\tau_{had-vis}\mu)$ arises from a Z-decay.

F.3 Likelihood charge flip estimation method

The likelihood method for electron charge flip estimation used in Ref. **[18]** will be presented here, with the description based on the internal note of Ref. [18]. The following derivation will use the notation $P(x;y)$ to mean the probability of x given a fixed y, where x often is an observable readout of data and y is a parameter such as average rate of events.

Defining the likelihood function

As in the tag-and-probe method described in paragraph F.2.1 one of the same-sign leptons observed in the Z mass window is assumed to be charge flipped. The probability $P(N_{SS};\lambda)$ to observe N_{SS} same-sign events given the expected number of charge flipped events λ is assumed to follow a Poissonian distribution, meaning that each charge flipped event occurs independently from each other and with a constant average rate. This Poissonian probability can be expressed according to

$$P(N_{SS};\lambda) = \frac{\lambda^{N_{SS}}e^{-\lambda}}{N_{SS}!}.$$

Where a *probability function* P gives the probability of an outcome given certain parameters, a *likelihood function* L gives the relative, unnormalized probability for a value of the parameters the given certain outcomes. A likelihood function $L(\lambda;N_{SS})$ can be constructed from taking the product of a probability function for all possible outcomes according to

$$L(\lambda;N_{SS}) = \prod_{N_{SS}} P(N_{SS};\lambda) = \prod_{N_{SS}} \frac{\lambda^{N_{SS}}e^{-\lambda}}{N_{SS}!}.$$

The λ that gives the largest L is the most likely λ based on the available data, and thus maximizing L in regards to λ will give the most probable average charge flip rate given the number of observed same-sign lepton pairs N_{SS}.

Defining charge flip probability in terms of (p_{T_i},η_i) bins

Charge flip probability ε depends on p_T and η. λ can thus be expressed in terms in charge flip probabilities $\varepsilon_{i/j}$, where i denotes the number of the (p_{T_i},η_i) bin of the leading lepton and j the bin number of the sub-leading. The charge flip probability is thus two-dimensional in p_T and η

bins. An example would be the choice of 3 bins inp_T and 2 binsη which would result in $2 \cdot 3 = 6$ bins in(p_T, η), with $i/j = 1, 2, ..., 6$. The choice of binning is important as more bins imply a greater resolution in the (p_T, η) dependency of charge flip. However, the fewer same-sign events found in each bin increases the statistical error of the resulting charge flip probabilities.

The likelihood for charge flip with the bin combination i, j, $L(\boldsymbol{\lambda}; \boldsymbol{N}_{SS})$, can thus be expressed as

$$L(\boldsymbol{\lambda}; \boldsymbol{N}_{SS}) = \prod_{i,,j} \frac{\lambda_{i,j}^{N_{SS}^{ij}} e^{-\lambda_{i,j}}}{N_{SS}^{ij}!}$$

where $\boldsymbol{\lambda}, \boldsymbol{N}_{SS}$ are vectors with $\boldsymbol{\lambda} = (\lambda_{11}, ...\lambda_{i_{max} j_{max}})$ and $\boldsymbol{N}_{SS} = (N_{SS}^{11}, ..., N_{SS}^{i_{max} j_{max}})$.

Defining the log-likelihood function

To simplify calculations a logarithm is often applied to the likelihood function giving the log-likelihood function $l(\boldsymbol{\lambda}; \boldsymbol{N}_{SS}) = \ln(-L(\boldsymbol{\lambda}; \boldsymbol{N}_{SS}))$, as it usually results in a simplified expression. It can be done as the likelihood function by design have one maximal value, and as the logarithm is a strictly increasing function l will attain its maximum value for the same parameter value as L attains its maximum. The log-likelihood of the charge flip probability can thus be expressed according to

$$l(\boldsymbol{\lambda}; \boldsymbol{N}_{SS}) = \ln(L(\boldsymbol{\lambda}; \boldsymbol{N}_{SS})) = \ln \prod_{i,,j} \frac{\lambda_{ij}^{N_{SS}^{ij}} e^{-\lambda_{ij}}}{N_{SS}^{ij}!} = \sum_{i,,j} \ln(\frac{\lambda_{ij}^{N_{SS}^{ij}} e^{-\lambda_{ij}}}{N_{SS}^{ij}!}) = \sum_{i,,j} (N_{SS}^{ij} \ln \lambda_{ij} - \lambda_{ij} - \ln(N_{SS}^{ij}!))$$

To extract the charge flip probability the expected number of charge flipped leptons for a leading lepton in bin i and a sub-leading lepton in bin j can be expressed in terms of charge flip probability $\varepsilon_{i,/j}$ according to

$$\lambda_{i,j} = \varepsilon_i(1-\varepsilon_j)N_{AS}^{ij} + \varepsilon_j(1-\varepsilon_i)N_{AS}^{ij} = (\varepsilon_i(1-\varepsilon_j) + \varepsilon_j(1-\varepsilon_i))N_{AS}^{ij} \tag{3}$$

where N_{AS}^{ij} is the total number of any-sign dileptons with the leading lepton in bin i and the sub-leading in bin j. Using equation 3 the log-likelihood function can be expressed according to

$$l(\boldsymbol{\varepsilon}; \boldsymbol{N}_{SS}, \boldsymbol{N}_{AS}) = \sum_{i,,j} (N_{SS}^{ij} \ln((\varepsilon_i(1-\varepsilon_j) + \varepsilon_j(1-\varepsilon_i))N^{ij}) - (\varepsilon_i(1-\varepsilon_j) + \varepsilon_j(1-\varepsilon_i))N_{AS}^{ij} - \ln(N_{SS}^{ij}!))$$

for the parameter vector $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_{i_{max}/j_{max}})$.

Parametrization

As N_{SS}^{ij} and N_{AS}^{ij} are known from data, the most likely charge flip probabilities $\varepsilon_{i/,j}$ in every (p_T, η) bin will be found when maximizing $l(\boldsymbol{\varepsilon}; \boldsymbol{N}_{SS}, \boldsymbol{N}_{AS})$ in regards to $\varepsilon_{i/,j}$. To simplify the maximization, the number of parameters can be reduced with the assumption that the η and p_T dependencies are independent from each other, giving

$$\varepsilon(p_T, \eta) = f(\eta) \cdot \sigma(p_T)$$

where $f(\eta)$ and $\sigma(p_T)$ are one-dimensional functions dependent only on η and p_T respectively. This assumption is based on the fact that η should affect the path through the tracking detector while p_T affect the curvature of the track, and while both of these should affect charge reconstruction separately there is no obvious dependency of the parameters.

A minimization algorithm[6] can then be run to find which $f(\eta)$ and $\sigma(p_T)$ gives the maximum likelihood for $l(\boldsymbol{\varepsilon}; \boldsymbol{N}_{SS}, \boldsymbol{N}_{AS})$. The effectiveness of minimization algorithms depend on not having a too large parameter space to search, and with many free parameters there is both the risk of finding a false minimum or simply not finding any stable minimum within a reasonable computational time. As $f(\eta)$ and $\sigma(p_T)$ are independent a way to reduce the number of free parameters is to require that $f(\eta)$ is normalized: This will make the algorithm first find the most likely $f(\eta)$ values, and then find the de facto scale factors $\sigma(p_T)$ that gives the most likely $\varepsilon(p_T, \eta)$. With N_η bins in η and N_{p_T} bins in p_T this gives a total number of free parameters N_{free} to

$$N_{free} = N_\eta + N_{p_T} - 1.$$

F.4 Method

F.4.1 Data set

For testing purposes a Monte Carlo $Z \to \tau\tau$ DxAOD sample with the EXOT0 derivation [3], requiring at least two light leptons in each event, was used. ROOT-readable ntuples were created from this using xAODAnaHelper which was expended from the previous SSDiLep analysis to include basic τ_{had} kinematics variables. A basic selection of events was applied to the sample, requiring exactly two τ_{had} with any charge to be present. Due to the EXOT0 derivation few events survived this cut: Few events contain both two τ_{had} and two light leptons $Z \to \tau\tau$ sample. The cutflow including cuts on b-jets are presented in figure 32 giving a sense of the scale of the dataset.

[6] Traditionally minimization algorithm are used to find the maximum of a functions it is a matter of convention whether to find a maximum or minimum.

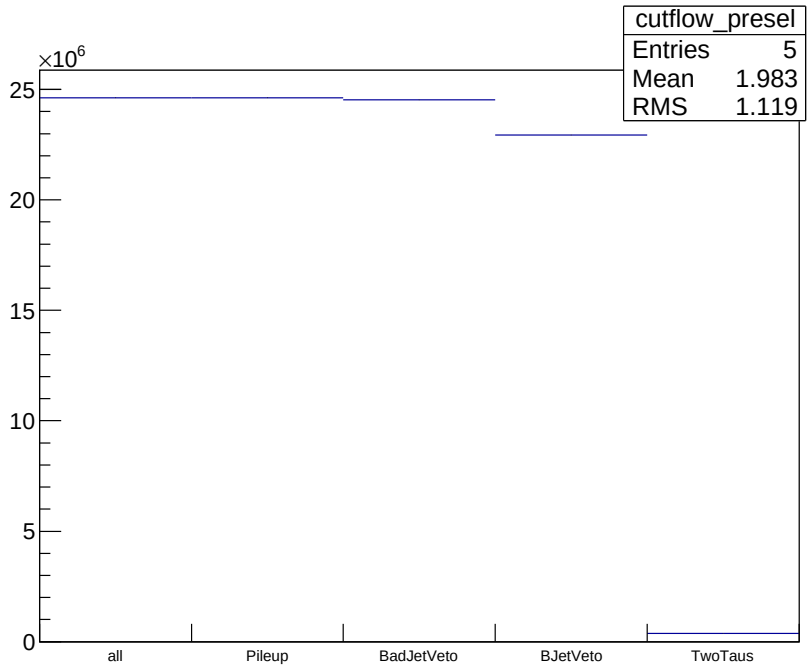


Figure 32: Cutflow diagram of the test ntuple.

F.4.2 Background removal

The assumption that the same-sign dilepton Z invariant mass peak arises due to only charge flipped events requires all background events not resulting from a Z decay to be removed. To remove the background the *sideband method* was used. It assumes that the background events found in the Z mass window can be estimated by looking at the background events outside the region. The invariant mass spectrum of the $2\tau_{had}$ was divided into three regions, one central region centered on the any-sign resonance $2\tau_{had}$ peak and two regions to each side of the center region. The background in the central region is removed by subtracting the events of the side regions from the center region. The width of the sidebands must therefore be equal to the width of the center region so as not to over- or under-estimate the background removed. The center of the resonance was determined by fitting a Gaussian function to the any-sign $m(2\tau_{had})$ spectrum using the ROOT fitting library, giving a mean of $\lambda_{center}(Z \to 2\tau_{had}) = 70.0 \pm 0.7$ GeV. The result of the fit presented in figure 33. For simplicity the Z mass window was set to $m_{Zwindow} = [40, 100]$ GeV with the sidebands set to $m_{SB,left} = [10, 40]$ GeV and $m_{SB,right} = [100, 130]$ for both the any-sign and same-sign invariant mass spectra.

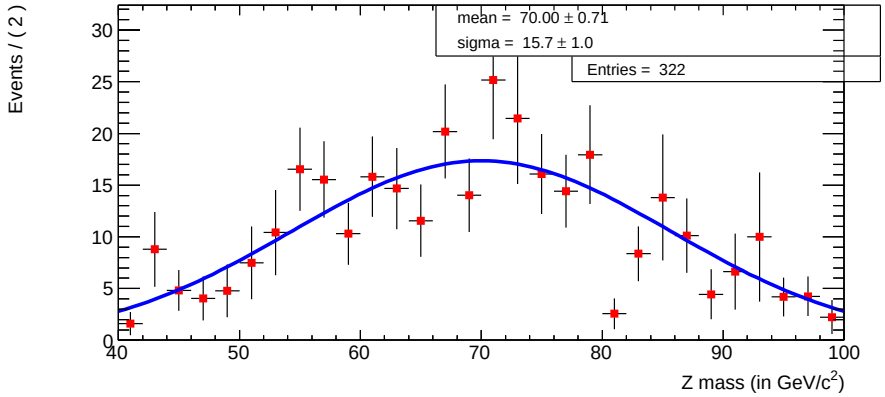
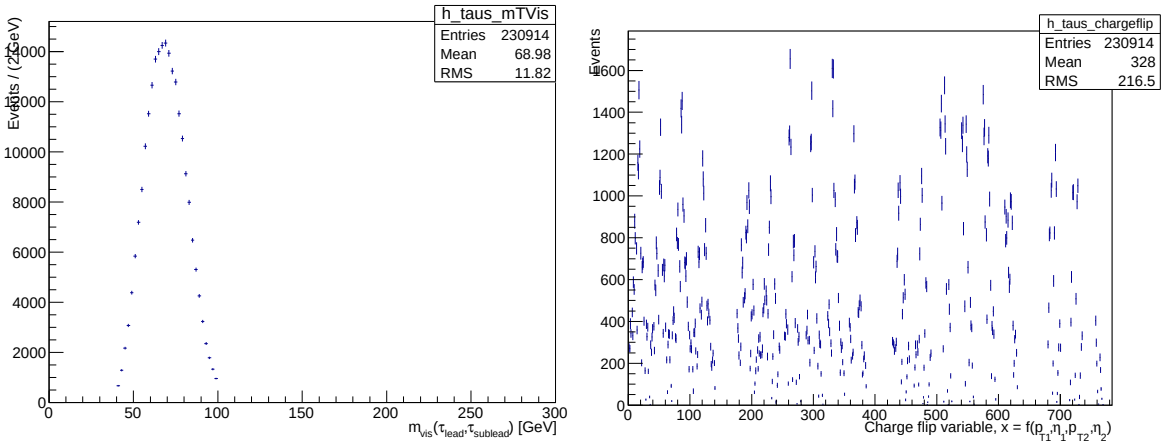


Figure 33: Result of the Gaussian fit of the any-sign $Z \rightarrow 2\tau_{had}$ resonance peak.

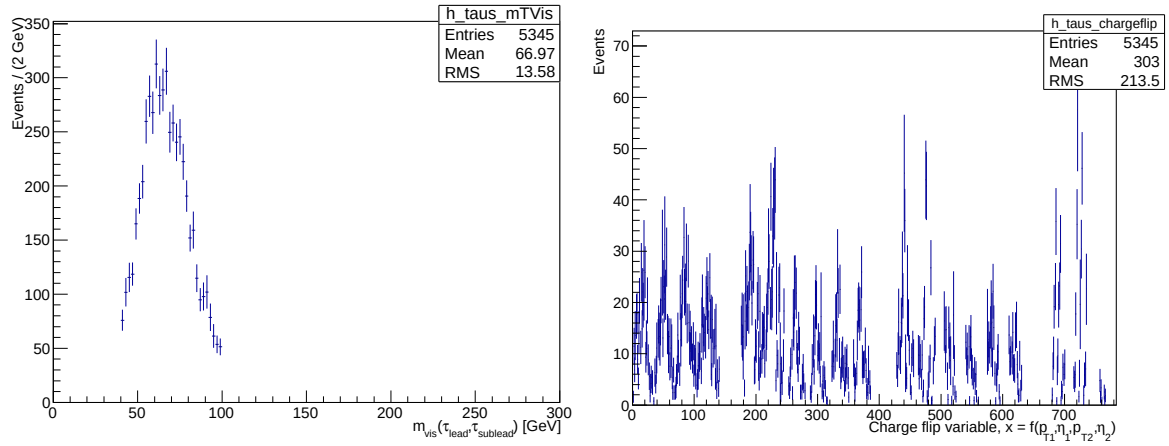
F.4.3 Log-likelihood maximization

The log-likelihood function is dependent on four kinematics variables, $(p_{T,1}, \eta_1, p_{T,2}, \eta_2)$ where $1, 2$ signifies the leading and subleading lepton respectively. To simplify the input to the minimization algorithm a charge flip variable $x = x(p_{T,1}, \eta_1, p_{T,2}, \eta_2)$, which could be output as a 1-dimensional ROOT-readable histogram. Dileptons from both the same-sign and any-sign Z mass window was used to construct two charge flip histograms, whose respective number of entries thus correspond to N_{AS} and N_{SS}. For background estimation two charge flip histograms constructed from the any-sign and same-sign sideband were constructed and subtracted from the Z mass window histograms. These four histograms were used as final input to the minimization algorithm. The isolated any-sign and same-sign Z mass window invariant mass spectra with their corresponding charge flip histogram is shown in figure 34.

The log-likelihood function was minimized using the ROOT minimization interface with the MINUIT2 package and the MIGRAD algorithm [4]. Several different parametrizations were tested to find one where the limited statistics of the data set would allow for a stable minimization: Too many bins in η and p_T resulted in no minimum found or different results with the same input. The final choice of the number of bins was $N_\eta = 6$ and $N_{p_T} = 4$ giving $N_{free} = 9$, with large bins at higher values of p_T and η to account for decreasing statistics at those values.



(a) Any-sign $2\tau_{had}$ selection.



(b) Same-sign $2\tau_{had}$ selection. Note that the same-sign Z mass window has been to narrowly defined as part of the left part of resonance peak is cut away.

Figure 34: The isolated visible invariant mass spectrum of $Z \rightarrow 2\tau_{had}$ (left) and the corresponding so-called *Binturong* diagram, plotting the charge flip parameter x (right) [26].

F.5 Result

The resulting charge flip plots for $f(\eta)$ and $\sigma(p_T)$ are presented in figure 35. Due to the limited statistics of the test sample few used for both the η and p_T dependency, and in the case of p_T a limit was put at 50GeV as the statistics in bins beyond that contained too few events to be for the minimization algorithm to be stable. One- and three-pronged τ_{had} have been plotted separately, with some statistics thus lost due to $Z \rightarrow \tau_{1-pronged}\tau_{3-pronged}$ not being used. The resulting charge flip probability is in the range $P(\eta, p_T) \sim 1.5\%$ for most bins. The result can be compared to the nominal charge flip plots for electrons used for the SSDiLep light lepton same-sign dilepton analysis presented in figure 36 for an example of the general look of the output of this method.

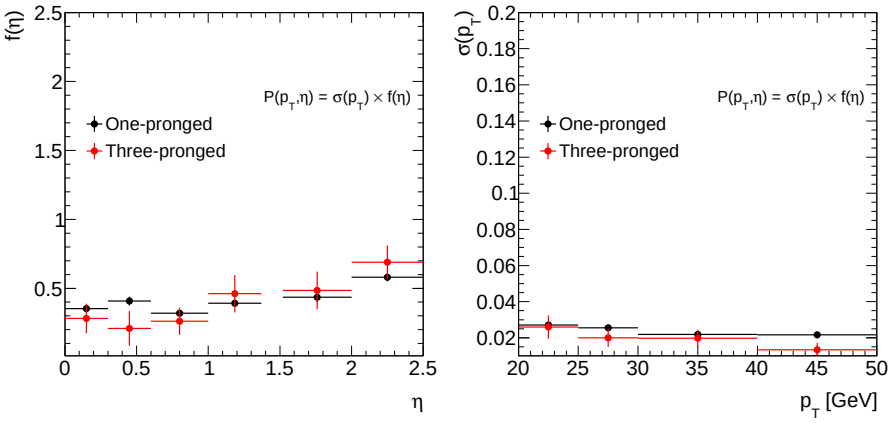


Figure 35: $f(\eta)$ and $\sigma(p_T)$ plots, divided up between one- and three-pronged τ_{had}. The crack region, $1.37 < \eta < 1.52$, is not included.

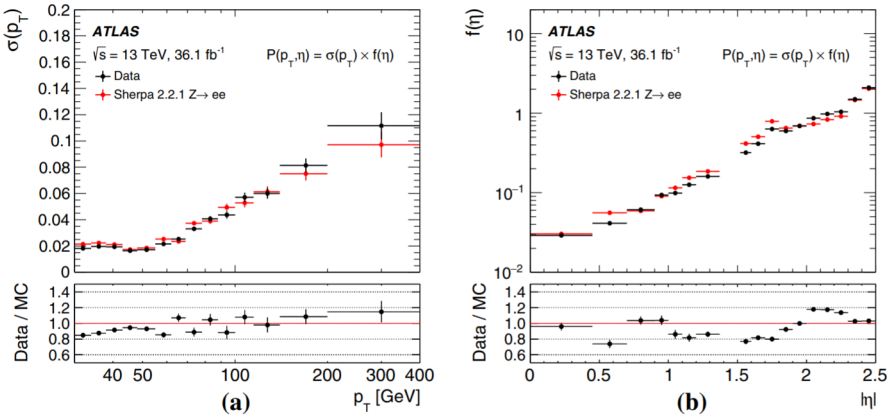


Figure 36: The nominal charge flip plots for electrons from Ref. [18].

F.6 Discussion

The resulting τ_{had} charge flip probabilities as seen in figure 35 is not an unreasonable order-of-magnitude estimation of what could be expected of the τ-lepton charge flip probability, and can be compared to the probability cited by CMS as $\varepsilon_{\tau_{had}} = 2.2\%$ [24]. It can therefore be concluded that the method can with only small modifications be applied to τ_{had}, even if this study gives no indication on the physical correctness of the result. For a complete study of the method several steps of validation should be conducted: Application of the method on real data with sufficient statistics, comparison of the result to the method applied to a simulated dataset, and a study on if with the attained charge flip probability gives a better modeling of simulation when compared to the data.

While the sideband method for background estimation is possible to implement using the visible invariant mass of the $2\tau_{had}$, it is highly problematic as the method is based on a well-defined

resonance peak with regions of pure background in the side bands. In the case of the $Z \rightarrow \tau\tau$ resonance peak, no such thing trivially exist due to the missing energy lost in ν_τ. While the visible invariant mass of $2\tau_{had}$ will give a resonance peak it will be significantly broadened and shifted to lower energies. In the case of data the ambient jet background will also be significant compared to the relatively clean invariant mass spectrum given by dielectrons or dimuons. Methods such as the collinear approximation might give a better-defined peak but will still have an inherently large tail of the distribution for higher masses. More complex methods for $2\tau_{had}$ mass reconstruction, such as the Missing Mass Calculator [28], might give a well-defined resonance peak but will not solve the problem of the jet background. For this data-driven charge flip estimation method to work a more detailed study of the $Z \rightarrow \tau\tau$ resonance and ways to isolate it from the background is therefore needed.

Another factor to take into account forτ-lepton charge reconstruction is that 3-pronged τ_{had} charge reconstruction depends on resolving three charged tracks compared to one, and as such must be handled differently than the 1-pronged case. The choice made for this study to ignore the events mixing one- and three-pronged τ_{had} was one made for the convenience of having to modify the existing method less. For τ-leptons it can be argued that the additional dimension of prongedness must be added to the parametrization, thus doubling the number of parameters that need to be maximized but accessing all available events. However, one- and three-pronged τ_{had} match each other within errors in almost all bins. While this is partly due to the low statistics and large errors of $\tau_{3-pronged}$ of the used test sample, the fact that no notable difference is seen prompts the question if there is a need for a separation in prongedness.

There is no clearcut answer on which method would give the best result regarding the τ-lepton charge: While it is enticing to use the powerful likelihood methods as presented here, it seems ill-suited for τ_{had} in general: data-driven methods need to have a clear and background-free signal in data, something τ_{had} might not be able to provide to a sufficiently high level. Due to the expected small charge flip probability the merit of a detailed study of the phenomenon might also be wasted time in an analysis context. The proper approach might be to first do a simple truth or tag and probe study for an estimation of the magnitude of the charge flip background. If then a problematic miss-modeling between data and simulation in some same-sign analysis regions occur a more in-depth analysis might be needed.

References

[1] Figure 2, the ATLAS detector, February 2011. URL http://brock.physik.uni-bonn.de/atlas.php?lang=en.

[2] Figure 1, the Standard Model particles, September 2014. URL http://en.wikipedia.org/wiki/Standard_Model.

[3] EXOT derivations TWiki, June 2018. URL https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/DerivationframeworkExotics#EXOT_derivations.

[4] ROOT minimization package Minuit2, June 2018. URL https://root.cern.ch/doc/v608/Minuit2Page.html.

[5] TauFilter JIRA AGENE ticket, January 2018. URL https://its.cern.ch/jira/browse/AGENE-1482.

[6] SUSY derivations TWiki, January 2018. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SUSYxAODDerivationsr21.

[7] Exotics MC signal sample request procedure, February 2018. URL https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ExoticsMCRequestsHowTo.

[8] Leading order diagram of Tau lepton decay, September 2018. URL https://en.wikipedia.org/wiki/Tau_(particle).

[9] Grid job monitoring website BigPanDA, 2018. URL https://bigpanda.cern.ch.

[10] Morad Aaboud and Others. A search for lepton-flavor-violating decays of the Z boson into a τ-lepton and a light lepton with the ATLAS detector. apr 2018. URL http://arxiv.org/abs/1804.09568.

[11] Simon Arnling Bååth. A Search for Indications of Majorana Neutrinos in Same-sign Dimuon Events at the ATLAS Detector, 2014. URL https://lup.lub.lu.se/student-papers/search/publication/4813170.

[12] Simon Arnling Bååth. Tau-inclusive DCH signal sample TWiki, December 2017. URL https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TauInclusiveDCH13TeV.

[13] Simon Arnling Bååth. Tau-inclusive DCH MC request validation slides, December 2017. URL https://indico.cern.ch/event/688549/contributions/2830268/attachments/1577751/2493399/ValidationplotsAtLeastOneTauFilter_corrections.pdf.

[14] Simon Arnling Bååth, Katja Mankinen, and Miha Muskinja. DCH Signal sample, December 2017. URL https://gitlab.cern.ch/sarnling/DCHValidationTauInclusive.

[15] ATLAS Collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, Geneva, nov 2015. URL https://cds.cern.ch/record/2064383.

[16] ATLAS Collaboration. Search for doubly charged Higgs boson production in multi-lepton final states with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV. oct 2017. URL http://arxiv.org/abs/1710.09748.

[17] ATLAS Collaboration. xAODAnaHelpers Documentation. Technical report, 2018. URL https://media.readthedocs.org/pdf/xaodanahelpers/latest/xaodanahelpers.pdf.

[18] ATLAS Collaboration. Search for doubly charged Higgs boson production in multi-lepton final states with the ATLAS detector using proton-proton collisions at sqrt(s)=13TeV. *European Physical Journal C*, 78(3):199, mar 2018. ISSN 14346052. doi: 10.1140/epjc/s10052-018-5661-z. URL https://epjc.epj.org/articles/epjc/abs/2018/03/10052_2018_Article_5661/10052_2018_Article_5661.html.

[19] ATLAS Tau Working Group. Standardized Notation for tau Leptons at ATLAS, 2012. URL https://twiki.cern.ch/twiki/pub/AtlasProtected/EditorialBoardGuidelines/tau-notation-v1.0.pdf.

[20] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, jul 2013. ISSN 05503213. doi: 10.1016/j.nuclphysb.2012.10.003. URL http://arxiv.org/abs/1207.1303http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003.

[21] R. Brun and F. Rademakers. ROOT - An Object Oriented Data Analysis Framework, 2014. URL http://root.cern.ch/. V. 5.34/19.

[22] James Catmore. Light lepton DCH signal sample TWiki, January 2018. URL https://indico.cern.ch/event/472469/contributions/1982677/attachments/1220934/1785823/intro_slides.pdf.

[23] Wei Chao, Zong Guo Si, Zhi zhong Xing, and Shun Zhou. Correlative signatures of heavy Majorana neutrinos and doubly-charged Higgs bosons at the Large Hadron Collider. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 666 (5):451–454, 2008. ISSN 03702693. doi: 10.1016/j.physletb.2008.08.003. URL https://arxiv.org/pdf/0804.1265v3.pdf.

[24] CMS Collaboration. A search for doubly-charged Higgs boson production in three and four lepton final states at $\sqrt{s} = 13 \sim$ TeV. Technical Report CMS-PAS-HIG-16-036, 2017. URL http://inspirehep.net/record/1510936?ln=enhttps://cds.cern.ch/record/2242956.

[25] Atlas Collaboration. ATLAS detector and physics performance: Technical Design Report, 1. Technical Report May, 1999. URL http://cdsweb.cern.ch/record/391176/files/cer-0317330.pdf.

[26] L. Cosson, L. L. Grassman, A. Zubaid, S. Vellayan, A. Tillier, and Géraldine Veron. Genetic diversity of captive binturongs (Arctictis binturong, Viverridae, Carnivora): Implications for conservation. *Journal of Zoology*, 271(4):386–395, apr 2007. ISSN 09528369. doi: 10.1111/j.

1469-7998.2006.00209.x. URL http://doi.wiley.com/10.1111/j.1469-7998.2006.00209.x.

[27] F Baltasar Dos, Santos Pedrosa, E Banas, P Banerjee, S Banerjee, D Banfi, A Bangert, V Bansal, S P Baranov, S Baranov, A Barashkou, T Barber, E L Barberio, D Barberis, M Barbero, D Y Bardin, T Barillari, M Barisonzi, T Barklow, N Barlow, B M Barnett, R M Barnett, A Baroncelli, A J Barr, F Barreiro, and P Barrillon. The ATLAS Simulation Infrastructure. 2010. URL https://arxiv.org/pdf/1005.4568.pdf.

[28] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A new mass reconstruction technique for resonances decaying to di-tau. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 654(1):481–489, dec 2011. ISSN 01689002. doi: 10.1016/j.nima.2011.07.009. URL http://arxiv.org/abs/1012.4686http://dx.doi.org/10.1016/j.nima.2011.07.009.

[29] Pavel Fileviez Pérez, Tao Han, Gui-Yu Huang, Tong Li, and Kai Wang. Neutrino masses and the CERN LHC: Testing the type II seesaw mechanism. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 78(1), may 2008. ISSN 15507998. doi: 10.1103/PhysRevD.78.015018. URL http://arxiv.org/abs/0805.3536http://dx.doi.org/10.1103/PhysRevD.78.015018.

[30] Vincent Garonne. Rucio Documentation Release 1.13.2-24-gfc2333c. 2017. URL https://media.readthedocs.org/pdf/rucio/latest/rucio.pdf.

[31] B K. Gjelsten. Event generation filter ParentChildFilter, January 2018. URL https://svnweb.cern.ch/trac/atlasoff/browser/Generators/GeneratorFilters/trunk/GeneratorFilters/ParentChildFilter.h.

[32] Anthony Hawkins. *Search for beyond Standard Model physics with same-sign dileptons*. PhD thesis, Lund University, 2014.

[33] Michael Heldmann, Xin Chen, and Simon Arnling Bååth. Event generation filter TauFilter, January 2018. URL https://svnweb.cern.ch/trac/atlasoff/browser/Generators/GeneratorFilters/trunk/GeneratorFilters/TauFilter.h.

[34] K. Huitu, J. Maalampi, A. Pietilä, and M. Raidal. Doubly charged Higgs at LHC. *Nuclear Physics B*, 487(1-2):27–42, jun 1997. ISSN 05503213. doi: 10.1016/S0550-3213(97)87466-4. URL http://arxiv.org/abs/hep-ph/9606311http://dx.doi.org/10.1016/S0550-3213(97)87466-4http://linkinghub.elsevier.com/retrieve/pii/S0550321397874664.

[35] Partha Konar and Abhaya Kumar Swain. Reconstructing semi-invisible events in resonant tau pair production from Higgs. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 757:211–215, 2016. ISSN 03702693. doi: 10.1016/j.physletb.2016.03.070. URL https://arxiv.org/pdf/1602.00552.pdf.

[36] Bruno Lenzi. The physics analysis tools project for the ATLAS experiment. In *AIP Conference Proceedings*, volume 1504, pages 987–990, 2012. ISBN 9780735411227. doi: 10.1063/1.4771862. URL https://doi.org/10.1063/1.4771862http://aip.scitation.org/toc/apc/1504/1.

[37] Tadashi Maeno. PanDA: Distributed production and distributed analysis system for ATLAS. In *Journal of Physics: Conference Series*, volume 119, 2008. doi: 10.1088/1742-6596/119/6/062036. URL http://inspirehep.net/record/803950/files/jpconf8_119_062036.pdf.

[38] Katja Mankinen. Light lepton DCH signal sample TWiki, January 2018. URL https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/DoublyChargedHiggs13TeV.

[39] C. Patrignani and Particle Data Group. Tau. *Chin. Phys. C*, 40(100001), 2017. doi: 10.1088/1674-1137/40/10/100001. URL http://pdg.lbl.gov/2017/listings/rpp2017-list-tau.pdf.

[40] M. L. Perl, G. S. Abrams, A. M. Boyarski, M. Breidenbach, D. D. Briggs, F. Bulos, W. Chinowsky, J. T. Dakin, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Goldhaber, G. Hanson, F. B. Heile, B. Jean-Marie, J. A. Kadyk, R. R. Larsen, A. M. Litke, D. Lüke, B. A. Lulu, V. Lüth, D. Lyon, C. C. Morehouse, J. M. Paterson, F. M. Pierre, T. P. Pun, P. A. Rapidis, B. Richter, B. Sadoulet, R. F. Schwitters, W. Tanenbaum, G. H. Trilling, F. Vannucci, J. S. Whitaker, F. C. Winkelmann, and J. E. Wiss. Evidence for anomalous lepton production in e+-e- annihilation. *Physical Review Letters*, 35(22):1489–1492, dec 1975. ISSN 00319007. doi: 10.1103/PhysRevLett.35.1489. URL https://link.aps.org/doi/10.1103/PhysRevLett.35.1489.

[41] M.L. Perl. Evidence for, and properties of, the new charged heavy lepton. 1977 (April), 1977. URL https://inis.iaea.org/search/search.aspx?orig_q=RN:9393842http://inis.iaea.org/search/search.aspx?orig_q=RN:9393842.

[42] Vikram Rentala, William Shepherd, and Shufang Su. A Simplified Model Approach to Same-sign Dilepton Resonances. 2011. URL https://arxiv.org/pdf/1105.1379.pdf.

[43] Thomas G Rizzo. Doubly charged Higgs bosons and lepton-number-violating processes. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 25(5), 1982. ISSN 05562821. doi: 10.1103/PhysRevD.25.1355. URL https://journals.aps.org/prd/pdf/10.1103/PhysRevD.25.1355.

[44] Lara Schildgen. Feedback on ATLFASTII from TauCP. Technical report, 2018. URL https://indico.cern.ch/event/687850/contributions/2850851/attachments/1583844/2503623/ATLFAST_TauCP_SimulationGroupMeeting.pdf.

[45] Dimitrios Sidiropoulos Kontos. Monte Carlo production of proton-proton collision events using the ATLAS@Home framework, 2018. URL https://lup.lub.lu.se/student-papers/search/publication/8932453.

[46] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191(1):159–177, oct 2015. ISSN 00104655. doi: 10.1016/j.cpc.2015.01.024. URL http://arxiv.org/abs/1410.3012http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[47] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 tune. *European Physical Journal C*, 74(8):1–39, apr 2014. ISSN 14346052. doi: 10.1140/epjc/s10052-014-3024-y. URL http://arxiv.org/abs/1404.5630http://dx.doi.org/10.1140/epjc/s10052-014-3024-y.

Acronyms

ATLAS A Toroidal LHC Apparatus

DCH Doubly Charged Higgs

LHC Large Hadron Collider

LO/NLO Leading Order/Next to Leading Order

LRSM Left-Right Symmetric Model

MET Missing Transverse Energy

MC Monte Carlo

MMA Missing Mass Calculator

PDF Parton Density Function

QCD Quantum Chromodynamics

SM Standard Model

SSDiLep group The ATLAS Same-sign dilepton group

vev Vacuum expectation value