UNIVERSITÄT BONN
Physikalisches Institut

Sensitivity of ATLAS to Alternative Mechanisms of Electroweak Symmetry Breaking in Vector Boson Scattering $qq \to qq\ell\nu\ell\nu$

von
Jan W. Schumacher

Abstract An analysis of the expected sensitivity of the ATLAS experiment at the Large Hadron Collider at CERN to alternative mechanisms of electroweak symmetry breaking in the dileptonic vector boson scattering channel is presented. With the generalized K-Matrix model of vector boson scattering recently implemented in the event generator WHIZARD, several additional resonances are investigated. WHIZARD is validated for ATLAS use and an interface for the Les Houches event format is adapted for the ATLAS software ATHENA. Systematic model and statistical Monte Carlo uncertainties are reduced with a signal definition using events reweighted in the couplings g of the new resonances. Angular correlations conserved by WHIZARD are used in the event selection. A multivariate analyzer is trained to take into account correlations between the selection variables and thereby to improve the sensitivity compared to cut analyses. The statistical analysis is implemented with a profile likelihood method taking into account systematic uncertainties and statistical uncertainties from Monte Carlo. Ensemble tests are performed to assure the applicability of the method. Expected discovery significances and coupling limits for new additional resonances in vector boson scattering are determined.

Post address:
Nussallee 12
53115 Bonn
Germany

BONN-IR-2010-013
Bonn University
October 2010
ISSN-0172-8741

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Michael Kobel
2. Gutachter:	Prof. Dr. Norbert Wermes
Tag der Promotion:	27.09.2010
Erscheinungsjahr:	2010

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn unter `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

Contents

1 Introduction

1.1 Why Look Beyond the Standard Model

The Standard Model of particle physics (SM) [1, 2, 3] has enjoyed huge success in explaining fundamental particle interactions. Its very name bears witness to that fact.

However, there are still open questions that are not answered within that framework. Some are directly based on experimental observation, e.g.:

- How do particles acquire their masses?
- Is cosmological dark matter made up of as of yet undiscovered particles?

Others pertain to the systematics of the theory, e.g.:

- Is there a way to do with fewer free parameters?
- Are the fundamental forces unified at some energy scale?
- How does gravity fit into quantum theory?

How elementary particles acquire their masses is perhaps one of the most important open questions. Without an answer to this, the current SM is inconsistent. The flagship idea how to explain this, is the Higgs mechanism [4]. It breaks electroweak symmetry spontaneously by introducing a new scalar field with a continuum of lowest states. Phenomenologically, the excitations of this field introduce a new particle, the Higgs boson. On the experimental discovery of this particle hinges a large part of our current world view of particle physics. Finally finding this particle which has not been seen despite great effort in any previous experiment to date [5] was in important reason to build the LARGE HADRON COLLIDER (LHC). The Higgs boson has become so enshrined in modern particle physics thinking, that it is commonly termed the Standard Model Higgs boson, despite its as of yet elusive nature.

Even not finding the Higgs boson at the LHC would still be a huge boon to particle physics. Not finding would mean exclusion. Exclusion would mean that the electroweak symmetry breaking (EWSB) remains an open question. Without EWSB, the Standard Model predictions for vector boson scattering (VBS) $VV \to VV, V \in W, Z$ violate perturbative unitarity for longitudinally polarized W bosons at about 1.2 TeV center of mass energy. If no Higgs boson is found, then new physics beyond the Standard Model must be just around the corner on the TeV-scale accessible to the LHC.

If a light enough Higgs boson is present and found, it will guarantee unitarity conservation. VBS at high energies will still be very interesting, because it is closely related to EWSB. Any deviation of its cross section from the SM plus Higgs boson predictions would hint at new physics. If, on the other hand, the Higgs boson is not found, then new physics is guaranteed in this channel. A new effect would have to unitarize longitudinal W boson scattering or be non-perturbative.

Instead of searching for effects from a particular theory, one may consider effective theories. Faced with a channel in which new phenomenology is highly expected but also faced with many possible alternatives, generality is an advantage. The electroweak chiral Lagrangian (EWChL) [6] sees the Standard Model as an effective low energy theory. It allows additional anomalous couplings, modeling the low-energy effects of resonances just beyond the energies within reach. Assuming custodial symmetry [7, 8, 9] (an approximate symmetry related to the m^W/m^Z mass ratio) and CP conservation, there are two additional free parameters related to vector boson scattering. By themselves, anomalous couplings do not unitarize the scattering amplitudes. At LHC the VBS center of mass energies will be too high, necessitating an explicit unitarization procedure. There are different options, among them Padé [10, 11] and K-Matrix unitarization. Depending on parameters, they may predict a continuum or additional resonances in VBS.

In this thesis, the EWChL with K-Matrix unitarization was investigated as a model. It allows to include resonances of various spin and weak isospin configurations with arbitrary masses and couplings. Several specific models like a heavy SM Higgs boson or Padé unitarization are included as special cases. This makes it a very generic tool to look for new phenomenology in this channel. The particular channel that was considered is VBS with two leptons in the final state. The two leptons ensure a very clean signature in a mostly hadronic environment. A previous study using the related Padé unitarization has looked into the semileptonic channel where one vector boson decays hadronically.

The experiments at the LHC have taken their first data. However, for this channel, a luminosity on the order of $\mathscr{L} = 100\,\mathrm{fb}^{-1}$ is needed. This amount will take several years to collect. In the meantime it is important to use Monte Carlo simulations to develop the analysis techniques to be used when data becomes available. At the same time this allows to determine sensitivities and expected limits for searches. This strategy allows to find channels with best potential for discovery or limit setting.

The aim of this thesis is to give sensitivities and expected limits on K-Matrix resonances for A Toroidal LHC Apparatus (ATLAS) in the dileptonic vector boson scattering channel.

1.2 Overview

The present thesis is structured into multiple sections.

Section 1: Introduction

Section 2: Theoretical Basis In this section, the electroweak chiral Lagrangian and the K-Matrix model used for the description of vector boson scattering will be introduced. A discussion of the characteristics of the signal process follows.

Section 3: Experiment The LHC and the ATLAS experiment will be described and an overview of the event reconstruction given.

Section 4: Monte Carlo Generators To simulate the models under consideration, Monte Carlo event generators will be necessary. The matrix element generator WHIZARD will be introduced here. Its validation for official ATLAS use can also be found in this section.

Section 5: Monte Carlo Simulation Following this, the signal and background samples used in the analysis will be characterized. Special considerations necessary for signal definition will be made.

Section 6: Event Selection In this section the methods used to increase the signal fraction in the selected events will be explained. This includes the trigger, object selection and finally the event selection criteria themselves.

Section 7: Sensitivity and Limits With the preceding results available it will then be possible to do a statistical analysis taking into account systematic uncertainties and statistical uncertainties from Monte Carlo samples. The resulting sensitivities and limits will be determined and quoted.

Section 8: Conclusions and Outlook The final conclusions of this thesis will be given.

2 Theoretical Basis

2.1 The Case for an Effective Theory

In a Standard Model without Higgs bosons, some new phenomenon must engender the Goldstone bosons making up the longitudinal components of the vector bosons. They are directly associated with EWSB [12]. The one channel, therefore, that is almost certain to yield information on the EWSB, is the scattering of longitudinal W and Z bosons.

The Higgs mechanism [4] is only one way to to supply these additional degrees of freedom. It suffers from the fact that experimental limits on the Higgs boson mass have already excluded the most likely mass range according to electroweak precision data [5]. At the same time it does not explain the huge range of mass scales in the SM: About nineteen orders of magnitude from the known fermion masses to the Planck scale and several orders of magnitude down to the neutrino masses. The characteristic scale of EWSB, v, is conventionally taken to be[1]

$$v = (\sqrt{2} G_F)^{-\frac{1}{2}} \approx 246\,\mathrm{GeV} \tag{2.1}$$

If electroweak gauge symmetry is accidental, the size of the breaking terms is accidental as well. The ratio $v/M_{\mathrm{Planck}} \approx 10^{-17}$ seems arbitrarily small and no fundamental reason for its size can be given. Therefore, models that assert an exact symmetry, which is spontaneously broken at low energies, are generally preferred. Even so, the Higgs mechanism does not actually explain this number, it still is an arbitrary parameter.

Strongly coupled field theories on the other hand, may generate this scale from their dynamics due to normalization group running [13]. Models with dynamical EWSB include technicolor [14, 15, 16, 17] extended technicolor [18, 19] topcolor [20, 21] topcolor assisted technicolor [22, 23], condensation of neutrinos [24, 25, 26, 27] and top see-saw [28, 29]. This is not the end of it, additional possibilities are suggested by extra dimension models with Kaluza-Klein states of gauge bosons [30, 31] or little Higgs models [32].

To prepare for data in an experiment to test EWSB, there are two strategic options how to cope with the plethora of models available. One may either pick a model and search for specific signatures. For the SM Higgs boson, supersymmetric extensions and some other models this is done in-depth (for ATLAS cf. [33].) The other approach is to look for a generic class of deviations from the (naive) SM cross sections. For

---

[1] For readability, the natural units $c \equiv \hbar \equiv 1$ are omitted throughout the theory section.

that one needs a generic model with adjustable parameters. The more specific the predictions are, the better the experimental sensitivity can be optimized. The more models may be subsumed, the greater the utility.

2.2 Electroweak Chiral Lagrangian

In Reference [34] a generic, parametrized model for vector boson scattering is suggested, which combines a low energy effective theory with additional resonances and K-Matrix unitarization.

The effective Lagrangian describing physics up to the scale of about 1 TeV is called the electroweak chiral Lagrangian. It is an expansion in E/Λ rather than in the weak coupling constants. Λ is a cutoff value customarily chosen to be $4\pi v$. Known low energy properties of the SM need to be reproduced. $U(1)_Q$ and $SU(3)_C$ gauge invariance has to hold, as well as electroweak symmetry $SU(2)_L \times U(1)_Y$ for the vector boson couplings. The latter is broken by the fermion and boson masses.

To parametrize our ignorance about the symmetry breaking sector, a new matrix-valued field $\Sigma(x)$ is introduced. Under local $SU(2)_L \times U(1)_Y$ transformations it behaves as follows:

$$\Sigma \to U_L \Sigma U_R^\dagger \tag{2.2}$$

Given the Pauli matrices τ^a and gauge transformation parameters $\beta^a(x)$, then $U_L(x) = \exp\left(i \sum_{a=1}^{3} \beta^a(x)\tau^a\right)$ and $U_R(x) = \exp\left(i\beta^0(x)\tau^3\right)$. Parametrized as a unitary matrix, the field Σ may also be written as

$$\Sigma(x) = \exp\left(\frac{-i}{v}\mathbf{w}(x)\right) \tag{2.3}$$

where $\mathbf{w} = \sum_{a=1}^{3} w^a\tau^a$. The w^a serve as the Goldstone bosons needed for the longitudinal vector bosons. While a complete theory may contain multiple new fields instead of Σ, the scalars w^a must be present [35, 36, 37]. In this picture, the Higgs mechanism becomes a special case of Σ. Using this field, an effective Lagrangian may be defined, i.e. the electroweak chiral Lagrangian (EWChL) [6, 38].

Perturbative expansion of this Lagrangian accounts for all measurements in this sector so far. While at leading order (LO) all its parameters are determined by low-energy data, additional coefficients appear in next-to-leading order (NLO) which describe the high-energy behavior and are still to be determined. This is the region where the ATLAS experiment is expected to help shed light on. With the definitions [6]

$$\mathbf{V}_\mu = \Sigma(D_\mu\Sigma)^\dagger \text{ and } \mathbf{T} = \Sigma\tau^3\Sigma^\dagger \tag{2.4}$$

all couplings can be expressed. Weak isospin $SU(2)_C$ (“C” for custodial) is approximately conserved at low energies as evidenced by the W-Z mass ratio [7, 8, 9]

$$\rho \equiv \frac{m_W^2}{m_Z^2 c_w^2} \approx 1 \tag{2.5}$$

Considering only NLO terms respecting this symmetry, five parameters are left. Of these, three are well-constrained by LEP, two of which affect three-boson couplings. Only the two remaining are unconstrained and pertain to vector boson scattering [34, 6]

$$\mathcal{L}_4 = \alpha_4 \left(\mathrm{tr}\left[\mathbf{V}_\mu \mathbf{V}_\nu\right]\right)^2 \tag{2.6}$$

$$\mathcal{L}_5 = \alpha_5 \left(\mathrm{tr}\left[\mathbf{V}_\mu \mathbf{V}^\mu\right]\right)^2 \tag{2.7}$$

The low energy behavior modeled with the EWChL and the NLO terms (2.6) and (2.7) has to match up to the high energy region of new physics. Additional resonances from new physics are reflected in changed anomalous coupling parameters α_4 and α_5, effectively approximating the rising slope of a resonance out of energy reach. A table of shifts derived by integrating out the resonances at tree level can be found in [34]. While the values of α_4 and α_5 may indicate the type of any dominant resonance, measuring them is very difficult at a machine like the LHC[2]. A previous study for ATLAS has examined the discovery potential for the two parameters, and given an estimate of the sensitivity [41]. However, at the LHC the high energy region above the applicability of the EWChL itself cannot be ignored. The energy reach is simply too large. Two additional components are necessary to extend the model of Reference [34] to higher energies: A set of resonances from which one can be picked to model the first resonance of the new physics spectrum and a unitarization procedure to keep the rise of the cross section for high center of mass energies reasonable.

2.3 Resonances

Five resonances are added to the EWChL by the following additional terms [34]

$$\mathcal{L}_{sigma} = -\frac{1}{2}\sigma\left(M_\sigma^2 + \partial^2\right)\sigma + \sigma j_\sigma \tag{2.8}$$

$$\mathcal{L}_\phi = -\frac{1}{2}\left[\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\phi}\left(M_\phi^2 + \partial^2\right)\boldsymbol{\phi}\right] + \mathrm{tr}\left[\boldsymbol{\phi}\mathbf{j}_\phi\right]\right] \tag{2.9}$$

$$\mathcal{L}_\rho = \frac{1}{2}\left[\frac{M_\rho^2}{2}\mathrm{tr}\left[\boldsymbol{\rho}_\mu\boldsymbol{\rho}^\mu\right] - \frac{1}{4}\mathrm{tr}\left[\boldsymbol{\rho}_{\mu\nu}\boldsymbol{\rho}^{\mu\nu}\right] + \mathrm{tr}\left[\mathbf{j}_\rho^\mu\boldsymbol{\rho}_\mu\right]\right] \tag{2.10}$$

$$\mathcal{L}_f = \mathcal{L}_{\mathrm{kin}} - \frac{M_f^2}{2} f_{\mu\nu} f^{\mu\nu} - f_{\mu\nu} j_f^{\mu\nu} \tag{2.11}$$

$$\mathcal{L}_t = \mathcal{L}_{\mathrm{kin}} - \frac{M_t}{4}\mathrm{tr}\left[\mathbf{t}_{\mu\nu}\mathbf{t}^{\mu\nu}\right] + \mathrm{tr}\left[\mathbf{t}_{\mu\nu}\mathbf{j}_t^{\mu\nu}\right] \tag{2.12}$$

with the appropriate currents. Only couplings to the longitudinal gauge bosons are realized, as the transversal gauge bosons are not directly involved in EWSB. The spin and weak isospin configurations of the resonances are given in Table 2.1. For

---

[2] At the envisioned International Linear Collider ILC, a more precise measurement ought to be possible [39, 40]. However, this is at the expense of a lower energy reach.

Type	Spin J	Isospin I	Electric Charges	Width Γ/Γ_0
σ	0	0	0	6
ϕ	0	2	$--$, $-$, 0, $+$, $++$	1
ρ	1	1	$-$, 0, $+$	$\frac{4}{3}\left(\frac{v^2}{m^2}\right)$
f	2	0	0	$\frac{1}{5}$
t	2	2	$--$, $-$, 0, $+$, $++$	$\frac{1}{30}$

Table 2.1: *Overview of resonance properties: Spins, isospins and corresponding charges for the implemented resonances. Width coefficients to need to be multiplied by* $\Gamma_0 = g^2 m^3/64\pi v^2$.

the isoscalars σ and f, the physical fields are neutral, the isovector ρ has neutral and singly-charged components and the isotensors ϕ and t have in addition doubly-charged components.

Using the Goldstone boson equivalence theorem (GBET) decay widths into vector bosons in leading order of the electroweak coupling expansion may be calculated. These scale as functions of mass m and coupling g. Coefficients are given in Table 2.1. While technically possible to include more than one of the resonances, this is not done in this thesis, so that resonance masses m and g are quoted without index. In particular $g = 0$ signifies the absence of all resonances. The shifts in α_4, α_5 originally introduced by implicit resonances correspond to the rising edges of the resonances. If these are included explicitly, the anomalous coupling parameters should not be shifted to avoid double counting. While not strictly necessary, they are usually kept at zero for simplicity. For the resonant model, that leaves a coupling and a mass per resonance. By restricting oneself to one resonance at a time this boils down to a discrete choice out of five resonances and two continuous parameters.

2.4 K-Matrix Unitarization

Despite the additional resonances, the scattering amplitudes rise asymptotically with a power of s for large energies. Only in the case of a scalar isoscalar σ of coupling $g = 1.0$, the resonance compensates the low energy effective theory's rise exactly [34]. This is equivalent to a SM Higgs boson. Figure 2.1 illustrates this behavior. Unitarity requires that the normalized eigenamplitudes $a_{IJ} = \frac{1}{32\pi} A_{IJ}$ for spin J and weak isospin I stay on the Argand-circle in the complex plane

$$|a_{IJ}(s) - i/2| = 1/2 \tag{2.13}$$

Amplitudes calculated in finite order will not usually conform to this constraint. A complete theory would have to fix this. Before this is known, one may unitarize any amplitude manually.

A prescription how to do this, is available in the K-Matrix unitarization [42, 43]. An arbitrary amplitude $a(s)$ may be mapped onto an amplitude $\hat{a}(s)$ on the Argand-

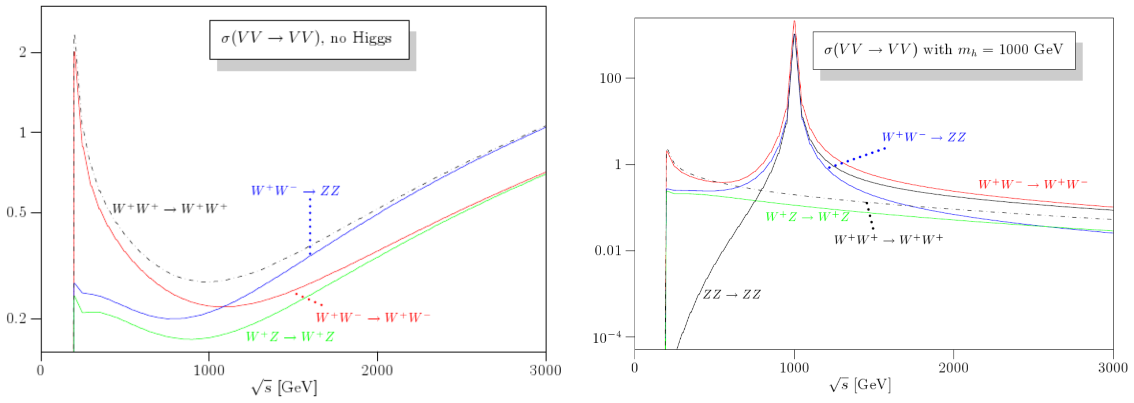


Figure 2.1: *Graphs of vector boson on-shell scattering. Left: Extrapolation of the low energy approximation without any unitarization. The amplitude scales with s. Right: A SM Higgs boson exactly cancels the divergence. (Graphs from Reference [34])*

circle by a simple transformation

$$\hat{a}(s) \equiv \frac{1}{\mathrm{Re}(1/a(s)) - i} \tag{2.14}$$

Figure 2.2 shows an illustration of this procedure. The effect on an amplitude $A(s) = 32\pi a(s) = s/v^2$ is convergence to the top of the Argand-circle

$$\lim_{s\to\infty} \hat{A}(s) = \lim_{s\to\infty} \frac{s/v^2}{1 - \frac{i}{32\pi v^2} s} = 32\pi i \tag{2.15}$$

Formally this corresponds to a resonance of infinite mass. The cross section asymptotically approaches saturation. A pole amplitude $A(s) = -c/(s-m^2)$ is transformed into a Breit-Wigner resonance. If c is dependent on s, the result is a Breit-Wigner with an s-dependent width. In particular, the s-dependent term in the tree-level amplitude of the Higgs boson like scalar isoscalar σ with a coupling of $g = 1.0$ [34]

$$A(s) = -\frac{m^2}{v^2} \frac{s}{s + m^2} \tag{2.16}$$

is mapped onto

$$\hat{A}(s) = -\frac{m^2}{s^2} \frac{s}{s - m^2 + im\Gamma\frac{s}{m^2}} \quad \text{with } \Gamma = \frac{m^2}{32\pi v^2} m \tag{2.17}$$

Another popular unitarization procedure is the Padé or inverse amplitude method [44]. It turns out to be a special case of the K-Matrix scheme applied to amplitudes with resonances. Typically the low energy effective theory parameters α_4 and α_5 are specified, defining a resonance of tunable mass. The width to mass ratio is set by the

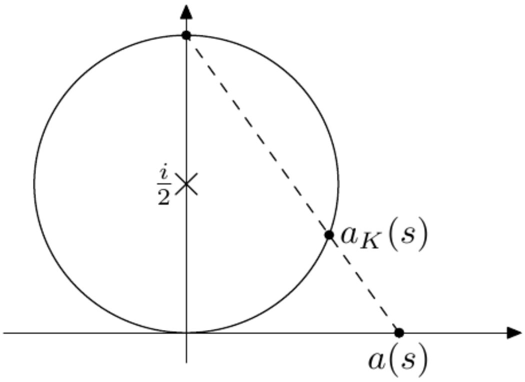


Figure 2.2: *Illustration of the K-Matrix projection $a_K(s)$ of a real scattering amplitude $a(s)$ on the Argand-circle. A divergent amplitude $a(s) \to \infty$ will asymptotically approach the point $0+i$. Resonance amplitudes pass it at the pole and continue down on the other side of the circle. (Figure from Reference [34])*

method and cannot be arbitrarily chosen. A previous study [33] has examined the ATLAS discovery potential for vector boson scattering using Padé unitarization in the semileptonic and $WZ \to \ell\nu\ell\ell$ channel.

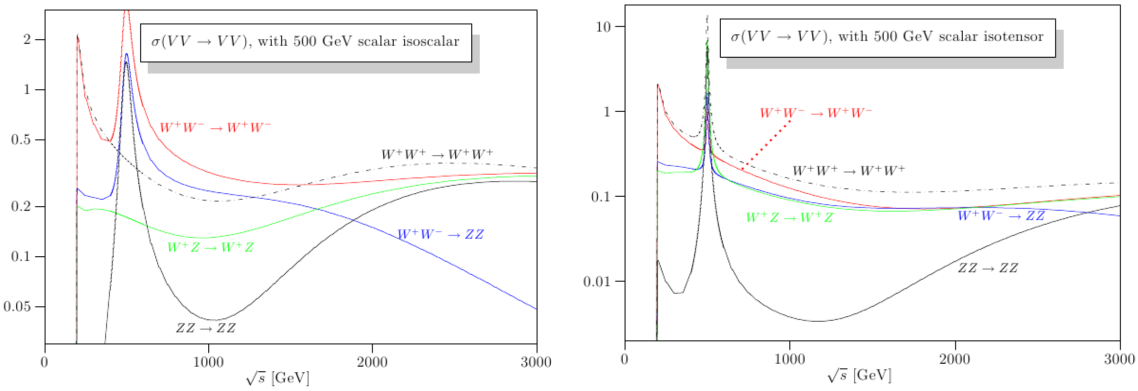
While the complete K-Matrix model including unitarization gives predictions for arbitrary center of mass energies of VBS, anything beyond the first resonance needs to be taken with a grain of salt. The main purpose of unitarization is not to replace a theory which models this region possibly with a rich additional spectrum but to make sure that the high energy cross section is not drastically overestimated. Otherwise, it may generate spurious sensitivity from the high energy tail instead of the resonance under investigation. This fact also motivates the choice of the null hypothesis in Section 4.2.5.

Figure 2.3 shows sample resonances for the complete K-Matrix model. Resonance widths strongly depend on the type. The high energy tails all tend to saturation due to unitarization. In the case of tensor resonances, additional terms $\propto s^2/M^2$ are visible before unitarization becomes dominant.

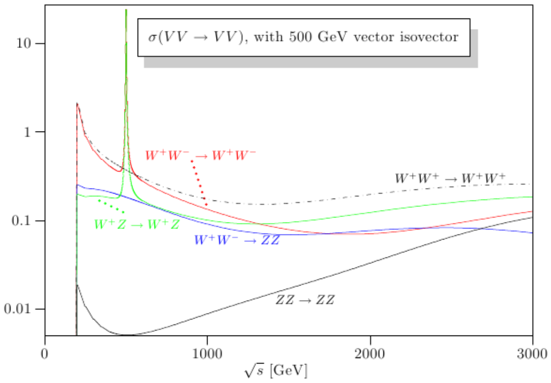
Several specific models may be closely approximated with this K-Matrix model (cf. [34, Appendix D].) The SM is one example, which has been discussed above. A coupling of $g = 1.0$ for a scalar isoscalar σ reproduces a Higgs boson in VBS. Scalar and vector resonances from the wide array of models introduced in [10, 11] may be reproduced. Similarly, the BESS model [45] may be implemented. For the Padé model, a mapping of α_4 and α_5 to the appropriate resonances in the K-Matrix model exists. Alternatively, a Padé scalar may be approximated[3] by a K-Matrix σ resonance of coupling $g = \sqrt{2/3} \approx 0.81$. These considerations give a rough idea which values of the couplings g are interesting to study experimentally. For a generic strong EWSB, a larger coupling up to $g \approx \sqrt{2\pi} \approx 2.5$ is reasonable as well [46].

---

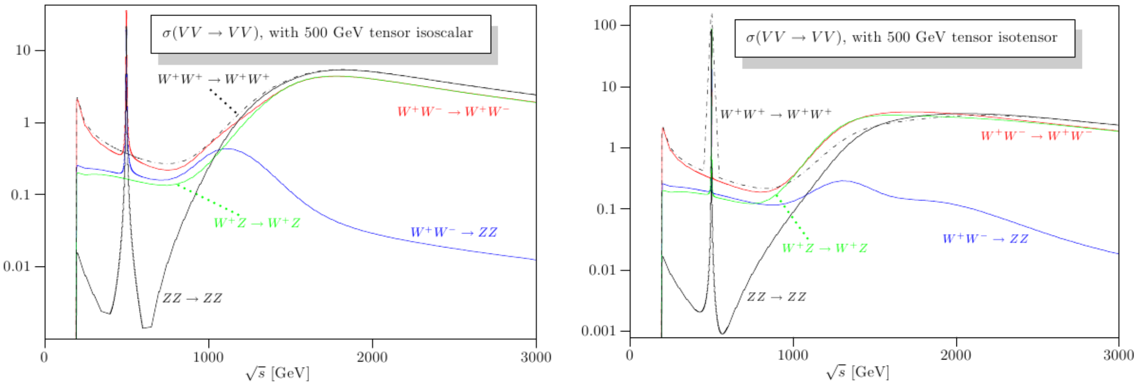
[3] An additional factor m^2/v^2 is found for the width Γ with respect to formula (129a) in [34]



(a) Scalar Isoscalar σ **(b)** Scalar Isotensor ϕ



(c) Vector Isovector ρ



(d) Tensor Isoscalar f **(e)** Tensor Isotensor t

Figure 2.3: *On-shell vector boson scattering cross sections in nanobarn for multiple resonances in the K-Matrix model. Widths vary widely between resonances, see also Table 2.1. While in particular the tensor resonances have strongly rising cross sections after the resonance, unitarization keeps them under control. (Graphs from Reference [34])*

2.5 Signal Characteristics at the LHC

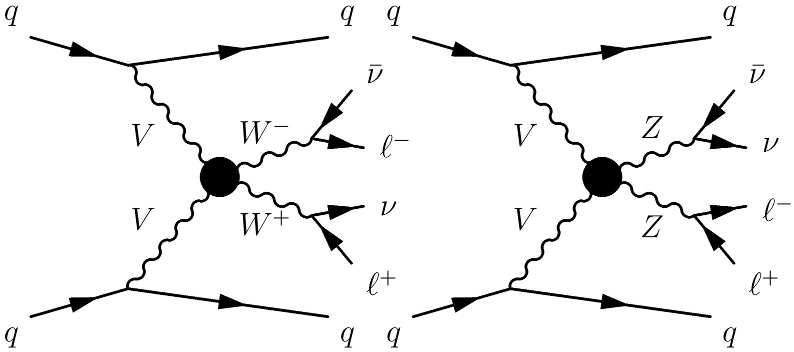


Figure 2.4: *Feynman diagrams of vector boson scattering $WW/ZZ \to \ell\nu\ell\nu$. Proton remnants not shown. Same-sign WW is also possible, but not shown.*

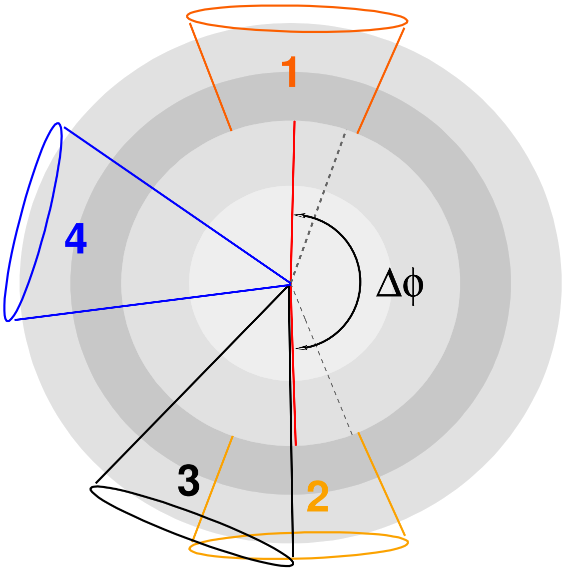
At the LHC vector boson scattering involves two quarks radiating off the initial vector bosons (cf. Figure 2.4.) In addition to the decay products of the bosons, two quark jets are expected to be seen in a typical event. The event shape is very similar to the related vector boson fusion (VBF) channels, which are prominent for a SM Higgs boson around $160\,\mathrm{GeV}/c^2$. Figure 2.5 shows a sketch of such an event in a schematic collider detector. The angle $\varphi \in [-\pi, \pi)$ is measured in the plane perpendicular to the beam axis. With the polar angle $\vartheta \in [0, \pi]$ measured from the beam axis in positive z direction, the pseudorapidity is defined as

$$\eta \equiv -\log\left(\tan\frac{\vartheta}{2}\right) \tag{2.18}$$

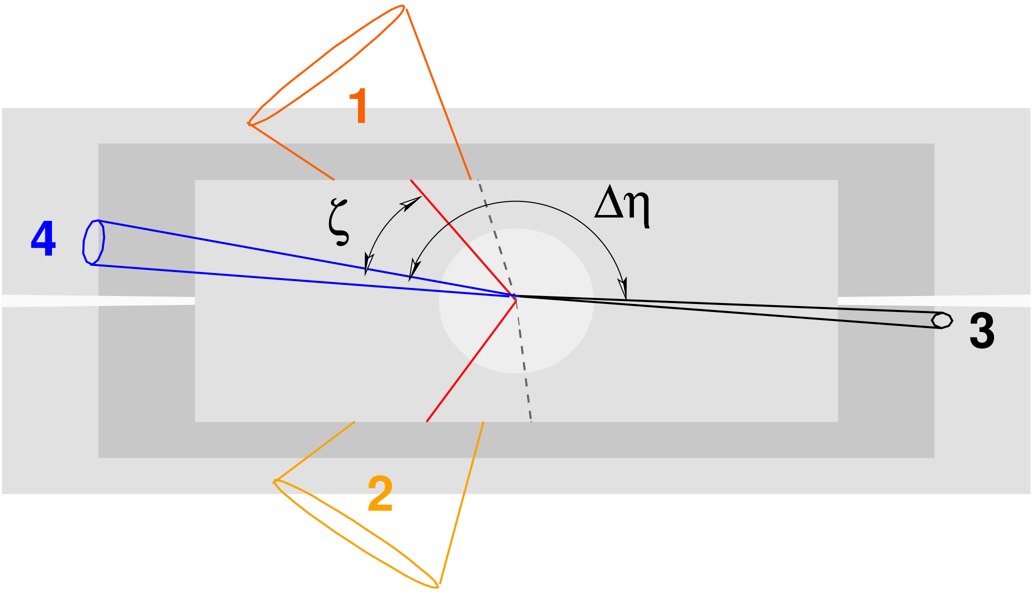
The two jets ("3" and "4" in Figure 2.5) tend to be energetic and in the forward direction, their separation in pseudorapidity $\Delta\eta^{\text{tagjets}}$ is usually large. As this property makes them a particular feature of VBS/VBF, they are commonly called tagging jets. Additional jets are less likely between the two tagging jets, because of the lack of color flow between the quarks in this electroweak process. Third jets display a gap in their pseudorapidity distribution (see for example [47, 48].) In a hadronic environment this is a very specific feature. QCD background processes tend to have a much more central distribution of additional jets and to contain more jets overall.

The vector boson decay products are relatively more central than the tagging jets. In the double leptonic channel investigated here, these are two charged leptons and two neutrinos, which may be produced from two W or two Z bosons. For massive resonances, the outgoing vector bosons are highly boosted such that the angular distributions of their decay products have similar angular correlations. In particular, the azimuthal angule separation of the leptons ("1" and "2" in Figure 2.5) $\Delta\varphi^{\ell\ell}$ has a strong peak at π (cf. Figure 4.3 in Section 4.1 and Figure 6.4 in Section 6.5.3.)

The lepton centrality ζ is a measure of how central the charged leptons are with respect to the tagging jets. Scattered bosons are expected to have a smaller absolute



(a) r-Z Schematic



(b) r-φ Schematic

Figure 2.5: *Schematic of a vector boson scattering event in a generic collider detector. "1" and "2" are charged leptons from vector boson decays. "3" and "4" are tagging jets (see text). The angular correlations* $\Delta\varphi^{\ell\ell}$, ζ *and* $\Delta\eta^{\text{tagjets}}$ *are inscribed for reference.*

pseudorapidity than the tagging jets. Analogously for their daughter leptons. A continuous variable capturing this condition is the lepton centrality ζ.

$$\zeta \equiv \min\left\{\min\{\eta_1^\ell, \eta_2^\ell\} - \min\{\eta_1^{\text{jet}}, \eta_2^{\text{jet}}\}, \max\{\eta_1^{\text{jet}}, \eta_2^{\text{jet}}\} - \max\{\eta_1^\ell, \eta_2^\ell\}\right\} \tag{2.19}$$

Without loss of generality, let $\eta_1^\ell \leq \eta_2^\ell$ and $\eta_1^{\text{jet}} \leq \eta_2^{\text{jet}}$. Then the definition of ζ simplifies to

$$\zeta = \min\{\eta_1^\ell - \eta_1^{\text{jet}}, \eta_2^{\text{jet}} - \eta_2^\ell\} \tag{2.20}$$

If both leptons are within the tagging jets, ζ is positive and equals the minimum $\Delta\eta$ of a lepton to a jet. If one or both are outside, ζ is negative and its absolute value is the largest $\Delta\eta$ of any outside lepton to a jet. For VBS events, this value tends to be larger than for many backgrounds (cf. Figure 4.3 in Section 4.1).

The two neutrinos almost certainly leave the detector without interaction. While the momentum they carry away is a useful characteristic of the signal, they also make the reconstruction of the invariant diboson pair mass $m^{\ell\nu\ell\nu}$ impossible. However, the angular correlation variables ζ and $\Delta\varphi^{\ell\ell}$ depend on the coupling g of any resonance included in the model and help mitigate this disadvantage.

Additional information on event selection is given in Section 6.

3 Experiment

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) [49] is a 14 TeV center of mass energy proton-proton collider located in the former tunnel of the Large Electron-Positron Collider (LEP) experiment. Its circumference is 27km. Figure 3.1 shows a schematic of the whole accelerator complex. Along its lengths it is instrumented with 1232 superconducting dipole magnets to keep the 7 TeV proton beams in their trajectories. Beams of the same particles and hence charge necessitate two antiparallel magnetic dipole fields. The dipole magnets have two bores with superconducting coils held together by a common yoke. Superconducting liquid helium keeps the magnets at 1.9K to allow peak fields of 8.33T.

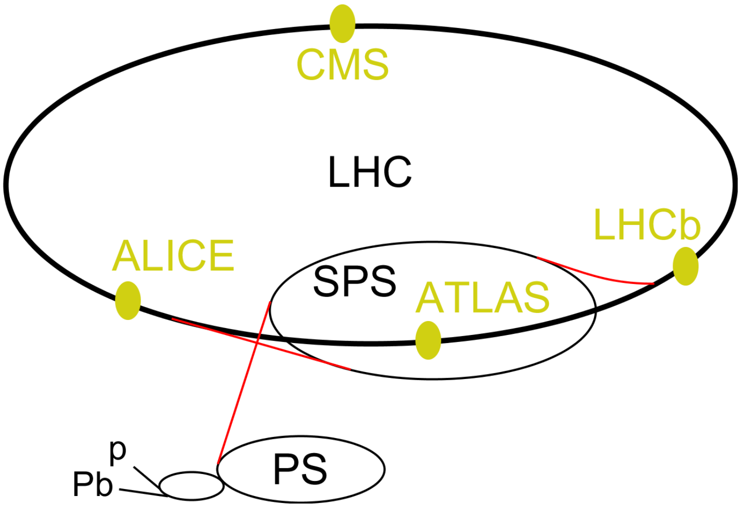


Figure 3.1: *Schematic of the LHC accelerator with pre-acceleration complex (from [50].) p/Pb: Proton/Lead linacs. PS: Proton Synchrotron, SPS: Super Proton Synchrotron, ALICE/CMS/LHCb/ATLAS: Experiments at interaction points.*

Eight superconducting radio frequency cavities per beam perform the main acceleration within the LHC ring. The mass of the proton allows much higher energies to be reached than the electrons of LEP whose synchrotron radiation makes this unfeasible with current technologies.

The instantaneous luminosity LHC was designed for is $\mathscr{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ which is expected to be reached after a couple of years of operation. With a bunch spacing of 25 ns, on average 23 proton-proton collisions will be in each bunch crossing.

As the LHC extends the current limits of technology and charts previously unknown territory, not all obstacles can be foreseen. After successfully recovering from a destructive magnet failure in 2009, at the time of writing the LHC is operating at 7 TeV center of mass energy with a peak luminosity of $\mathscr{L} = 10^{30}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

3.2 A Toroidal LHC Apparatus

A Toroidal LHC Apparatus (ATLAS) [51] is one of two large multi-purpose detector experiments located at the LHC interaction points. Figure 3.2 shows a cut-away view highlighting the main detector components.

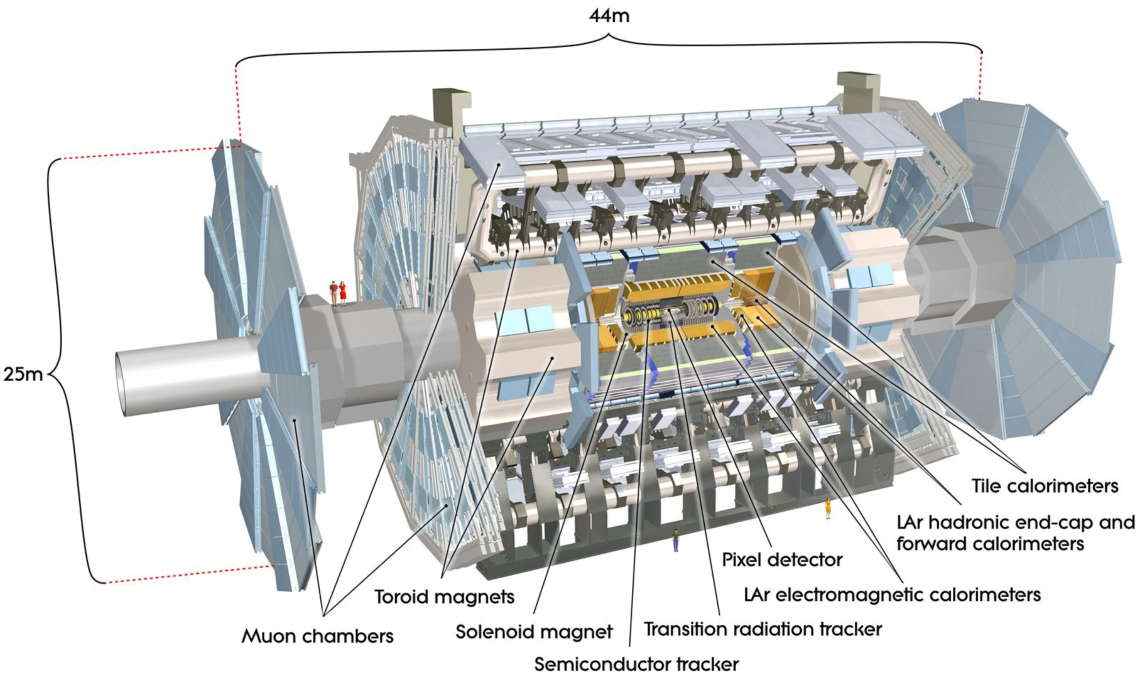


Figure 3.2: *Cut-away schematic view of the ATLAS detector (from [51].)*

The overall design is a typical barrel shape, with very large acceptance to ensure good $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, jet finding and tracking up to small angles relative to the beam axis. Most conspicuous is the muon system with its eponymous toroidal superconducting magnetic coils on the outside of the barrel and in the end caps. A superconducting solenoid encloses the inner detector.

With 25 m height and 44 m length it is the largest of the LHC experiments. A 25 ns bunch spacing corresponds to approximately 7.5 m at the speed of light which results in particles from two to three bunch crossing being in the detector at any time during normal operation.

3.2.1 Coordinate System

Typical coordinates used in a hadron collider experiments are pseudorapidity η and the azimuthal angle φ, which have already been introduced in Section 2.5. Pseudorapidity is a useful variable as it is closely related to rapidity and for massless particles actually identical. While the boost of the interacting partons is not known, differences in rapidity and hence approximately in pseudorapidity are Lorentz invariant.

Additionally, a Cartesian coordinate system is defined. The x and y coordinates are in a plane transversal to the beam axis. Positive x points towards the center of the ring, y points upwards and z is oriented along the beam axis to complete a right-handed system.

Two other useful quantities are the distance measures $\Delta\varphi$ and ΔR in $\eta \times \varphi$ space. Consider two coordinates (η_1, φ_1) and (η_2, φ_2). The distance in η is trivially $\Delta\eta \equiv \eta_1 - \eta_2$. The azimuthal angle φ has a period of 2π, however. Accordingly $\Delta\varphi$ is defined as

$$\Delta\varphi \equiv \varphi_1 - \varphi_2 + n\pi, \text{ where } n \in \mathbb{Z} \text{ such that } \Delta\varphi \in [-\pi, \pi) \tag{3.1}$$

In this text, if not otherwise noted, $\Delta\varphi$ always denotes the absolute value. A two dimensional distance ΔR may be defined analogous to the common Euclidean distance.

$$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\varphi^2} \tag{3.2}$$

Transversal coordinates play a particular role at hadron colliders, as the initial boost of the partons is unknown. Instead of the momentum p, often therefore the transversal momentum p_T is considered:

$$p_\mathrm{T} \equiv \sqrt{p_x^2 + p_y^2} \tag{3.3}$$

3.2.2 Inner Detector

The Inner Detector (ID) is closest to the interaction point inside the 2T of the superconducting solenoid (Figure 3.3.) Its main purpose is particle tracking and measurement of charged particle momentum from the trajectory in the magnetic field. For each design luminosity collision, about 1000 particle tracks are expected in the tracker.

Pixel Detector The innermost detector is produced in pixel technology to achieve high resolution and low occupancies. Three cylindrical layers make up the barrel region which is capped at both ends with three disks each. Coverage is up to $|\eta| = 2.5$ defining the limit of inner detector lepton identification. Within that range, the pixel detector provides potentially three high precision points per track. The

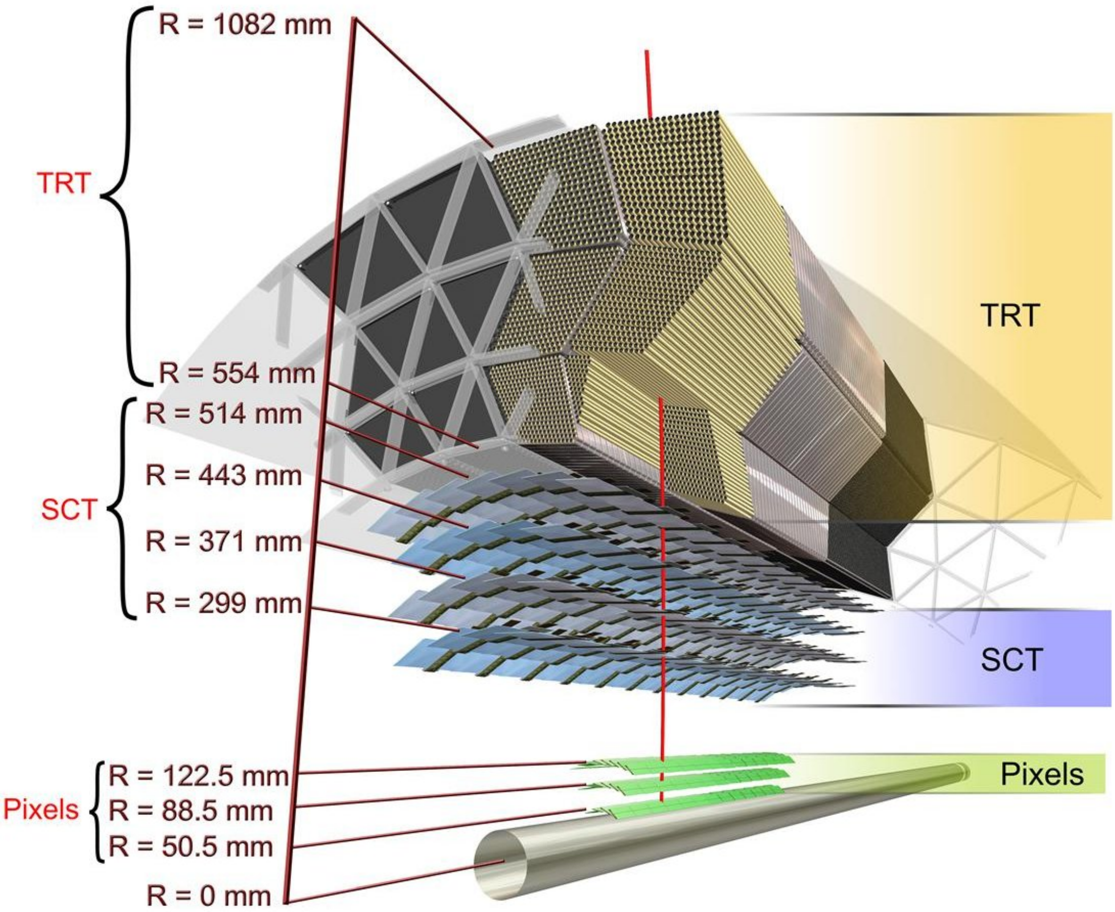


Figure 3.3: *Cut-away schematic view of the ATLAS Inner Detector (from [51].) A single high-energy particle is traversing all subdetectors.*

innermost layer is as close as possible to the beam line to achieve optimal resolution for vertexing. Flavor tagging in particular benefits from this, but also secondary vertex reconstruction to suppress background from pileup.

1744 modules contain 47232 pixels each, which are slightly reduced to 46080 read out channels per module due to combined channels at the edges of chips, totaling more than 80 million channels. This corresponds to about half of the total read out channels of the whole detector, even though the Pixel Detector is the most central and hence smallest subdetector. Most pixels have a pitch of 400μm in the beam direction z and 50μm in the transversal direction φ.

A flip-chip method is used to connect the read out chips directly on the back of the sensor chips with bump bonds. The large amount of electronics needed on the whole area of the detector leads to a large power consumption per volume. This necessitates an appropriate cooling capacity to carry off excess heat. To keep dark currents and noise down even for irradiated modules that need increased depletion voltage, the temperature is kept at −5 to −10°C, further increasing the demands on the cooling system. As the modules and support consist of material in which particles may undergo multiple scattering, material reduction and performance of inner detector components have inherent trade-offs.

Silicon Microstrip Tracker Next in r to the Pixel Detector is another silicon based detector, the Silicon Microstrip Tracker (SCT). It is made up of four cylindrical layers and nine disks at each end. Coverage is up to $|\eta| = 2.5$. Its purpose is analogous to the pixel detector, supplying four additional points per track typically. The two subdetector share cooling.

Due to its larger radius, occupancy is naturally lower per area, allowing the use of one dimensional silicon strips for measurement of (r, φ) in the barrel and (ϑ, φ) in the discs. To reconstruct the missing coordinate, each layer is subdivided into two layers of modules installed at an angle of ≈ 40mrad to one another, such that two hits from adjacent modules define a space point. 15912 modules of 768 strips each make up the active components of the detector.

Transition Radiation Tracker The outermost part of the inner detector is the Transition Radiation Tracker (TRT). Gas-filled straw tubes of 4 mm diameter contain a gold-plated tungsten wire to pick up the gas ionization signature of passing particles. Drift time accuracy is $\approx 130\mu$m. Varying dielectric coefficients at material boundaries in the detector trigger additional transition radiation. Electrons result in a particularly strong radiation which is detected with a higher signal threshold in the readout, allowing improved electron identification. The TRT cover extends to $|\eta| = 2.0$ and registers typically seven to ten high-threshold hits per electron above 2 GeV/c.

This technology is much cheaper than silicon and at the same time it is less dense, reducing the amount of material in the inner detector compared to a full silicon system. On the other hand, it cannot cope with the same amount of tracks as a silicon detector, as larger active subcomponent areas lead to a higher occupancy per readout channel.

3.2.3 Calorimetry

Calorimeters measure the total energy of particles, ideally stopping them completely. Whereas the ideal tracker should be massless, calorimeters need a lot of material to minimize leakage. Calorimeter in ATLAS sample the particle showers, interleaving active detector and passive absorber components. They are not compensating, which necessitates a calibration between electromagnetic and hadronic response.

The main barrel calorimeters are located outside of the solenoid magnet containing the inner detector. Energy loss in the coil is corrected for by a presampler just inside. A cut-away view of the calorimeter system is shown in Figure 3.4.

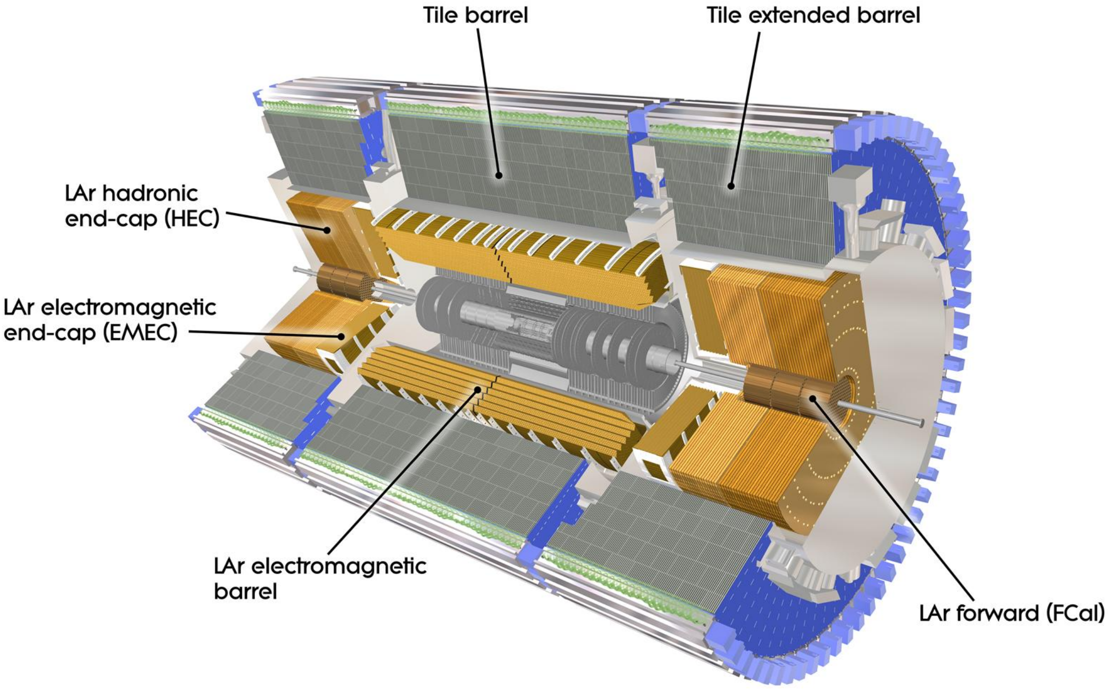


Figure 3.4: *Cut-away schematic view of the* ATLAS *calorimeter system (from [51].) The Inner Detector is drawn in gray.*

LAr Electromagnetic Calorimeter The LAr Electromagnetic Calorimeter (EMCal) uses liquid argon as active detector medium. It is has an accordion geometry, with alternating absorber and electrode waves (cf. Figure 3.5.) This leads naturally to a very uniform response in φ. By changing the folding angle and wave amplitude, the liquid argon gap can be kept constant in r. Both this geometry and the use of a liquid material lead to a very good overall uniformity. Using liquid argon makes the technology very radiation hard at the expense of needing a cryostat.

As the name implies, electromagnetically interacting particles are usually stopped, i.e. their showers contained in the EMCal. Hadronically interacting particles tend to leave most energy in the adjacent hadronic calorimeters. Shower shapes are another important distinguishing feature of electromagnetic particles, wherefore the EMCal has a particularly high granularity down to $\approx (0.025 \times 0.025)$ in $(\Delta\eta \times \Delta\varphi)$.

Overall, depending on η, two to three module layers cover up to $|\eta| = 3.2$. The barrel part ends at $|\eta| = 1.475$, the end cap (EMEC) begins with some overlap at $|\eta| = 1.375$.

Hadronic Calorimeters The next layer of ATLAS consists of the hadronic calorimeter (HCal) which stops particles that are only hadronically interacting. As such it

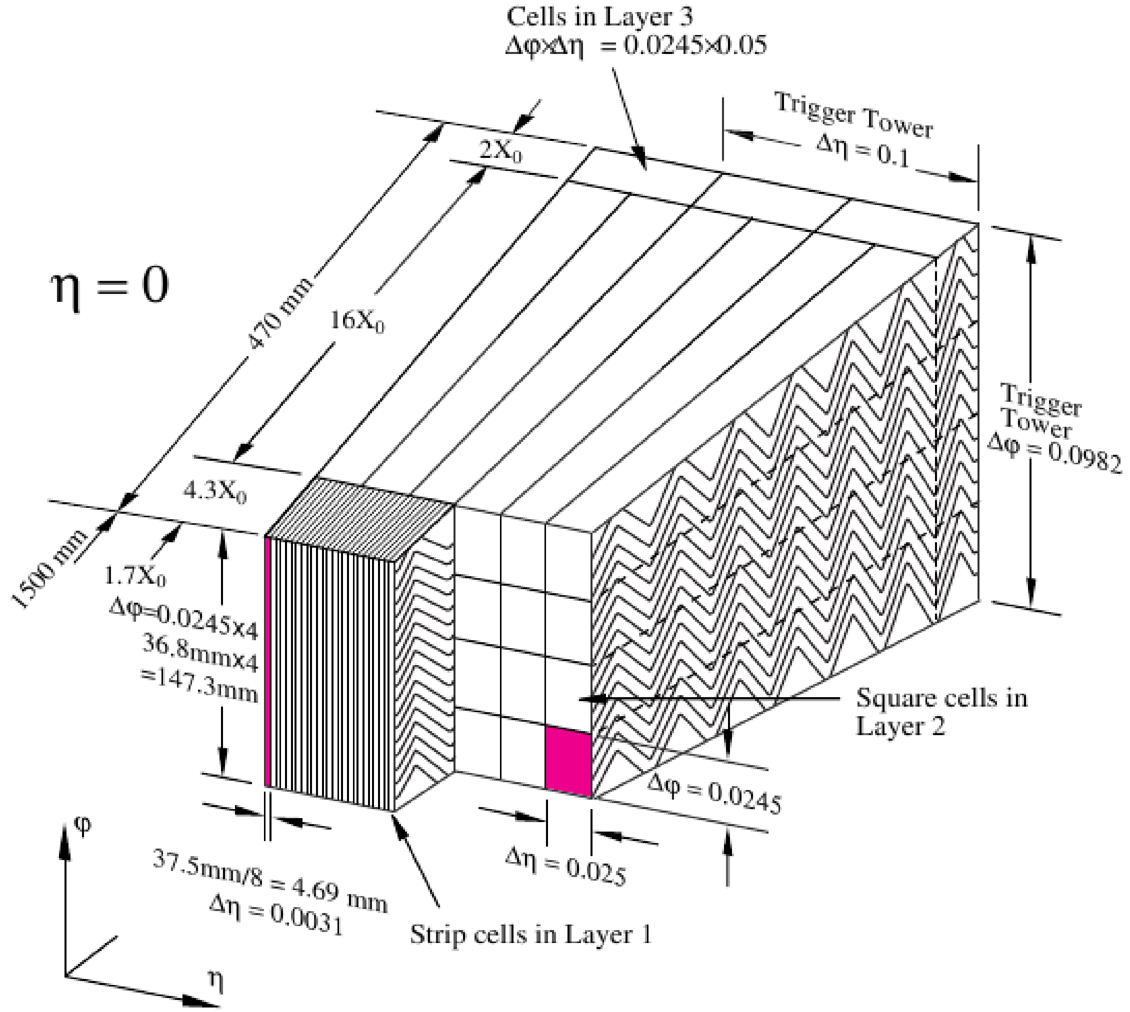


Figure 3.5: *Schematic of the accordion structure in the LAr Electromagnetic Calorimeter (from [51].)*

is very important for the measurement of hadronic jets. Only muons and weakly interacting particles should escape it on a regular basis. The barrel part is called Hadronic Tile Calorimeter (HTC) and is made of steel absorbers and plastic scintillator tiles. Scintillation light is read out with photomultipliers connected via fiber optics. For the LAr Hadronic Endcap Calorimeter (HEC) liquid argon technology was chosen to improve radiation hardness in the more exposed forward regions. It shares the cryostat with the EMEC and forward calorimeter.

The segmentation is chosen less fine than for the EMCal, reaching $\approx (0.1 \times 0.1)$ in $(\Delta\eta \times \Delta\varphi)$. The HTC has a coverage up to $|\eta| = 1.7$, the HEC covers $1.5 \leq |\eta| \leq 3.2$, matching the total EMCal coverage.

LAr Forward Calorimeter The Forward Calorimeter (FCal) has coverage from $3.1 \leq \eta \leq 4.9$. It will have to cope with the largest radiation doses due to its position and is therefore also situated in the liquid argon end cap cryostats. Heat dissipation is a particular problem for the modules facing the interaction point. Copper is used here as absorber to maximize heat transfer. Tungsten serves with its high density in the modules behind to contain the showers.

At this extreme forward position, angular and pseudorapidity differences translate to very small areas, making a segmentation in $(\Delta\eta \times \Delta\varphi)$ impractical. Instead, the FCAL is segmented in $(\Delta x \times \Delta y)$, with a minimum of about $0.75\,\mathrm{cm} \times 0.625\,\mathrm{cm}$.

3.2.4 Muon Spectrometer

Energetic muons are the only charged particle to routinely escape the calorimeters. The muon spectrometer measures their tracks in the outermost layers of the main ATLAS detector in the cavern. To get a second momentum measurement, muon tracks are bent in toroidal magnetic fields both in the barrel and end caps. Figure 3.6 contains an overview of the muon system.

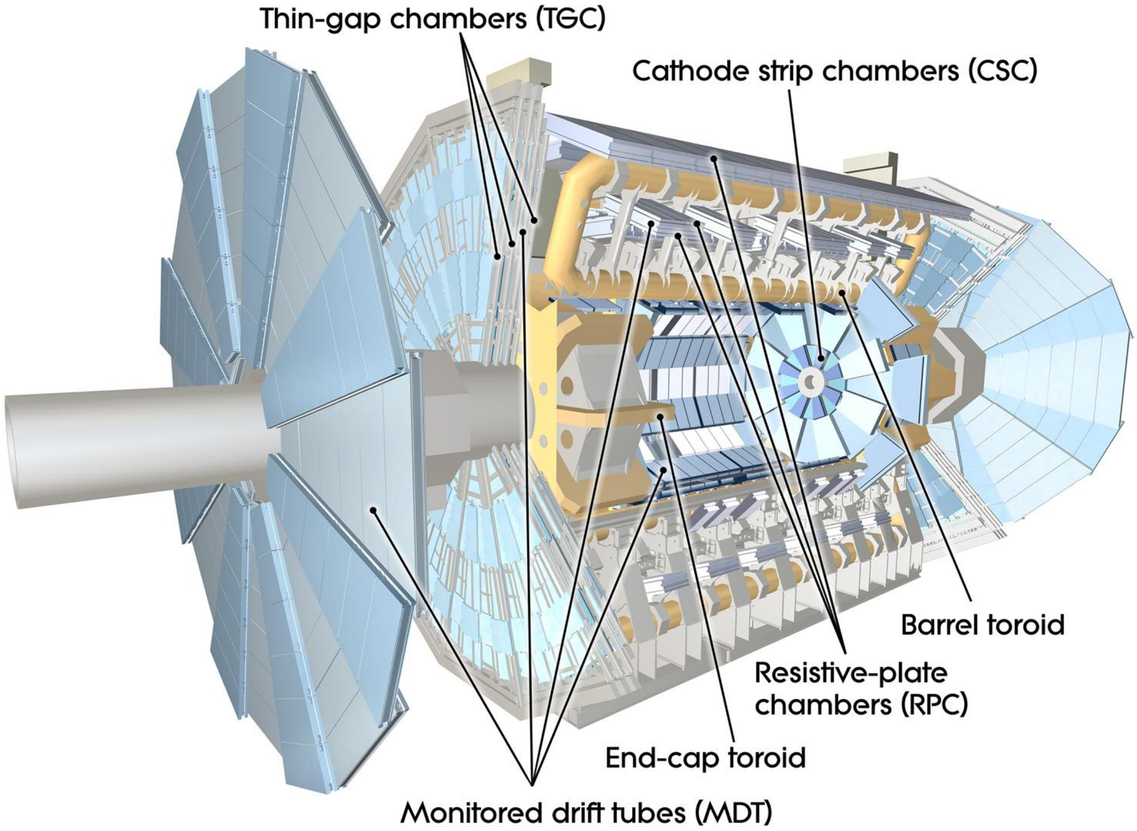


Figure 3.6: *Cut-away view of the ATLAS muon system (from [51].)*

Precision Tracking System Offline measurements by the muon system use information from the precision tracking system. Its overall coverage is up to $|\eta| \leq 2.7$. Most of that is instrumented with Monitored Drift Tubes (MDT) of about 3 cm in diameter. Particles passing through ionize the gas mixture of 93% Ar and 7% CO_2, leading to an avalanche amplification picked up by the tungsten-rhenium wire in the center. Per track about 20 measurements in the barrel and/or in the end caps are typical.

In the innermost forward layer in the range $2 \leq |\eta| \leq 2.7$, there is too much occupancy for drift tubes. The slow collection time degrades performance. Here, Cathode-Strip Chambers (CSC) are used. These multi-wire proportional chambers average only 4 measurements per track but have a faster time response and can cope with the higher rates.

Muon Trigger System In addition to the precision tracking, the muon system has a dedicated hardware trigger system with additional detectors. It is split into Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) in the end cap region because of the latter's better rate capabilities. A simple but fast hardware logic selects hits from potential energetic muons for triggering within the coverage of $|\eta| \leq 2.4$.

3.2.5 Luminosity Measurement System

Luminosity may be measured by multiple detectors in ATLAS. Two detectors outside of the main detector installation are primarily targeted at luminosity measurements. The Luminosity measurement using Cerenkov Integrating Detector (LUCID) is placed 17 m from the interaction point. It measures inelastic proton-proton collisions at low angles detecting Cerenkov light in its tubes filled with C_4F_{10}.

The Absolute Luminosity for ATLAS (ALFA) detector is located at 240 m from the interaction point to measure forward elastic proton-proton scattering. This process is well-understood theoretically and connected to the total cross section via the optical theorem. ALFA uses scintillating fibers as active material. To achieve an accuracy below ≈ 5% in the long run, the detector has to be placed very close to the beam to match a scattering angle of $3\,\mu$rad from the interaction point. This is within normal LHC beam emittance and only possible in runs with special beam optics. The detector is placed within roman pots connected to the beam pipe and may be moved close for measurements.

3.2.6 Trigger System

Due to technical and financial constraints it is unfeasible to write all of the 40 million events per second to mass storage for off-line analysis. Neither the storage capacity nor the bandwidth are nearly sufficient. At a hadron machine, the ratio of interesting events to background events is very large. The rarity of many types of events is the reason for the large luminosity target of the LHC in the first place. A trigger system that only writes out preselected events for later analysis reconciles the luminosity requirements with computational constraints.

At the end of the readout chain events may be written to mass storage at a rate of up to 200 Hz. These are selected in the 3rd level of the trigger system after standard offline reconstruction on commodity hardware. This is the event filter which works technically like a precut in an analysis. Average latency is four seconds.

The event filter can only accept an event rate of about 3.5 kHz itself. Still on commodity hardware, a reduced reconstruction is performed with simpler algorithms on a reduced part of the detector information. This happens in the 2nd level of the trigger (L2) with an average processing time of only 40 ms.

The regions of interest (ROI) which are read out and form the bases of the L2 trigger decision are selected at the lowest level trigger (L1). Available readout bandwidth limits the L1 rate to 75 kHz upgradeable to 100 kHz. The maximal admissible latency is $2.5\,\mu$s, otherwise an event is discarded. The L1 trigger is implemented in specialized hardware. In the case of the muon system, this includes active detector subsystems as mentioned above. Mostly the added hardware is using information from detectors already used for normal readout. All detector systems must, however, include buffers in the front-end electronics to cache event information until the L1 trigger decision arrives. This buffer length defines the admissible latency.

3.3 Object Reconstruction

Measurements of the ATLAS detector as of any other multi-purpose detector at a collider do not translate one-to-one into particle four-momenta which are of interest for the theoretical evaluation. Instead, particles have to be reconstructed from the raw data read out from the millions of detector channels.

Basic final state objects in the detector are reconstructed using algorithms that are centrally defined for all of ATLAS. This ensures a common baseline but is also necessary to keep the amount of data to be distributed for analysis manageable. Baseline reconstruction algorithms for the objects used in this thesis are introduced in this chapter. These are electrons, muons and hadronic jets. Neutrinos would also be of interest, but as weakly interacting particles, they do not register in the detector. A useful quantity called missing transverse energy or $E_{\mathrm{T}}^{\mathrm{miss}}$ in the context of hadron colliders can give some information about neutrinos.

3.3.1 Hadronic Jets

Quarks or gluons in the final state of a hard process like the two quarks in $qq \to qq\ell\nu\ell\nu$ are confined by the strong interactions. Though charged, they do not produce a track in the detector as they never reach active material as free quarks. Instead, they hadronize producing color singlet states which can can propagate. Additional decays and QCD Bremsstrahlung lead to a multitude of particles that are produced from a high-p_{T} final state quark or gluon. They tend to be collimated around the original momentum of the colored particle which gives the conglomerate an experimental signature. Taken together they make up what is called a jet of particles.

The charged particles in a jet produce tracks that may be found in the tracker. In the calorimeters clusters of energy can be reconstructed. For a hadronic jet of reasonably high energy a large fraction of the energy will be found in the hadronic calorimeter.

Experimentally, the definition of a jet cannot be unequivocal, because it is not possible to clearly determine if a particular track or cluster corresponds to a particular parton. Instead, several pragmatic definitions are available that collect energy deposits that are close to one another in momentum or position space. What an algorithm finds is per definition an experimental jet. Here, a seeded cone algorithm is used which operates on calorimeter towers.

Towers in ATLAS are (0.1×0.1) in $(\Delta\eta \times \Delta\varphi)$ logical regions of the calorimeter. Energy is summed in the cells contained in any tower. When tower edges cut through cells, they are summed proportionally. Noise in the calorimeters may lead to negative tower energies. As this must be unphysical and is not foreseen in the algorithms, they are summed with neighboring towers until only positive entries remain. Each tower is assigned a four-momentum corresponding to a massless particle of the tower's energy.

The algorithm first looks for seed towers with more than 1 GeV of energy. Around each seed in a cone of $\Delta R = 0.4$ the tower four-momenta are added. If the total momentum direction deviates too much from the seed tower, it is taken as a new cone axis and the process repeated until stable cones are found. The jets may acquire mass due to the addition of massless four-momenta.

Resulting cones may overlap. If their overlap in transversal energy exceeds 50% they are merged, otherwise the towers are assigned uniquely to the jet with the closest axis.

On generator level, the algorithm is the same, but cannot operate on towers. It uses stable final state particles instead, except those that would not register in the detector.

3.3.2 Electrons

Electrons produce a track and a collimated energy deposition in the electromagnetic calorimeter. For electrons with energies above $20 - 30$ GeV, calorimeter information is not improved by the p_{T} measurement from the track.

The algorithm is seeded using a sliding window algorithm with an aperture of 5×5 cells in the middle layer of the EMCal. From the seed a fixed size cluster is reconstructed. In the barrel, the cluster is 3×7 in the middle layer. For the endcaps the cluster size is 5×5.

Reconstructed energy is corrected for dead material. Electron candidates are saved and may be subjected to further quality cuts in the selection. These include E/p ratio, shower shape variables, lateral and longitudinal profiles and the quality of the match to the Inner Detector track.

3.3.3 Muons

Muons in the energy range interesting for this analysis are minimally ionizing. This immediately has the consequence that their energy cannot be measured well in the calorimeter alone. Instead, muon reconstruction mostly relies on track finding in the Inner Detector and for high-p_T muons in the Muon Spectrometer. The latter is dominating the resolution for muons over $30\,\mathrm{GeV}/c$. For the very highest momenta, performance degrades, as the sagitta to measure p_T comes closer to the MDT track resolution.

For combined muons a track has to be found both in the ID and the muons system. This limits the coverage to the $|\eta| \leq 2.5$ of the ID. A service gap at $\eta \approx 0$ and at the feet of the detector introduces some inefficiency in the muon system.

3.3.4 Missing Transverse Energy

Some weakly interacting particles, i.e. SM neutrinos may escape undetected. This should show up in the total process momentum which vanishes in the lab frame. However, at a hadronic machine the proton remnants typically escape down the beam pipe, making total momentum inaccessible. The final state of the hard process is often measured, but the initial state has an unknown boost in z. This leaves transverse momentum. The transverse vector needed to balance p_T is called missing transverse momentum $p_\mathrm{T}^\mathrm{miss}$.

It is much more exact to estimate $p_\mathrm{T}^\mathrm{miss}$ from the calorimeter rather than adding up track p_T. The measured quantity then is an energy which engenders the term transverse energy and missing transverse energy $E_\mathrm{T}^\mathrm{miss}$. This is different from the missing energy variable at a lepton collider where initial and final states may be reconstructed more completely. In this text, $p_\mathrm{T}^\mathrm{miss}$ and $E_\mathrm{T}^\mathrm{miss}$ will be used interchangeably, either emphasizing the theoretical or experimental quantity.

$E_\mathrm{T}^\mathrm{miss}$ is taken from the sum of all energy clusters in the calorimeters. Several corrections are applied. Firstly, muons are minimally ionizing but may carry off a large amount of momentum when produced in the hard process. Energy loss in the cryostat needs to be taken into account. Its effect may reach 5%. Finally all cells are correlated to reconstructed objects, with individual corrections applied for each species.

Measurement of $E_\mathrm{T}^\mathrm{miss}$ is one reason why hermetic calorimetry has been an important goal in the design of ATLAS.

3.3.5 Flavor Tagging

To distinguish between jets from light quarks or gluons, one can use the life time of B hadrons formed from b quarks or similarly for c quarks. A typical decay length for a B hadron is on the order of $400\,\mu\mathrm{m}$. Taking the boost into account, the flight path in the lab frame is on the order of some millimeters.

Methods to tag jets according to the original quark flavor use the fact that tracks in the jet tend to come from the decay of the B hadron. They do not point to the primary vertex but to a secondary vertex some way off.

The first step is to select tracks of sufficient quality. Seven precision hits in the silicon part of the tracker are required, two of these in the pixel detector, one in the inner pixel layer. Track transverse momentum has to exceed $1\,\mathrm{GeV}/c$. The impact parameters d_0 and z_0 at the point of closest approach to the primary vertex must not exceed $|d_0| = 1\,\mathrm{mm}$ in transverse direction and $|z_0 - z_v| \sin\vartheta$ in longitudinal direction, where z_v is the primary vertex position and ϑ the polar angle of the track.

Impact parameters are signed, negative signs signifying impacts behind the primary vertex. Negative values are not physical but may result from resolution effects. The impact parameter distribution differs for light jets and those containing b-quarks. The latter tend to higher values. The IP3D tagger uses both impact parameters for discrimination.

This may be refined by using explicit secondary vertex reconstruction for the subsequent decay of B hadrons. From the tracks that have a high impact parameter significance, pairs forming good vertices are selected. The invariant mass is used to reject K^0_S, Λ, hyperons and conversions.

The combined tagger IP3D+SV1 used impact parameter information and secondary vertex reconstruction into one weight. While for early data such an advanced tagger may be problematic, in the time frame of this analysis it seems a reasonable choice.

4 Monte Carlo Generators

In order to actually make predictions for what can be seen in the experiment given a certain model, collision events have to be simulated. From the perturbation theory point of view, all begins with a hard interaction. In a proton-proton machine like the LHC, as in any hadron collider, the hard interaction is not immediately accessible. Instead, one has to first move from the proton to the parton level. Only the initial state of two colliding protons are well-known in the machine: 7 TeV energy each in the laboratory frame. As the configuration is symmetrical, the center of mass frame is the laboratory frame and the sum of energies gives the total center of mass energy of 14 TeV. At the small distances probed with these energies, partons are clearly visible in the processes. The initial state of interacting partons is unknown, due to the complexity of the proton. Parton distribution functions (PDF) give their longitudinal momentum distribution as a function of the momentum transfer in the hard process. An event generator will have to pick initial state partons for the desired hard process from these distributions. At this point, the hard interaction itself may be simulated using knowledge of the appropriate matrix elements. Most processes can only be simulated at leading order, some very interesting ones are available at higher orders of perturbation theory. To simulate additional radiation, a parton shower model modifies the results from the hard interaction. Depending on the setup some particles may decay if this has not been absorbed into the matrix element itself. To get back from parton level to particles that are not confined, resulting partons need to be combined to hadrons in the hadronization step. At this point, the simulated event is still considered to be at generator level. Next, interactions in the detector and electronic response would have to be simulated.

4.1 Whizard and Pythia

The WHIZARD event generator [52, 53] is currently the only generator available that is able to simulate a K-Matrix model in vector boson scattering with additional resonances. It includes an automatic tree level matrix element generator to simulate the hard process. Parton showering and hadronization is done externally in other generators. In this thesis, PYTHIA [54] was used for this purpose. At the same time, PYTHIA contains a K-Matrix implementation in its own right according to References [55, 56] and served as a baseline comparison. It is not able to simulate additional resonances but anomalous couplings α_4 and α_5 (cf. Section 2.2) are available.

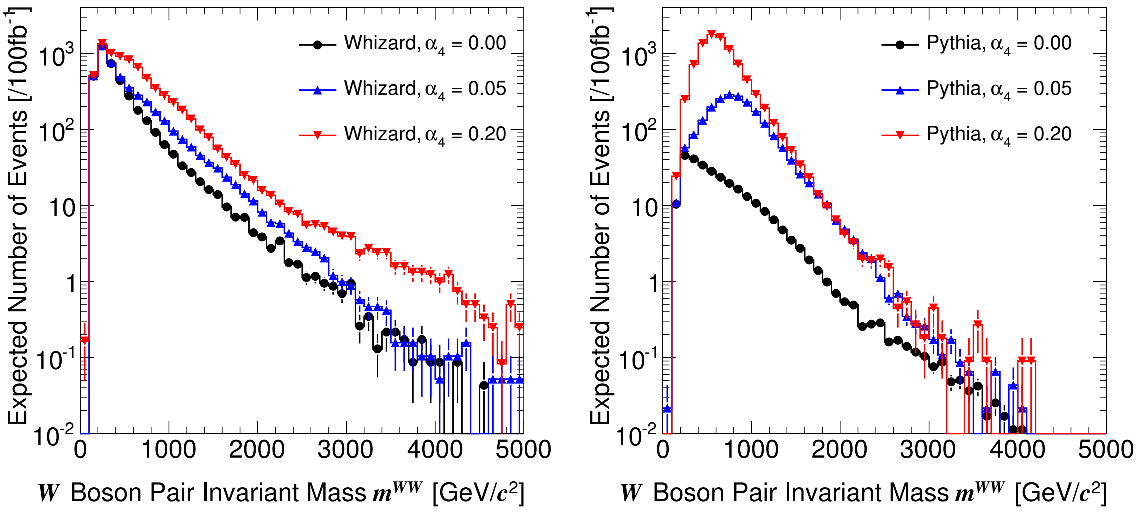
The generators simulate the hard process with different techniques. PYTHIA is designed as library of two-to-two processes, WHIZARD is a full matrix element generator. Vector boson scattering $qq \to qq\ell\nu\ell\nu$ is a two-to-six process. While this can directly be done in the matrix element by WHIZARD, PYTHIA needs to split this into several phases. The $WW \to WW$ process (ignoring Z bosons for the time being) is simulated using the matrix element. The W boson decay to leptons is factorized out. On the initial state side, the W are taken from the effective W approximation [57, 58, 59, 60, 61] (EWA). This is conceptionally similar to a W PDF in the proton.

These differences have a couple of consequences for the simulation. In EWA, the W bosons have to be on-shell. Also, information about angular correlations is lost in PYTHIA. The largest effect, however, is that WHIZARD generates all possible processes that connect the initial to the final state instead of just WW scattering. Many of these do not even have vector boson scattering topology. EWA and other effects are therefore intrinsically entangled in a realistic simulation. To see the effect of the EWA alone, WHIZARD has to be configured to also only simulate $WW \to WW$ scattering with EWA switched on or off (see also figures in Reference [34, p. 29].)

A comparison between WHIZARD's and PYTHIA's K-Matrix implementation was performed for this thesis using only anomalous couplings. In WHIZARD, the reduced process $qq \to qqWW$ was simulated. Large values of α_4 were chosen to highlight the effects. If such a model is realized in nature, the expected scale is closer to $1/16\pi^2 \lesssim 0.01$. Figure 4.1 shows the invariant WW pair mass, m^{WW}, on generator level for WHIZARD and PYTHIA. ROOT [62], IPYTHON [63] and PYTHON [64] were used to prepare these and other figures. Due to the presence of irreducible background in the latter generator's events, a complete agreement would have been surprising. In particular, an additional continuum contribution should be present in the WHIZARD samples. Total cross sections are therefore not expected to agree either. However, the additional contribution due to increased anomalous coupling should agree better. This is shown in Figure 4.2 where for each generator, the histogram at $\alpha_4 = \alpha_5 \equiv 0.0$ has been subtracted. Agreement is, if far from perfect, much better. Noticeable generator differences remain.

The behavior is different in the case of angular correlations. In Figure 4.3 the lepton centrality ζ and the azimuthal angle separation of the leptons $\Delta\varphi^{\ell\ell}$ are shown. With increased anomalous couplings, WHIZARD shows changing correlations, while the distributions for PYTHIA are completely unchanged. Both these values are experimentally interesting, as they are accessible in the double leptonic channel where m^{WW} cannot be reconstructed.

For the models under investigation in this thesis, WHIZARD is currently the only generator available. Its ability to preserve angular correlations make it also interesting for models that are currently implemented in PYTHIA.



(a) Whizard **(b)** Pythia

Figure 4.1: *Distributions of the invariant W boson pair mass m^{WW} for different values of the anomalous coupling α_4. Left: Histograms for* Whizard*. Right: Histograms for* Pythia*. The additional coupling adds to the total cross section for both generators but the shapes are different.*

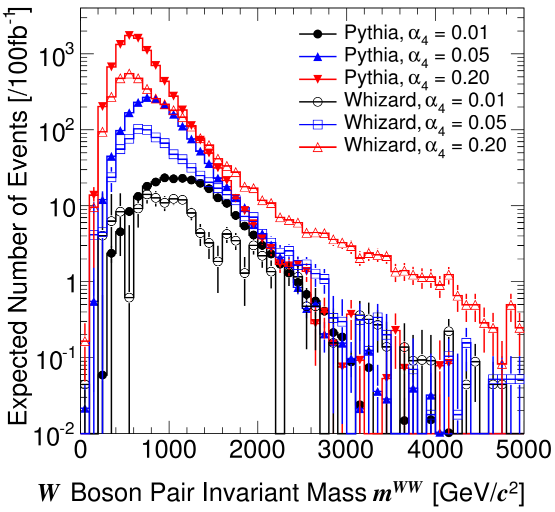
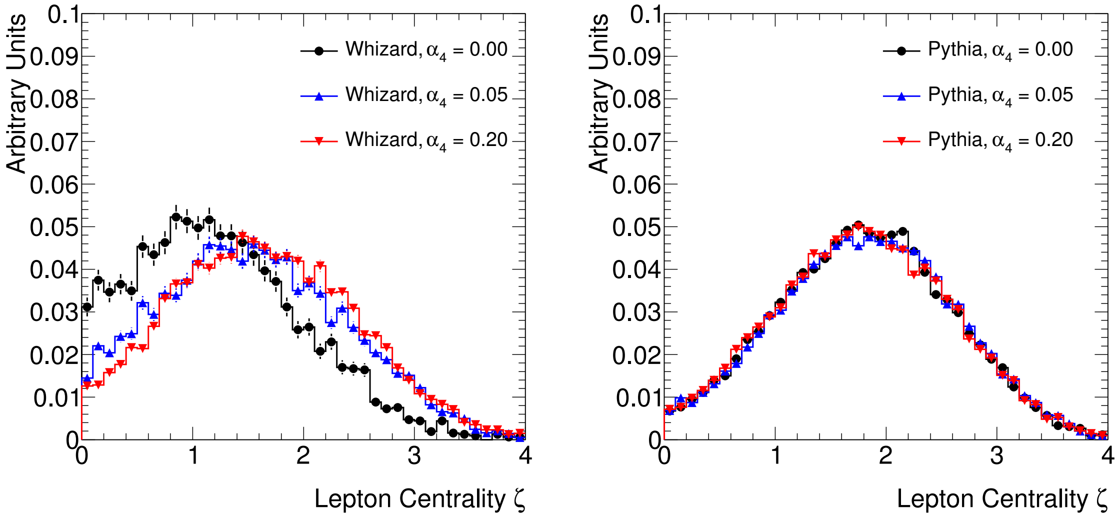
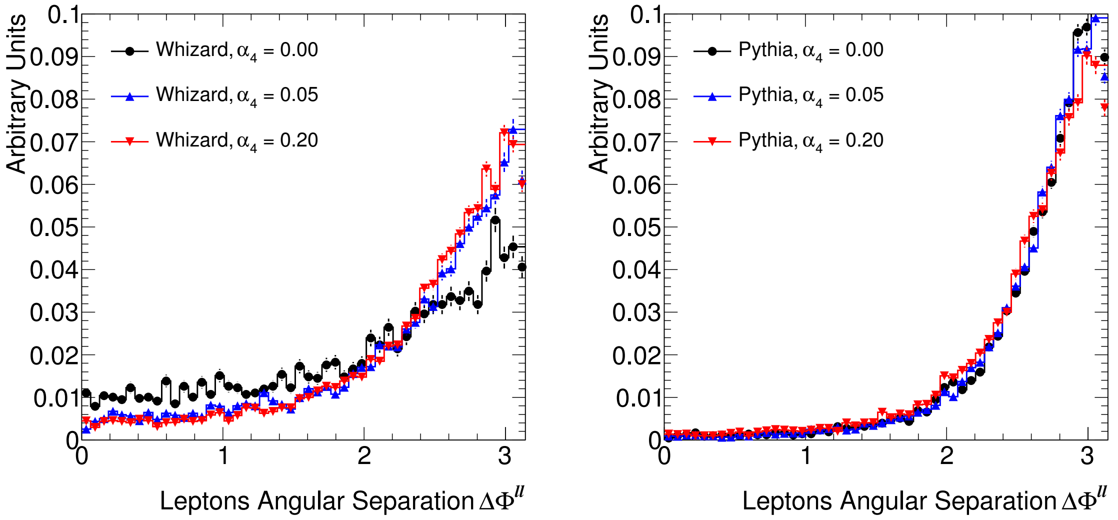


Figure 4.2: *Distributions of the invariant W boson pair mass m^{WW} for different values of the anomalous coupling α_4 after subtraction of the continuum at $\alpha_4 = 0$. Histograms for both generators are superimposed. Differences remain, but the shapes are much closer than without subtraction of $\alpha_4 = 0$.*



(a) Whizard ζ **(b)** Pythia ζ



(c) Whizard $\Delta\varphi^{\ell\ell}$ **(d)** Pythia $\Delta\varphi^{\ell\ell}$

Figure 4.3: *Angular correlation variables for different values of the anomalous coupling α_4. Histograms are scaled to unity to emphasize shape. Left: Plots for* Whizard*. Right: Plots for* Pythia*. While anomalous couplings change angular correlations in* Whizard*, this is not modeled in* Pythia*.*

4.2 Whizard Validation

4.2.1 Interface

WHIZARD had not been in official ATLAS use before the work on this thesis started. Originally it was targeted at the future International Linear Collider (ILC.) In order to be able to use it in ATLAS for central sample production it had to be validated and integrated into the ATLAS software. As WHIZARD was able to write out the hard process event information as an ASCII (text) dump of the Les Houches Accord (LHA) [65], it could be read in using interfaces for other generators. LHA is not well defined in ASCII, originally being specified only as a set of FORTRAN common blocks. As the latter are not easily stored and passed except within a single FORTRAN process, a plethora of slightly divergent LHA text formats had been used for different generators. With the advent of the Les Houches Event File (LHEF) [66] format, which uses an XML like markup language, a more standardized format had become available. To be future proof it was decided to port an existing interface for MADGRAPH to be used as a generic LHEF interface for ATHENA[1] in the context of the present thesis. This interface is now recommended for all generic LHEF input in ATLAS.

WHIZARD in the current version has a problem with writing out initial state quark flavors in LHEF. It is triggered when flavor summation is used for quarks. For example in the process $qq \to qq\ell\nu\ell\nu$ one might want to consider the three lightest quarks ($q \in \{u, d, s\}$). It is possible to configure WHIZARD to automatically perform the necessary flavor summation for matrix element calculation. When doing so, however, the wrong initial state quark flavors are written into the LHEF. PYTHIA, which is used for the parton shower implementation and hadronization, crashes because of inconsistent flavors. It is expected to be fixed in the next major release of WHIZARD. In the meantime, a workaround has been developed for use in this analysis, which uses WHIZARD's ability to add many processes to a single run. Instead of relying on WHIZARD's flavor summing automatism, an external program was developed to generate all combinations for an arbitrary list of quark (and lepton final state) flavors.

4.2.2 Quark Flavor Scaling

The manual workaround for flavor summation explained above has an undesirable side effect. While the phase space integration is done individually at subprocess level, for event generation all phase space results from that step have to be loaded into memory at the same time. Instead of generating the processes $uu \to qqe\nu_e e\nu_e$, $u\bar{u} \to q\bar{q}e\nu_e e\nu_e$, etc. at a time, all of them are generated together. With all flavors in the initial state and all lepton flavor combinations in the final state, this exceeded machine architecture memory limits for $q \in \{u, d, s, c\}$. The total cross section may be recovered, but individual events can only be generated with the three lightest quarks $q \in \{u, d, s\}$. Internally WHIZARD would be able to handle this much

---

[1] ATHENA version 14.2.25 was used throughout this thesis.

more intelligently, by using the redundancies in the phase space calculation. Even then, computation time would be immense. Splitting the processes allows the use of multiple computer cores in parallel.

It is desirable to recover information from the c quark, if possible. As all light flavors are approximated as massless, they are expected to yield identical distributions. Any difference would have to come from the PDF and combinatorics. This was tested for the reduced process $qq \to qqWW$ where event generation is possible for any considered number of quarks. Samples with four quark flavors ($q \in \{u, d, s, c\}$) and two flavors ($q \in \{u, d\}$) were compared. Results for the invariant w pair mass m^{WW} and the lepton azimuthal angle separation $\Delta\varphi^{\ell\ell}$ are shown in Figure 4.4. For the m^{WW} histogram, VBS cuts from Reference [41] were applied to reduce the continuum and focus on the resonance:

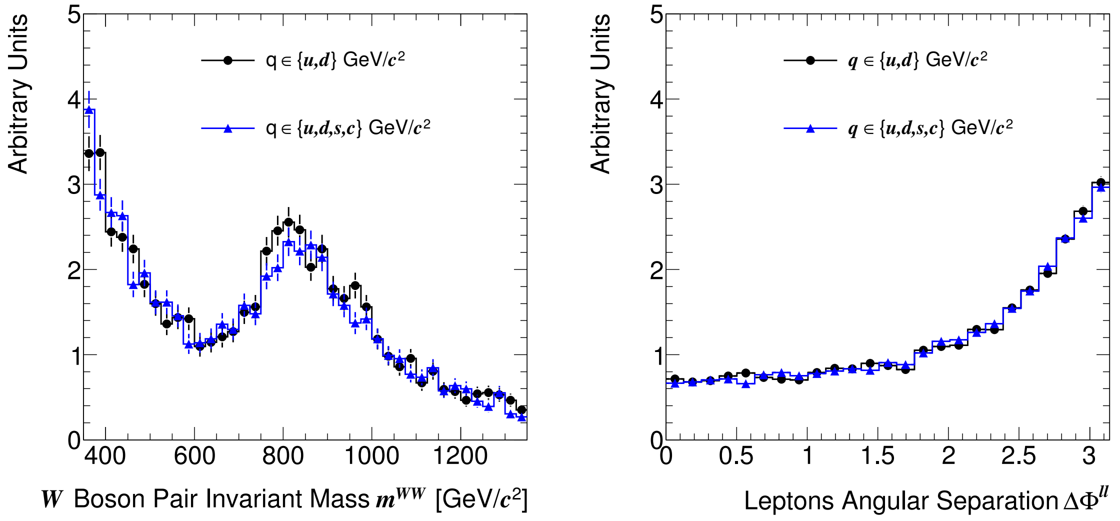
- two leptons with transverse momentum $p_{\mathrm{T}1,2} \geq 35.0\,\mathrm{GeV}/c$
- two tagging jets with
  - transverse momentum $p_{\mathrm{T}1} \geq 60.0$, $p_{\mathrm{T}2} \geq 24.0\,\mathrm{GeV}/c$,
  - energy $E_1 \geq 600.0\,\mathrm{GeV}$ and $E_2 \geq 400.0\,\mathrm{GeV}$
  - invariant mass $m^{\mathrm{tagjets}} \geq 1080.0\,\mathrm{GeV}/c^2$
  - pseudorapidity separation $\Delta\eta^{\mathrm{tagjets}} \geq 4.4$
- lepton centrality $\zeta \geq 0.0$

Differences in shape are within statistical uncertainties. Additional quarks only add to total cross section. In the analysis, therefore, the distributions generated for u, d and s quarks were scaled to the cross section estimated from u, d, s and c. Samples used in Sections 4.1 and 4.2 contain u and d quarks unless otherwise noted.

4.2.3 Process

To validate the signal process in WHIZARD, a standard to compare against had to be found. A validation of the formulaic correctness had already been done by the developers of the generator. For use in the experiment, validation had to be extended to the whole generation chain including experimental software up to but not including the simulation of the detector. Technically this is an integration test of the generator and experimental software. Moreover, apart from model availability, differences between generators due to their overall design have to be under control.

As there is no implementation of the model under investigation other than WHIZARD, a boot-strapping approach to testing was necessary. A heavy SM Higgs boson may be generated both with PYTHIA and WHIZARD giving a common starting point. However, a lot of the generator differences, which are model independent should become visible in this process as well. To illustrate this, the distribution of the invariant mass of the WW boson pair is plotted for both generators in Figure 4.5a. WHIZARD has been restricted to the process $qq \to qqWW$, which forces the final state W bosons on

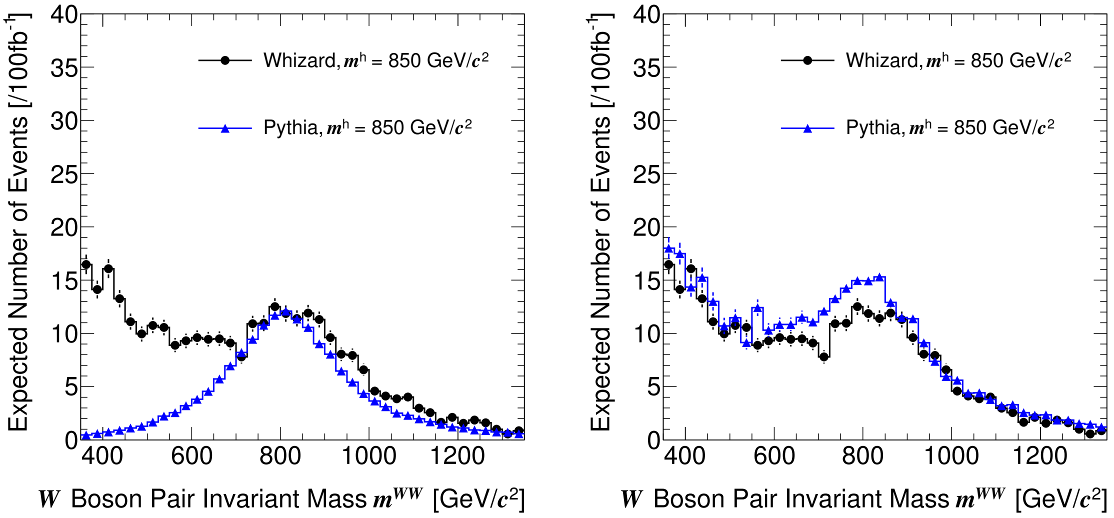


(a) m^{WW} **(b)** $\Delta\varphi^{\ell\ell}$

Figure 4.4: *Distributions of the invariant W pair mass m^{WW} and the lepton azimuthal angle separation $\Delta\varphi^{\ell\ell}$ for the process $qq \to qqWW$ in* Whizard, *$W \to \ell\nu$ in* Pythia. *Compared are the quark sets $q \in \{u, d\}$ and $q \in \{u, d, s, c\}$. Histograms are normalized to a common area. Within statistical uncertainties the shapes agree.*

the mass shell just as in the case of Pythia. The mass of the Higgs boson was set to $m^h = 850\,\text{GeV}/c^2$. While the mass peaks are at the expected position, Whizard has an additional continuum contribution for irreducible background. This continuum may also be generated independently, for example using a K-Matrix model without additional resonances or anomalous couplings. By adding it to the pure resonance events from Pythia the plots in Figure 4.5b were derived which agree much better. Remaining differences include Whizard still generating a two-to-four instead of a two-to-two process, which allows any intermediate W bosons to be off-shell.

The K-Matrix formalism has the nice property that it contains other models as special cases. Among those is the SM with a Higgs boson. A scalar isoscalar resonance σ with a coupling strength of $g = 1.0$ behaves like a Higgs boson of the same mass. To test the K-Matrix implementation against the better studied SM implementation, a heavy Higgs boson of $m^h = 850\,\text{GeV}/c^2$ was compared to an equally heavy K-Matrix σ resonance within Whizard. As now the same generator chain is used, the differences between the models themselves should become apparent. In the course of this test, it was found that the two models are not in fact absolutely identical. The WW mass distribution for the two models is shown in Figure 4.6. While the resonance is clearly visible in both cases, the differences in lineshape are too large, given that the same generator is used. An explanation was found in the K-Matrix unitarization. The σ resonance at $g = 1.0$ behaves just like a Higgs boson in terms of couplings and Feynman rules. Unitarization is still applied, though. This is desirable



(a) Raw Output **(b)** Continuum Added to Pythia

Figure 4.5: *Distributions of W boson pair invariant mass m^{WW} (with VBS cuts) for SM with a Higgs boson at $m = 850\,\mathrm{GeV}/c^2$. PYTHIA and WHIZARD differ widely on the left. After adding a WHIZARD K-Matrix continuum to PYTHIA on the right, agreement is more reasonable.*

in so far as consistency within the parameter space of the K-Matrix implementation is concerned, but leads to an additional phase in the resonance. This in turn changes the lineshape from a Breit-Wigner distribution with a constant width to a Breit-Wigner with an s-dependent width (cf. Equation (2.17)).

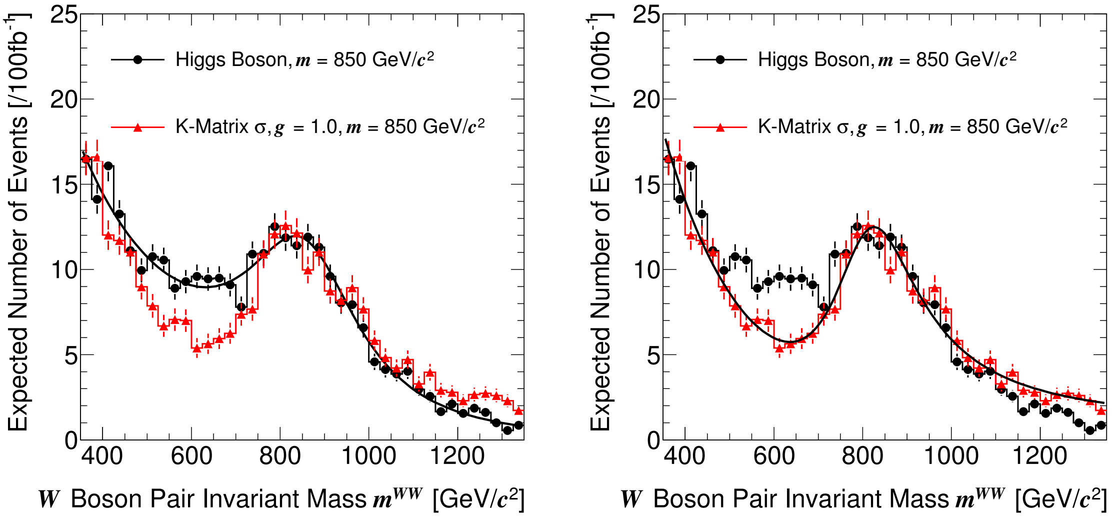
$$\sigma^{\text{SM VBS}} \propto \frac{1}{(s-m^2)^2 + m^2\Gamma^2} + c\exp\ldots \tag{4.1}$$

$$\sigma^{\text{K-Matrix VBS}} \propto \frac{s^2}{(s-m^2)^2 + m^2\Gamma^2\frac{s^2}{m^4}} + c\exp\ldots \tag{4.2}$$

Using the correct Equations (4.1) for SM and (4.2) for K-Matrix plus an exponential term for the continuum, the generator output was well-described. Fits are shown in Figures 4.6a and 4.6b.

4.2.4 $m^{\ell\nu\ell\nu}$ Distributions for Example Resonances

One advantage of WHIZARD is its ability to generate the full six fermion final state in the matrix element. For the following discussion, the complete process $qq \to qq\ell\nu\ell\nu$ was generated. One result of that is, that the vector bosons are not directly accessible anymore even on generator truth level. To get a handle on the resonance mass, the invariant boson pair mass m^{WW} had to be replaced by the invariant mass of the leptons and neutrinos from the hard process $m^{\ell\nu\ell\nu}$.



(a) Fit to Higgs Boson Lineshape

(b) Fit to K-Matrix σ Lineshape

Figure 4.6: *Distributions of W boson pair invariant mass m^{WW} (with VBS cuts) for SM with a Higgs boson at $m = 850\,\mathrm{GeV}/c^2$ and a K-Matrix model with a scalar isoscalar resonance σ of coupling $g = 1.0$ and $m = 850\,\mathrm{GeV}/c^2$. Both are generated with* WHIZARD. *The lineshapes differ due to additional unitarization in the K-Matrix sample. Left: SM is fitted with the Breit-Wigner model in Equation* (4.1). *Right: Fit to K-Matrix using the s-dependent model in Equation* (4.2).

In Figure 4.7a, the $m^{\ell\nu\ell\nu}$ distribution for samples with a scalar isoscalar resonance σ at $m = 850\,\mathrm{GeV}/c^2$ is shown for three values of the coupling strength. σ is the widest of the resonances, so wide in fact, that the peak is shifted to lower energies. That is an effect of the s-dependent width in conjunction with the steep continuum background. For larger masses, this is even more pronounced. Examples at $m = 1150\,\mathrm{GeV}/c^2$ are shown in Figure 4.7b. On the opposite end of the width spectrum, the very narrow tensor isoscalar resonance f displays a negligible shift (Figure 4.8 at $m = 850\,\mathrm{GeV}/c^2$). Table 4.1 gives a list of the widths of the example resonances for reference.

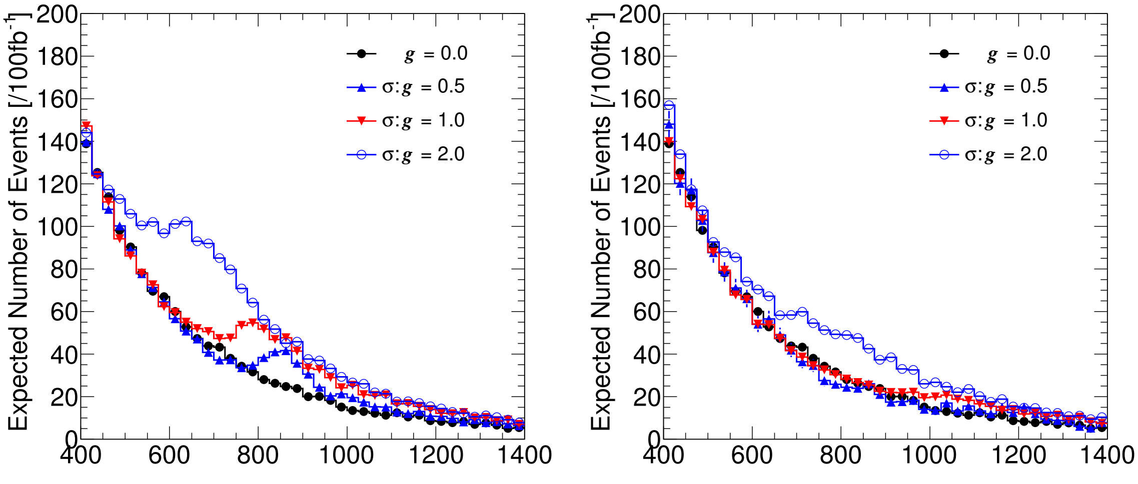
4.2.5 Null Hypothesis

The K-Matrix resonance samples of WHIZARD are generated according to a model extending the Standard Model. To investigate the sensitivity of an experiment it is necessary to compare hypotheses. The signal hypothesis is given by a K-Matrix resonance model. For the background hypothesis there is some ambiguity here, as vector boson scattering in the Standard Model without a Higgs boson is not well defined for the energies in question.

One could choose to use the Standard Model with a light Higgs boson present as

Resonance	Mass m [GeV/c^2]	Coupling g	Width Γ [GeV/c^2]
σ	850	0.5	75
σ	850	1.0	300
σ	850	2.0	1200
σ	1150	0.5	190
σ	1150	1.0	750
σ	1150	2.0	3000
f	850	0.5	2.50
f	850	1.0	10
f	850	2.0	40

Table 4.1: *Example widths for very wide and very narrow resonances. Values calculated using Table 2.1.*



(a) $m = 850\,\mathrm{GeV}/c^2$ **(b)** $m = 1150\,\mathrm{GeV}/c^2$

Figure 4.7: *Distributions of $\ell\nu\ell\nu$ invariant mass $m^{\ell\nu\ell\nu}$ of K-Matrix models with a scalar isoscalar resonance σ at different couplings g. This resonance is the widest, its maximum quickly moving up the slope of the continuum to lower masses with increasing coupling. Left: Mass of σ $m = 850\,\mathrm{GeV}/c^2$. Right: Mass of σ for $m = 1150\,\mathrm{GeV}/c^2$.*

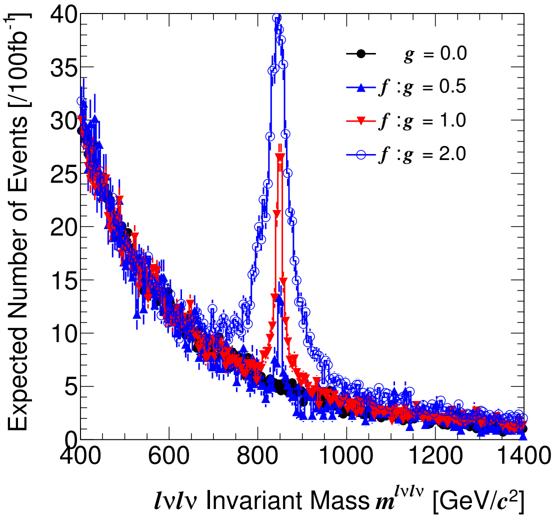
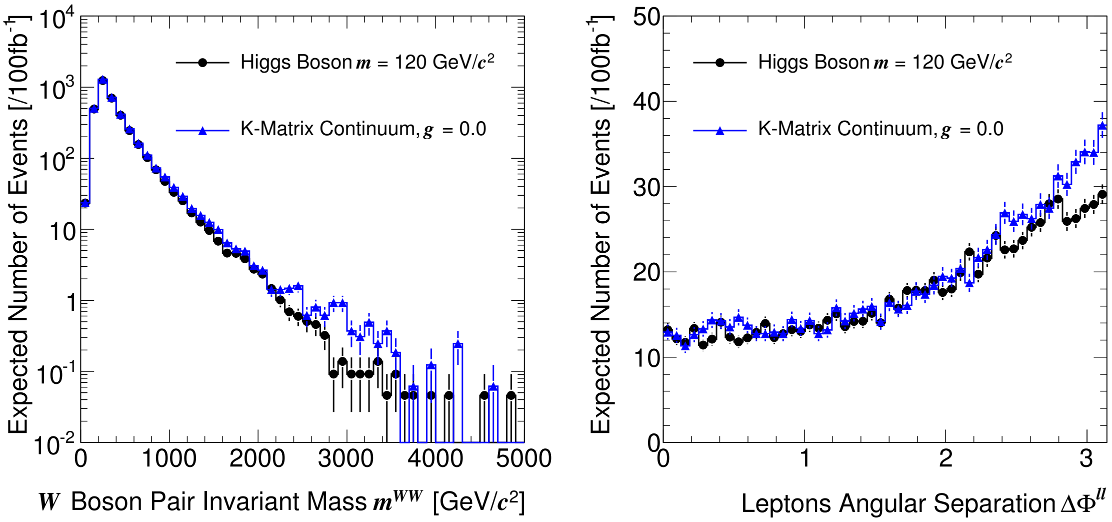


Figure 4.8: *Distributions of $\ell\nu\ell\nu$ invariant mass $m^{\ell\nu\ell\nu}$ of K-Matrix models with a tensor isoscalar resonance f of mass $m = 850\,\mathrm{GeV}/c^2$ at different couplings g. This resonance is comparatively narrow resulting in a well-defined peak. Binning has been modified accordingly.*

null hypothesis. This has the advantage to be relatively well-defined. At the same time a Higgs boson with a mass below the vector boson pair production threshold would not interfere too much with the channel. The exact choice of mass is uncritical.

In this thesis, another alternative is preferred. The null hypothesis is defined as the K-Matrix unitarized continuum without an additional resonance. While strictly speaking this choice compares two different extended models with one another rather than the current Standard Model to an extension, it also reduces the dependence on the unitarization scheme and focuses on the resonances. K-Matrix unitarization is conservative insofar as it tries to minimally change the amplitudes. Still, it does have a slightly increased cross section at high energies compared to the Standard Model with a light Higgs boson. More importantly, the $\Delta\varphi^{\ell\ell}$ distribution is a bit more signal like. Both effects are shown in Figure 4.9. This suggests that in addition to the effects from a potential resonance, the unitarized continuum is shifted closer to what a resonance with small coupling g would look like. By using K-Matrix unitarization in both signal and background hypotheses, a part of the effect of unitarization cancels. This is done in this analysis to focus on the presence of additional resonances. Using a SM with light Higgs boson as null hypothesis would yield slightly better sensitivities for the K-Matrix resonance model.



(a) m^{WW} **(b)** $\Delta\varphi^{\ell\ell}$

Figure 4.9: *Distributions of W boson pair invariant mass m^{WW} and lepton azimuthal angle separation $\Delta\varphi^{\ell\ell}$ for* WHIZARD *null hypotheses. Option 1: SM with a light Higgs boson $m = 120\,\mathrm{GeV}/c^2$. Option 2: K-Matrix continuum without additional resonances. Left: There is only a slight disagreement at highest values of m^{WW} which may be be due to statistical uncertainties. Right: The $\Delta\varphi^{\ell\ell}$ distribution of the K-Matrix is a bit more signal like than the SM Higgs boson.*

5 Monte Carlo Simulation

5.1 Signal and Irreducible Background Samples

As a matrix element generator, Whizard by default simulates all tree level subprocesses for any combination of initial and final state particles. Though technically possible to restrict the allowed Feynman graphs to a subset, the result is not necessarily well-defined. In the process $qq \to qq\ell\nu\ell\nu$ for example one might want to restrict the model to the vector boson scattering graphs. Doing this, however, will give a model that violates unitarity [46] even more strongly than before. Therefore, to see effects related to unitarity one needs to generate all irreducible background with the signal. Processes containing QCD vertices are an exception, as they can be effectively switched off by setting the strong coupling constant to zero (cf. Section 5.1.)

5.1.1 Whizard K-Matrix Resonant $qq \to qq\ell\nu\ell\nu$

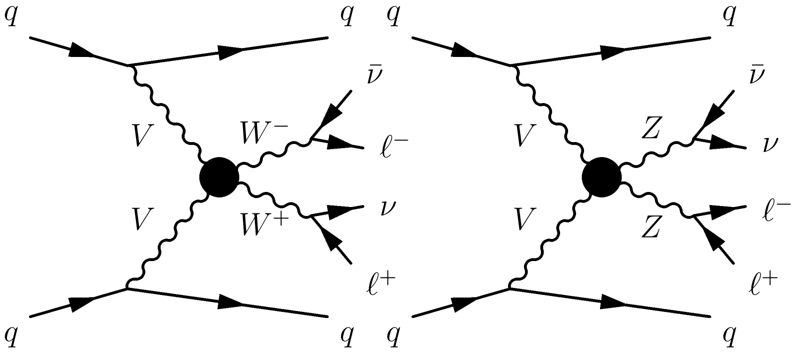


Figure 5.1: *Feynman diagram of a vector boson scattering graph in the process $qq \to qq\ell\nu\ell\nu$. The $VV \to VV$ vertex is changed by the presence of additional K-Matrix resonances. The lepton/neutrino configuration is drawn for scattering into a $WW/ZZ \to \ell^+\ell^-\nu\nu$, but same-signed W bosons are also included.*

The signals searched for in this analysis are any of the additional resonances in the vector boson scattering channel introduced in 2.3.

$$qq \to qq\ell\nu\ell\nu \tag{5.1}$$

A Feynman graph of the signal-like process at tree level is shown in Figure 5.1. Many other background Feynman graphs are also included but not shown. For each of the resonances σ, φ, ρ, f and t, two mass points $m = 850\,\mathrm{GeV}/c^2$ and $m = 1150\,\mathrm{GeV}/c^2$

Resonance	Mass m [GeV/c^2]	Coupling g	Cross Section [fb] (at $\sqrt{s} = 14\,\text{TeV}$)	Number of Events
σ	850.0	2.0	78.0	6000
f	850.0	3.0	88.3	6000
φ	850.0	2.0	78.4	6000
ρ	850.0	4.0	74.0	6000
t	850.0	4.0	89.7	6000
σ	1150.0	3.0	79.4	6000
f	1150.0	4.0	82.3	6000
φ	1150.0	3.0	76.2	6000
ρ	1150.0	4.0	66.3	6000
t	1150.0	4.0	74.8	6000

Table 5.1: *List of* WHIZARD *signal samples containing K-Matrix resonances. Cross sections at pp-center-of-mass-energy* 14 TeV *are given at generator level before detector simulation and acceptance cuts. Only events with two leptons of* $|\eta| < 2.7$ *and* $p_\text{T} > 15\,\text{GeV}/c$ *were accepted. Couplings* g *of samples are chosen to result in a comparable cross section.*

were chosen close to the masses investigated in Reference [33]. A large coupling g was chosen for each of the ten resulting resonance type and mass combinations, such that the resulting cross sections of all signal samples are similar. Later reweighting to lower couplings (cf. Section 5.3) should cover the expected range of sensitivities. The effect of the Higgs boson was removed by setting its mass to infinity. All samples were generated for a center of mass energy of 14 TeV. Table 5.1 contains a list of these samples.

WHIZARD was used to generate the matrix elements at tree level for each of the resonances. Only one resonance was switched on at a time. The strong coupling constant α_s was set to zero to suppress graphs containing QCD vertices. Quark flavors included in the samples were limited to u, d and s due to computational constraints. Since the phase space integration is done individually, the cross section including all four flavors could be reconstructed from individual log files. For all WHIZARD samples, the given cross section corresponds to quark flavors $q \in \{u, d, s, c\}$. The heavier quarks were not simulated due to their small contribution. A veto on b-jets is performed during event selection which further reduces the expected effect of the heavy flavors. On the other hand there is a possibility that a veto against b-jets might also affect c-jets. In principle, the selection efficiency for signal might be lowered for the c-quark contribution but the effect is believed to be small and is therefore neglected. Neither the cross section itself nor its uncertainty were available from literature and had to be taken from the generator's own phase space integration. The numbers are therefore given in leading order as the simulation itself. Statistical uncertainties from the integration are on the order of 1‰ allowing them to be neglected. Scale uncertainties are much larger and will be revisited in the section on systematic

	[GeV/c^2]
m^t	172.5
m^W	80.403
Γ^W	2.141
m^Z	91.1876
Γ^Z	2.4952

Table 5.2: *Additional mass and width parameters used in* WHIZARD

Type	Cross Section [fb] (at $\sqrt{s} = 14\,\text{TeV}$)	Number of Events
WHIZARD EW	63.8	6000
WHIZARD EW+QCD	382.7	29000

Table 5.3: *List of* WHIZARD *irreducible background samples with K-Matrix unitarization. Cross sections at pp-center-of-mass-energy* 14 TeV *are given at generator level before detector simulation and acceptance cuts. Only events with two leptons of* $|\eta| < 2.7$ *and* $p_\text{T} > 15\,\text{GeV}/c$ *were accepted.*

uncertainties (7.2.3.)

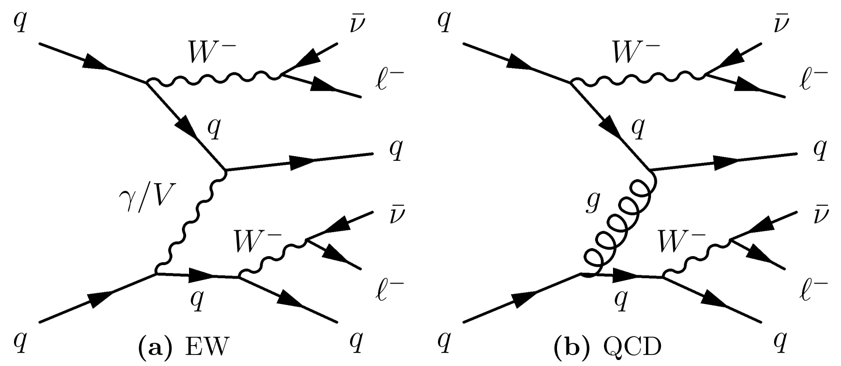
As parton density function, the implementation of CTEQ6L [67] from the LHAPDF [68, 69, 70] library was used. The Z mass was chosen as PDF scale. Leptons and light quarks were treated as massless. Additional mass and width parameters are given in Table 5.2.

The output is in the LHEF format and was read into the ATLAS software ATHENA for parton showering and hadronization by PYTHIA. τ lepton decays and final state radiation were performed by TAUOLA [71] and PHOTOS[72]. All decays of the τ leptons were allowed, although most hadronically decaying τ leptons were filtered out by subsequent cuts. After hadronization a multi-lepton filter was applied to the truth event. Only events that contained two leptons within $|\eta| < 2.7$ with $p_\text{T} > 15\,\text{GeV}/c$ each were processed any further. The detector response was simulated with GEANT4[73, 74]. Finally, standard reconstruction algorithms were run on the fully simulated events.

5.1.2 Whizard Standard Model $qq \to qq\ell\nu\ell\nu$

Two samples were generated for the irreducible background. The electroweak sample was set up exactly like the signal samples above, with the exception that no K-Matrix resonance was enabled. All other parameters, including the whole simulation chain were the same. This sample completely overlaps with each signal sample. For the QCD and electroweak sample, QCD vertices were allowed. A summary is given in Table 5.3.

Typical Feynman graphs for electroweak and QCD processes are shown in Figure 5.2.



(a) EW **(b)** QCD

Figure 5.2: *Feynman diagram of background subgraphs in the process $qq \rightarrow qq\ell\nu\ell\nu$. On the left, electroweak and on the right QCD background. Instead of two W bosons, Z bosons may be emitted as well, if one decays to two neutrinos.*

5.2 Reducible Background Samples

In addition to the irreducible backgrounds that have exactly the same tree level initial and final state as the signal process, there are some reducible backgrounds to take into consideration. By their nature, interference between these and the signal is restricted to higher orders of perturbation theory and individual processes may be assumed to add up. The importance of a particular background is given by its influence on an analysis' sensitivity. Sensitivity in turn is determined by the difference of histogram bins between models. A background is per definition not dependent on this difference. Therefore, the more events of a background end up in a relevant histogram bin N_i^{back} after event selection, the more problematic it is. Large expectation values may be either due to a large initial cross section or a high selection efficiency. The former is intrinsic to the background in question, the latter depends how similar the events look in the detector compared to signal like events.

The signal has a final state of $qq\ell\nu\ell\nu$. Neutrinos are invisible in the detector but their presence can be inferred by large $E_{\text{T}}^{\text{miss}}$. Charged leptons with a large p_{T} are relatively rare in hadronic collisions and can be measured well. The two quarks that are left tend to form two highly energetic jets. For a process to look signal like, it would have to have two charged leptons of large p_{T} or objects that can fake one or two and two energetic jets. Additional jets may be present as the signal may include them due to higher order effects (simulated by the parton shower only) or additional interactions in the event.

Next to leading order cross section calculations are available for the backgrounds. These are used for normalization, despite the fact that for signal only a leading order calculation from the generator is available. This is considered acceptable, as it uses the best known value each time. For handling of uncertainties it is already necessary to make the assumption that for any order in perturbation theory, calculations estimate the same value and the uncertainties are correctly estimated. Going from one order to a higher order, should then decrease uncertainty. This argument is not strictly frequentist, but a necessary offshoot of the perturbation method. It is important, however, not to pick an estimate with a view on results. This can very easily lead to bias. Here, next to leading order estimates tend to be higher than leading order for the background. If there is a bias from not using leading order throughout, it is to pessimistic results.

Tables 5.4 and 5.5 contain a list of background samples and their properties. Details follow in the sections on individual background processes.

Type	Cross Section [pb] (at $\sqrt{s} = 14\,\mathrm{TeV}$)	Total Weight	Equivalent Luminosity [fb^{-1}]
$t\bar{t}$	60.5	1374910.0	22.7
Wt	3.67	59482.0	16.2
$W + 0j(W \to \mu\nu)$	110.0	19999.0	0.182
$W + 1j(W \to \mu\nu)$	84.2	29949.0	0.356
$W + 2j(W \to \mu\nu)$	87.9	98000.0	1.11
$W + 3j(W \to \mu\nu)$	75.0	149748.0	2.0
$W + 4j(W \to \mu\nu)$	42.7	69999.0	1.64
$W + 5j(W \to \mu\nu)$	24.7	29749.0	1.2
$W + 0j(W \to e\nu)$	147.0	19950.0	0.136
$W + 1j(W \to e\nu)$	96.2	29998.0	0.312
$W + 2j(W \to e\nu)$	96.1	99999.0	1.04
$W + 3j(W \to e\nu)$	79.4	149997.0	1.89
$W + 4j(W \to e\nu)$	43.4	69956.0	1.61
$W + 5j(W \to e\nu)$	24.9	29975.0	1.2
$W + 0j(W \to \tau\nu)$	20.3	10000.0	0.493
$W + 1j(W \to \tau\nu)$	19.5	15000.0	0.768
$W + 2j(W \to \tau\nu)$	20.3	50000.0	2.47
$W + 3j(W \to \tau\nu)$	16.2	72700.0	4.49
$W + 4j(W \to \tau\nu)$	9.97	35000.0	3.51
$W + 5j(W \to \tau\nu)$	6.18	15000.0	2.43

Table 5.4: *Table of background samples. Cross sections at pp-center-of-mass-energy* 14 TeV *are given at generator level (after parton shower and fragmentation). Only $t\bar{t}$ is weighted. For all other backgrounds total weight is equivalent to the number of generated events. Cuts on tagging jet candidates and two leading leptons were applied on generator level (cf. Section 5.2.1.)*

5.2.1 Top Pair Production $t\bar{t}$

Leading order Feynman graphs are shown in Figure 5.3. Top pairs are expected to be ubiquitous at the LHC. At the same time, top quarks decay almost exclusively into a bottom quark and a W boson. Bottom quarks form jets in the detector, but may also contain highly boosted charged leptons. They tend to be aligned with

Type	Cross Section [pb] (at $\sqrt{s} = 14\,\mathrm{TeV}$)	Total Weight	Equivalent Luminosity [fb^{-1}]
$Z + 0j(Z \to \mu\mu)$	7.44	39748.0	5.34
$Z + 1j(Z \to \mu\mu)$	6.81	79999.0	11.8
$Z + 2j(Z \to \mu\mu)$	7.59	98899.0	13.0
$Z + 3j(Z \to \mu\mu)$	7.12	99750.0	14.0
$Z + 4j(Z \to \mu\mu)$	3.78	48850.0	12.9
$Z + 5j(Z \to \mu\mu)$	2.11	30000.0	14.2
$Z + 0j(Z \to \tau\tau \to \ell\ell)$	0.316	59383.0	188.0
$Z + 1j(Z \to \tau\tau \to \ell\ell)$	0.287	45500.0	158.0
$Z + 2j(Z \to \tau\tau \to \ell\ell)$	0.344	338748.0	986.0
$Z + 3j(Z \to \tau\tau \to \ell\ell)$	0.299	217011.0	726.0
$Z + 4j(Z \to \tau\tau \to \ell\ell)$	0.182	32499.0	178.0
$Z + 5j(Z \to \tau\tau \to \ell\ell)$	0.106	16000.0	151.0
$Z + 0j(Z \to ee)$	12.0	39000.0	3.25
$Z + 1j(Z \to ee)$	8.06	75000.0	9.31
$Z + 2j(Z \to ee)$	8.75	104996.0	12.0
$Z + 3j(Z \to ee)$	7.18	99998.0	13.9
$Z + 4j(Z \to ee)$	4.13	50000.0	12.1
$Z + 5j(Z \to ee)$	2.15	29954.0	13.9

Table 5.5: *Continuation of Table 5.4: Background samples. Cross sections at pp-center-of-mass-energy* 14 TeV *are given at generator level (after parton shower and fragmentation). Only $t\bar{t}$ is weighted. For all other backgrounds total weight is equivalent to the number of generated events. Cuts on tagging jet candidates and two leading leptons were applied on generator level (cf. Section 5.2.1.)*

the jet axis and have therefore usually other particles in close vicinity. This lack of isolation is well-suited to reject leptons from this source. The two W bosons may decay leptonically. If both do, there are two highly boosted and usually isolated leptons in the event. The corresponding neutrinos give rise to $E_{\mathrm{T}}^{\mathrm{miss}}$. Two quark jets are available to mimic the jets of the signal. In the fully leptonic channel, the final state of top pair production has almost the same particle content as the signal. Most notably, the quarks are always bottom quarks. Due to the long lifetime of B mesons, heavy flavors can be identified in the detector. This is one of the important methods to reduce the $t\bar{t}$ content in the selected events.

Current measurements of the $t\bar{t}$ cross section need to be extrapolated to the center of mass energy $\sqrt{s} = 14\,\mathrm{TeV}$ of the LHC. By the time this analysis will be applied to experimental data, $t\bar{t}$ cross section measurements at that energy will be available. Until then, theoretical predictions have to be used to normalize the samples. The theoretical calculation recommended by ATLAS is published in [75] and includes next to leading order and next to leading log corrections. It predicts the cross section for

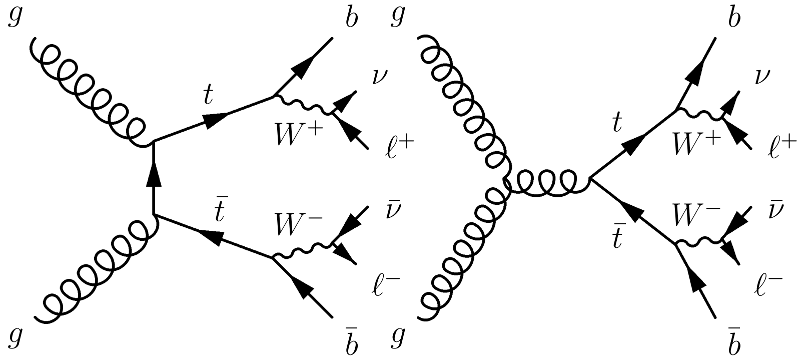


Figure 5.3: *Feynman graphs of $t\bar{t}$ production. At* LHC *energies, gluon fusion dominates.*

all non-fully hadronic channels at $\sqrt{s} = 14\,\text{TeV}$ to be

$$\sigma = 833^{+52}_{-39}\text{pb} \tag{5.2}$$

Often, background samples are centrally produced on the grid and made available to all analyses within ATLAS. This ensures that results are comparable and that resources are shared. For this analysis, centrally produced $t\bar{t}$ samples were not available in sufficient quantities. A sample of 32.5 million $t\bar{t}$ events produced with MC@NLO [76, 77] had been made available from the ATLAS top group but only at matrix element level. Therefore, this background had to be processed specifically for this analysis.

As a next to leading order generator MC@NLO produces events with negative weights. While such an event is not physical, the distributions produced by filling histograms with the given weights have the correct expectation values. Only weights 1.0 and −1.0 occur. The raw input from MC@NLO was read into Athena for parton shower and hadronization by Pythia in the same configuration as for the signal including use of Photos and Tauola. A multi-lepton filter similar to the filters for signal was applied. Only events with two leptons are allowed to pass. In order to increase reusability of the samples for other analyses, the filter was relaxed to accept leptons of $p_{\mathrm{T}} \geq 10\,\text{GeV}/c$. In addition a truth filter on the two jets was applied taking advantage of the rapidity gap in signal events, suppressing a large part of the background. For each event, a list of truth jets was selected and overlap with leptons and photon removed. Any jet was removed for which a lepton or photon existed with the following properties:

- $p_{\mathrm{T}} > 15\,\text{GeV}/c$
- $|\eta| < 2.5$
- $\Delta R^{\text{jet},\gamma/\ell} < 0.05$

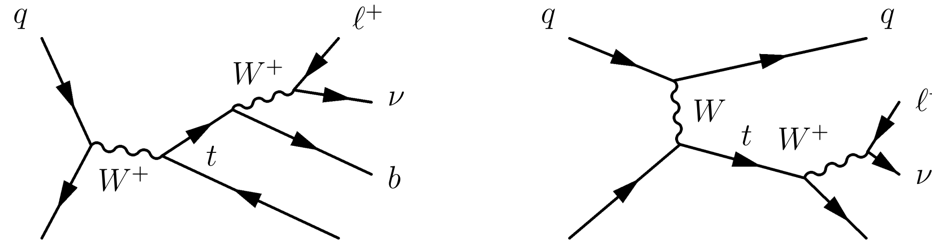
Only events containing a pair of jets j_1 and j_2 out of this list, fulfilling the following conditions were allowed to pass:

- $p_\mathrm{T} > 15\,\mathrm{GeV}/c$ for both
- $|\eta| < 5.0$ for both
- invariant mass of pair $m^{jj} \geq 300\,\mathrm{GeV}/c^2$
- pseudorapidity separation $\Delta\eta > 2.0$

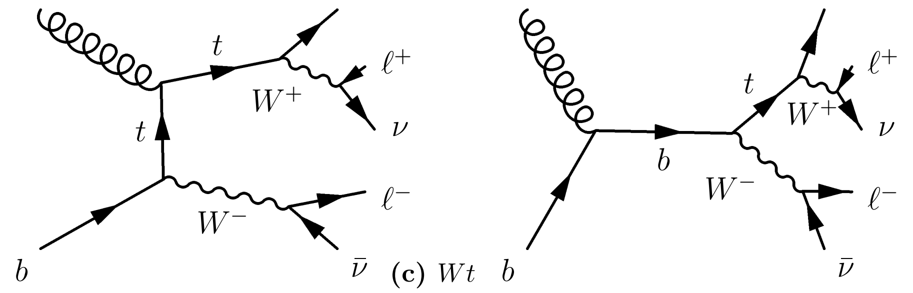
After filtering, the total weight of events left was 1374910.0 which corresponds to approximately $\mathscr{L} = 23\,\mathrm{fb}^{-1}$ integrated luminosity.

Detector simulation was performed using Atlfast-II to decrease computational requirements. Atlfast-II is a detector simulation that uses a reduced complexity description of the detector to combine high accuracy with speed. It is much closer to the full Geant4 simulation than the original fast detector simulation which only smeared truth particles according to detector resolutions.

5.2.2 Single Top Production Wt



(a) s-Channel **(b)** t-Channel



(c) Wt

Figure 5.4: *Feynman graphs of single top production. Only Wt production is a major background for this analysis.*

Single top production comes in three flavors, t-channel, s-channel and Wt production. Feynman graphs of all three processes at tree level are shown in Figure 5.4. Cross sections are taken from [78] according to ATLAS recommendations. All calculations are next to leading order and assume at least one lepton in the final state.

Their values are:

$$\sigma^{\text{t-channel}} = 246.6^{+11.8}_{-12.2}\,\text{pb} \tag{5.3}$$

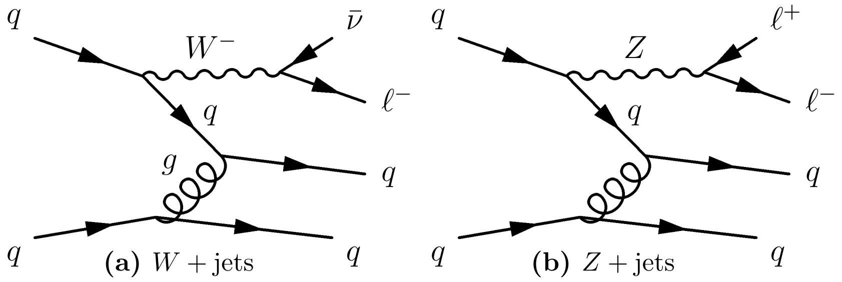
$$\sigma^{\text{s-channel}} = 10.65^{+1.12}_{-1.02}\,\text{pb} \tag{5.4}$$

$$\sigma^{Wt} = 66.2^{+2}_{-2}\,\text{pb} \tag{5.5}$$

While the t-channel has the largest cross section, it only contains one W which could decay leptonically. At the same time it has at least one bottom quark in the final state which may be vetoed. It is not expected to have a sufficiently high selection efficiency to become noticeable in the analysis. The s-channel both lacks a second isolated lepton and has two bottom quarks. In addition its cross section is more than an order of magnitude lower and it can also be neglected. Only Wt is close enough to the signal to warrant further analysis. It contains a W boson from the top decay and a second one from its production. When both decay leptonically, the right number of charged leptons and neutrinos is in the final state. However one jet is missing and the other one is from a bottom quark. From now on, out of the single top processes, only Wt will be considered.

As event generator for Wt, AcerMC [79] was used. It is leading order for this process and produces unweighted events. 1.2 million events were produced. As parton density function CTEQ6L was used as for all other samples. Parton shower and hadronization was done in Pythia as for all other samples. The same generator level cuts were applied as for $t\bar{t}$, i.e. the multi-lepton filter and the filter on jets. After cuts, 59482 events were left corresponding to about $16\,\text{fb}^{-1}$.

5.2.3 $W+$jets and $Z+$jets Production



(a) $W+\text{jets}$ **(b)** $Z+\text{jets}$

Figure 5.5: *Feynman graphs of* $W+$jets *and* $Z+$jets *production. Despite the differences in final state, the large cross sections warranted making sure this background was suppressed.*

$W+$jets and $Z+$jets production are two other potential backgrounds. Feynman graphs are shown in Figure 5.5. Both have large cross sections and a weak boson which may decal leptonically. $W+$jets is disadvantaged because it has only one real lepton. Despite its large cross section its jets do not tend to have enough p_T to make it relevant at the end. $Z+$jets may have two isolated leptons in the final state, which makes it potentially much more dangerous. However, the lepton pair's invariant mass is determined by the Z resonance, which is far lighter than the resonances in the signal.

Unlike the other samples which had to be simulated or even completely generated from scratch for this analysis, $W+$jets and $W+$jets samples were available from central production. These include only QCD but not electroweak production which has a much lower cross section. The generator used is ALPGEN [80] which is capable of using MLM matching (cf. [81] for an overview) to reconcile additional hard jets from the matrix element with the parton shower. The generator does not simulate additional internal lines and hence produces unweighted events.

Only leptonic decays of the bosons have been considered, in the case of Z bosons only decays into charged lepton. The samples are split up into different final states including lepton flavor and jet multiplicities. Up to five additional jets from the matrix element had been considered. As parton density function CTEQ6L was used as for all other samples. On parton level, a cut on the invariant mass $m^{\ell\ell} < 200\,\mathrm{GeV}/c^2$ of the lepton pair from the Z boson decay was applied. While this is unfortunate, masking potential lepton pairs of large invariant mass, it is unlikely that these could be produced in large quantities from the Z resonance. Any effect from fake leptons or measurement resolution is not affected by this cut. After generation, parton shower and hadronization had been performed via ATHENA using the HERWIG [82, 83] code. PYTHIA could not have been used due to an incompatibility between the MLM matching scheme and the new parton shower in PYTHIA 6.4. TAUOLA and PHOTOS had been used for tau decays and weak boson final state radiation. After hadronization, the same jet cut was applied and two leptons of $p_\mathrm{T} \geq 10\,\mathrm{GeV}/c$ were preconditioned as for the top quark backgrounds.

Leading order cross section estimates are available from the generator. As these backgrounds are removed early in the analysis chain, corrections from next to leading order calculations are too small to enter in the result and can be neglected.

Note that diboson production is covered by the WHIZARD irreducible background albeit not including additional hard jets from the matrix element.

5.2.4 Pileup

For all samples, a realistic detector simulation utilizing either GEANT4 or ATLFAST-II was performed for the process in question. Multiple interactions in the same parton called the underlying event are simulated by the parton shower generators and considered to be part of the same process. In addition to these interactions there are other interactions that may register in the detector. The vacuum in the beam pipe is excellent but not perfect. Interactions with the beam gas may produce particles in the detector in addition to those of an interesting event. Similarly, the activation of the detector and cavern by the intense radiation of the LHC leads to further background.

The most important effect, however, is caused by pileup. In-time pileup are proton-proton collisions during a bunch crossing in addition to an interesting event. The high instantaneous luminosity of the LHC is expected to lead to ultimately 23 additional events in an average bunch crossing. For SuperLHC the number may well be ten times larger. Out of time pileup stems from collisions that have happened in the bunch crossings before or after a particular event. Due to the integration time of the calorimeter which is larger than the bunch crossing interval entries may add up.

Additional particles in the detector may change performance. Most of the effect is to be expected in the very forward and backward directions of the detector and low p_{T} objects. As opposed to the heavily filtered interesting events, pileup events are made up of average LHC events. In a hadronic machine this is dominated by QCD interactions with relatively little p_{T} and very few isolated leptons in the tracker. Such events are called minimum bias events, alluding to the fact that such events are read out when only minimal filtering by triggers and cuts is performed.

All these effects can be simulated in dedicated samples. However, at the time this analysis is performed, no pileup samples had been available for the needed instantaneous luminosities. Even for lower luminosities, samples had only been done for signal, not for the backgrounds. The effect of pileup was consequently ignored in the samples used here. As of the time of writing a follow-up study is under way to investigate the effect of low luminosity pileup. The loss of sensitivity that has been seen so far is moderate [84].

5.3 Event Reweighting

5.3.1 Method

The phenomenological model under discussion has quite a few important parameters. For each resonance there is a mass m and coupling g making a total of 10 new parameters. One of these resonance would be the first to become visible in ATLAS, so this thesis looks only at one resonance at a time greatly reducing the parameter space points to be looked at. By focusing on two example mass points, a manageable number of parameter points results. Left are the couplings g for which a small spacing is desirable in order to determine discovery and limit sensitivities in terms of g. One can easily arrive at some 10-20 points in g to to get a reasonably fine spacing. If one simply generates each parameter point anew, the computational requirements scale linearly. That is unfortunate, as Monte Carlo simulations are computationally quite expensive. Typically, however, the full detector simulation is much slower than the physics event generation step. Therefore, one may save a lot of simulation time by using a reweighting technique to reuse the fully simulated events.

In first approximation an event generator produces events that are drawn uniformly from possible final states in phase space. The probability density is not equal everywhere of course, but depends on the matrix element and the phase space factor. In order to still get correct distributions, each event is assigned a weight proportional to the local probability density. One may simply fill in histograms from this weighted distribution, making sure that each entry is weighted using the given weight and possibly an overall normalization factor. For detector level studies this is quite efficient, as every event that is generated is used and adds some bit of information.

Including detector simulation, the trade-off is different. A more expensive step follows generation that one would not want to run disproportionally often for events of low weight. So instead the events are unweighted by rejecting some of them: An event is retained with a probability proportional to its weight. Clearly, if an event with weight w is only kept with the probability w, it needs to be assigned a new weight to reflect its lower abundance. The change in abundance for each event is $\frac{1}{w}$, thus all events have now weight one.

Having produced some fully simulated events for a particular model, one can apply the (un-)weighting scheme above to a new problem. A change in any parameter affecting the matrix element of a process, say a change from g to g', affects the probability density for the final state configurations. However, for the same set of final state particles, the states live in the same phase space with the same variables. With the exception of areas with a vanishing probability density, two sufficiently large samples at g and g' will contain the same kind of events, just with different abundance. One can thus get from one to the other sample by changing the abundance. Alternatively, as has been illustrated above, one can adjust the weights retaining all events.

An important condition is, that for no part of the phase space the probability density may vanish in the source process (coupling g) where it is present in the destination (coupling g'). For finite sample sizes, this expands to the condition that the probability densities must not increase too much from g to g'. Otherwise there are just not enough events in the source to describe the destination well. For reweighting couplings, one tends to get better results by weighting from larger to smaller couplings [41]. Masses are for the same reasons difficult to change, as shifting mass peaks will lead to large changes in phase space density.

Technically, the needed weights are already calculated within the generator before unweighting. It is simply the ratio between the event weight at g' to the weight at g, i.e. $\frac{w'_i}{w_i}$. WHIZARD can reprocess events it has produced and output events with the correct reweighting applied. These weights were mapped to the correct events in the output of the ATLAS simulation. Unfortunately, the standard event format does not contain a unique event identifier for events that spans all generation steps. The mapping between the events at generator level and their weights to the events at detector simulation level had to be developed specifically for this analysis.

Signal samples were generated at a high coupling g, specific to each resonance signal. Reweighting was used to reach all other g' down to $g = 0$. All original samples are unweighted, therefore $w_i \equiv 1, \forall i$, simplifying the weight expressions.

5.3.2 Validation

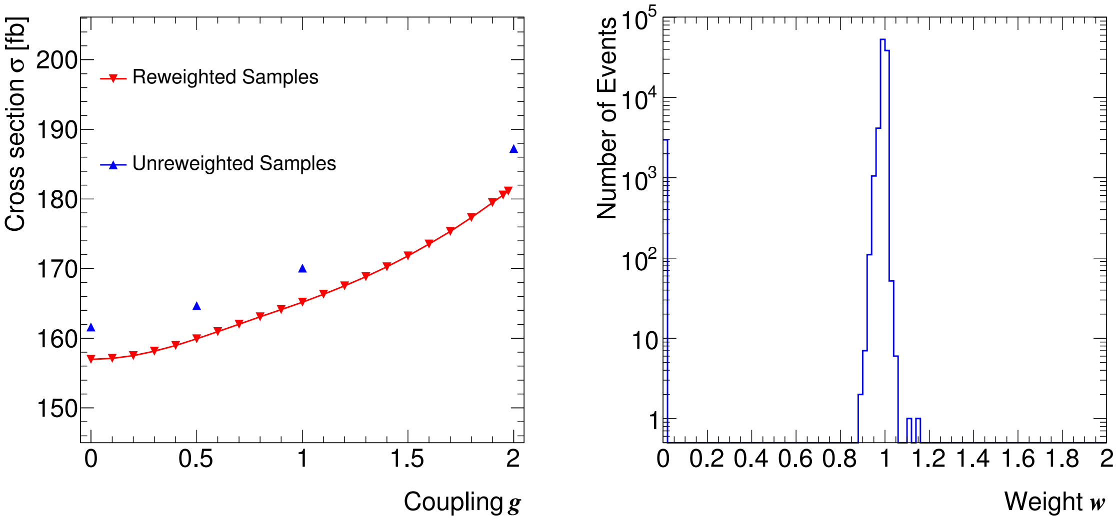
The method introduced above was applied to vector boson scatting events generated by WHIZARD in the model including the scalar isoscalar σ at mass $m = 850\,\mathrm{GeV}/c^2$

and coupling $g = 2.0$. Multiple reweighting were performed down to $g = 0.0$. Additionally, the points $g = 0.0$, $g = 0.5$ and $g = 1.0$ were generated directly for comparison.

Different couplings give rise to different sample cross sections. This change has to be reflected in the reweighted samples as well as directly generated ones. For the latter, the estimated sample cross section at tree level is calculated by the generator in the phase space integration step. For reweighted samples, the integration is not redone explicitly. However, the new weights are themselves samples of the phase space density ratios between the old and the new phase space. Their average is thus an estimator of the ratio between the old and the new cross sections $\sigma(g)$ and $\sigma(g')$ respectively. The new estimate for the cross section can be written in terms of the old cross section and the weights, using $w_i \equiv 1$.

$$\sigma(g') = \frac{\sum_{i=1}^{N} w'_i}{N}\sigma(g) \tag{5.6}$$

In Figure 5.6a both cross sections calculated for reweighted samples according to the given equation are shown as well as the cross sections from direct phase space integration for unreweighted samples. Error bars for statistical fluctuations are negligible and are hidden by the markers.



(a) Cross Section Estimates **(b)** Weight Distribution

Figure 5.6: *Left: Cross section estimates $\sigma(g)$ versus coupling g for scalar isoscalar resonance σ, mass $m = 850\,\mathrm{GeV}/c^2$, no additional generator level cuts. No correction for incorrectly vanishing weights has been applied. Right: Weight distribution for the reweighted sample at $g' = 1.975$. The peak at zero is an unphysical artifact.*

Clearly, the two sets of cross sections of reweighted and unreweighted samples do not match. The reweighted samples seem to be shifted by a constant amount to lower cross sections. A first hint for the cause of this effect comes from the distribution of weights for the sample only slightly reweighted from $g = 2.0$ to $g' = 1.975$ plotted in Figure 5.6b. In addition to the distribution of weights close to 1.0, there is a conspicuous peak at exactly at 0.0 that is not explicable by numerical rounding alone. The ratio of events with vanishing weight to other events corresponds to the ratio of the cross section shift to the calculated cross section, marking a faulty reweighting as the culprit. This a bug in WHIZARD that has been first discovered in this context. At the time of writing it has not been fixed in the used branch. However, it seems not to be correlated with any physical property of the events in question. This allows the simple work-around to ignore all events with vanishing weights from the average in Equation (5.6). The main disadvantages are some additional complexity in the code and a slightly increased statistical fluctuation. As can be inferred from the ratio of correct to incorrect cross sections, the fraction of broken events is $\approx 3\%$, which is quite manageable.

A corrected plot is shown in Figure 5.7. The sample cross sections agree between reweighted and unreweighted samples. Any residual pulls are small compared to the cross section resolution achieved. As a further test, the truth level distribution of the lepton azimuthal angle separation $\Delta\varphi_{\ell\ell}$ was compared at $g = 1.0$ using an unreweighted sample generated at that coupling, and one that was reweighted down from $g = 2.0$. A fiducial cut on lepton $|\eta| < 2.5$ was imposed and the histograms normalized to unity. In Figure 5.8 the results are plotted, showing agreement.

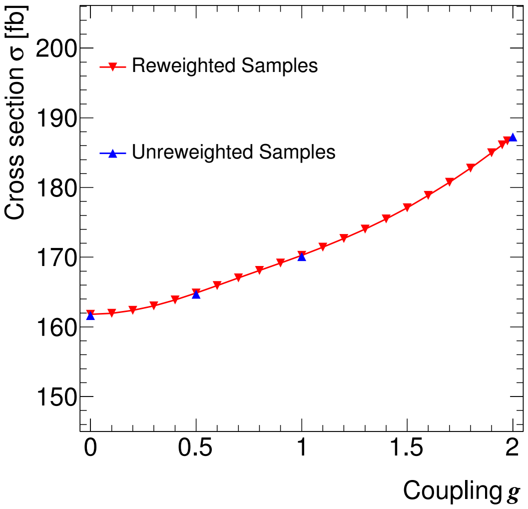


Figure 5.7: *Cross section estimates $\sigma(g)$ versus coupling g for scalar isoscalar resonance σ, mass $m = 850\,\mathrm{GeV}/c^2$, no additional generator level cuts. Events with vanishing weights have not been considered in the determination of the cross section.*

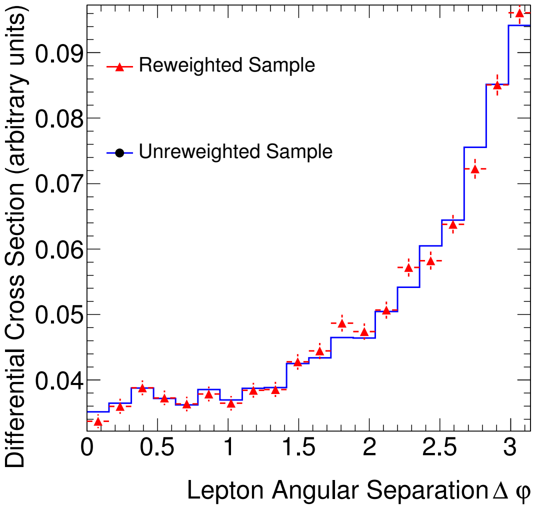


Figure 5.8: *Truth level distribution of the lepton azimuthal angle separation* $\Delta\varphi_{\ell\ell}$ *at* $g = 1.0$ *for a scalar isoscalar resonance* σ*. Reweighted sample originally at* $g = 2.0$*.*

5.4 Signal Definition

5.4.1 Conventional $s + b$

To be most general, in order to simulate a model of particle physics one would have to simulate collision events for the whole model together. There would have to be a generator of Standard Model physics producing Standard Model physics events throughout all of the phase space. Similarly for any alternative model. There would be no way around this, if there was large interference between processes. In order to compare the two model one would have to generate lots of events and then compare at the very end some variable that is different between the two cases. This is if at all possible quite unwieldy. In most of the phase space of the models that are to be compared, there is no change. An analysis would try to focus on events where a change is expected. Practically that means cutting out large parts of phase space. It would therefore be useful to not have to generate all the useless events in the first place. One feature of perturbative models is that they tend to decompose into approximatively additive processes. An equivalent formulation would be that there are processes that do not interfere with one another. This is only approximately true in most cases, but the approximation tends to be quite good for many practical examples.

Using this, it is possible to compose a sample of a complete model by adding up subsamples for non-interfering processes. For a particular analysis, most of the possible processes will not have to be included at all, as far as it is clear that they will be removed by the event selection. Of the remaining processes some can be shared between models. A search for a new particle will often require some processes present in the Standard Model and the new model called backgrounds. The new model will often contain only one additional process called signal for the purposes of that particular analysis.

In the present analysis this technique can be applied directly for the reducible backgrounds. Processes with the relevant final state $qq\ell\nu\ell\nu$ have to be treated specially, as they are not completely separable (cf. Sections 5.1 and 4.2.5.) To arrive at a more conventional signal definition, where at least the expectation values in histograms of signal and background add up, some additional steps are necessary. A procedure has to be defined and tested how to subtract the background components from the signal samples.

5.4.2 Recovering $s + b$ at Histogram Level

The reducible background samples together with the QCD sample constitute an estimate of the Standard Model. In terms of a histogram with M bins $i \in [1, \ldots, M]$ and bin entries N_i this can be rephrased mathematically.

$$\begin{aligned} N_i^{\mathrm{SM}} &= N_i^{\mathrm{EW+QCD}} + N_i^{t\bar{t}} + N_i^{Wt} + N_i^{Z+\mathrm{jets}} + \ldots \\ &= N_i^{\mathrm{EW+QCD}} + N_i^{\mathrm{backgrounds}} \end{aligned} \tag{5.7}$$

The electroweak background is contained completely in the EW+QCD sample and therefore is not counted twice. EW+QCD overlaps partly with each signal sample. This overlap corresponds to the electroweak sample. Before a sample of a signal model can be assembled complete with all backgrounds, this overlap has to be removed.

It is not possible to do this event by event, but instead it has to be done at the histogram level. Always subtracting the histograms of EW from signal histograms suggests itself as a straightforward method. In terms of bin entries N_i^{KM} a model with an additional K-Matrix resonance (bin entries N_i^{res}) then becomes:

$$N_i^{\mathrm{KM}} = N_i^{\mathrm{res}} - N_i^{\mathrm{EW}} + N_i^{\mathrm{EW+QCD}} + N_i^{\mathrm{backgrounds}} \tag{5.8}$$

Comparing to the Standard Model equation above, one can now separate off the signal part N_i^{signal}.

$$N_i^{\mathrm{signal}}(g) = N_i^{\mathrm{res}}(g) - N_i^{\mathrm{EW}} \tag{5.9}$$

To illustrate this fact, the total cross section at generator level is shown in Figure 5.9 for various resonances versus coupling g. Per definition, any resonant sample reweighted to a coupling of $g = 0.0$ is equivalent to EW and should match up accordingly in the leftmost point as has been seen already in the similar Figure 5.7. This is the case, supporting this method. The residual difference are on the percent level and hence much greater than the statistical uncertainty. They are assumed to be due to global normalization uncertainties.

While Equation (5.9) is straightforward and defines an acceptable estimator, it can be improved upon. In Figure 5.9 the statistical uncertainty on the total cross section is very low, since there were 100,000 events at truth level available for making the plot.

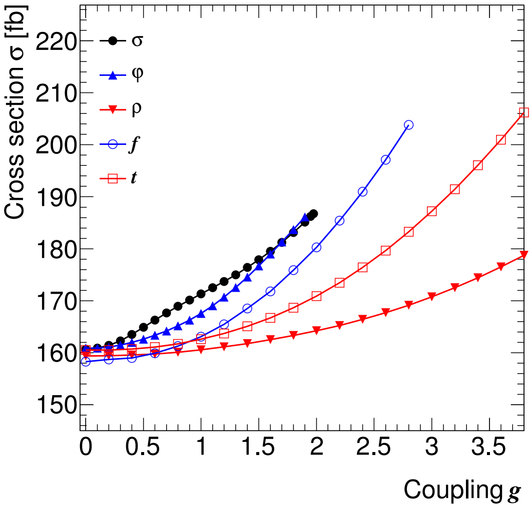


Figure 5.9: *Cross section prediction $\sigma(g)$ versus coupling g for various resonances of mass $m = 850\,\mathrm{GeV}/c^2$. For $g = 0$ the graphs converge to the same point, as this is equivalent to the EW cross section without resonances.*

In full simulation, however, only 6000 events are available. After analysis cuts the number of events left is on the order of 60 to 150. Statistical uncertainties then start to play a role. In anticipation of Section 6, a typical analysis configuration was run on fully simulated events. The resulting total event yield y after cuts constitutes in terms of the just introduced nomenclature a histogram of one bin with $y \equiv N_0$. This bin value y is shown in Figure 5.10 for different resonance samples each reweighted to various couplings g and EW. Error bars correspond to statistical errors. Within statistical uncertainties the EW and K-Matrix resonance samples still match up at coupling $g = 0.0$. In the development of this method it was noted, however, that the reweightings of each sample all contained the same events. The points of each individual graph in Figure 5.10 are highly correlated and have a lower relative point-to-point uncertainty than the statistical uncertainty reflected in the error bars.

This correlation can be used to reduce the uncertainty on the the signal prediction $N_i^{\text{signal}}(g)$. First, consider the definition of (5.9). EW is a completely independent sample, such that its statistical uncertainty fully impacts on the uncertainty of the difference in Equation (5.9). Let $\text{var}(N)$ denote the variance of N.

$$\text{var}(N_i^{\text{signal}}(g)) = \text{var}(N_i^{\text{res}}(g)) + \text{var}(N_i^{\text{EW}}) \tag{5.10}$$

As argued above, a resonant sample reweighted to a coupling $g = 0.0$ is itself an estimate of the EW background. In particular, $N_i^{\text{res}}(g = 0.0)$ is an estimator of N_i^{EW}. Substituting this in Equation (5.9) gives the following alternative signal definition

$$N_i^{\prime\text{signal}}(g) = N_i^{\text{res}}(g) - N_i^{\text{res}}(0.0) \tag{5.11}$$

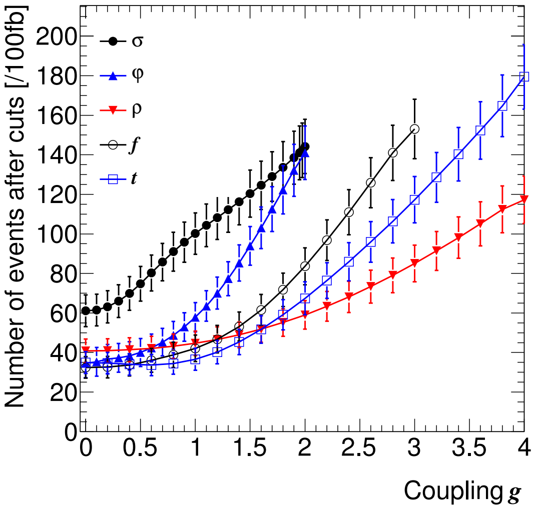


Figure 5.10: *Number of events expected after cuts versus coupling g for multiple resonance samples of mass $m = 850\,\mathrm{GeV}/c^2$. For $g = 0$ the graphs converge to the same point, but statistical fluctuations are large.*

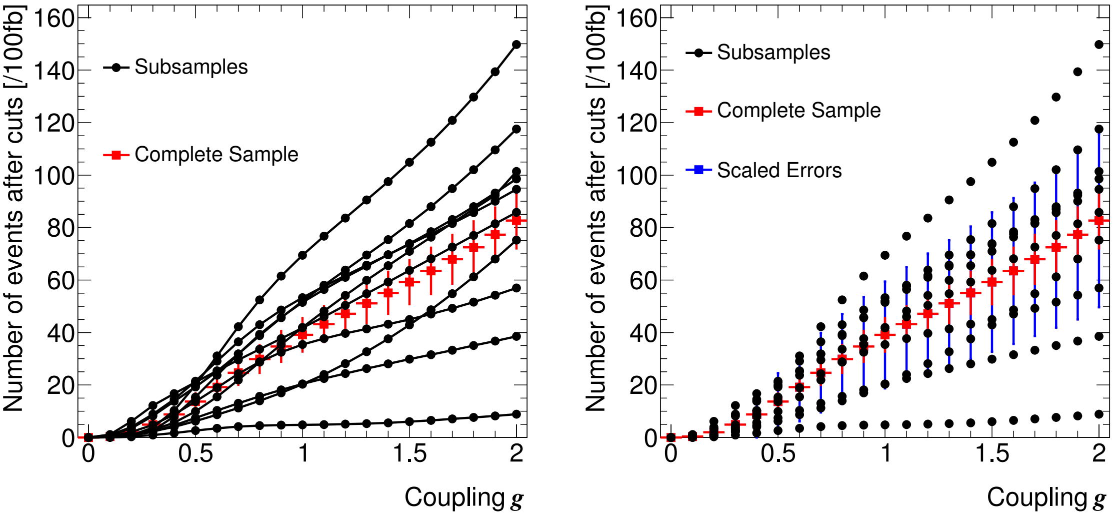
which has the same expectation value as the old one. The variance is much reduced, however, as fluctuations in $N_i^{\mathrm{res}}(g)$ are highly correlated with fluctuations in $N_i^{\mathrm{res}}(0.0)$.

$$\mathrm{var}(N_i^{\prime\mathrm{signal}}(g)) = \mathrm{var}(N_i^{\mathrm{res}}(g) - N_i^{\mathrm{res}}(0.0)) \tag{5.12}$$

To test Equation (5.12), a subsample test was performed. The signal sample containing a scalar isoscalar resonance σ was split into $l = 10$ subsamples of equal size. For all subsamples and the complete sample, the expected signal yield after analysis cuts, $s^{\mathrm{MC}}(g) \equiv N_i^{\prime\mathrm{signal}}(g)$, was calculated according to Equation (5.11). The results are shown in Figure 5.11a. There is a strong correlation between points of the same subsample which has been highlighted by connecting the points of one samples. Any apparent structure can be explained by this correlation. The red points correspond to the whole sample. Error bars in red reflect uncertainties calculated according to Equation (5.12). If the uncertainty is described correctly, it should be related to the variance of the subsample yields $s_j^{\mathrm{MC}}(g)$, where $j \in \{1, \dots, 10\}$ via the relation

$$\mathrm{var}(s^{\mathrm{MC}}(g)) = \frac{1}{l-1}\mathrm{var}(s_j^{\mathrm{MC}}(g)) \tag{5.13}$$

The offset of one to $l - 1$ removes the bias due to the fact that the mean of the subsamples is estimated as well. In Figure 5.11b the same results are shown with enlarged blue error bars ($\times\sqrt{l-1} = \sqrt{9}$) to describe the relative uncertainty on a subsample rather than the uncertainty on the full sample. A visual check was performed on the result, noting that about a third of the subsample results lies outside the error bars as expected.



(a) **(b)**

Figure 5.11: *Number of events in signal expected after cuts versus coupling g for a scalar isoscalar resonance σ. Mass $m = 850\,\mathrm{GeV}/c^2$. Result at $g = 0$ has been subtracted. Subsamples in black are shown with connecting lines on the left to point out correlations. The complete sample is shown in red. On the right, error bars scaled to the expected statistical uncertainty of the subsamples are added.*

With those results a particular signal is defined according to Equation (5.11) with uncertainty according to Equation (5.12). While this does not correspond anymore directly to a signal sample, it fits the usual approach to split histogram entries into background and signal components b and s. In particular, it allows a reasonable definition of s^{MC} as a function of g. Since even with reweighting, the number of points that can be simulated is finite, points in-between are linearly interpolated. The uncertainty $\sigma^{\mathrm{s}}(g)$ if defined as above for each simulated point and similarly interpolated. Note that it does not describe the point to point fluctuations for all pairs of couplings $g \neq g'$ because of correlations. It only does, if one of the two is zero as then s^{MC} vanishes.

6 Event Selection

6.1 Trigger

Readout will be an unprecedented challenge in LHC experiments. While collisions will be produced in the detectors at a rate of 40 MHz [49] this induces an amount of data that is impossible to write to tape for off-line analysis. Instead, the rate of events that can be expected to be available for off-line analysis in ATLAS is between 75-100 Hz. This requires almost three orders of magnitude rate reduction by the trigger (cf. Section 3.2.6).

In an analysis, it is important to take into account the effect of the triggers, just as one needs to take into account any other detector effect or precut. For the Monte Carlo simulation that would be done ideally on an event by event basis. This way, events which would not have passed the relevant triggers can be rejected to get a correct estimate of the yields. Not all samples used here had been available in full detector simulation and thus some did not contain the trigger simulation.

The problem is mitigated by the particular event topology. Vector boson scattering at high energy produces very energetic particles. In the two-lepton channel that results in two very highly boosted leptons, trigger efficiencies tend to form a plateau for particles of high p_T. As long as the cuts used later on in the analysis guarantee, that all passing events are in the plateau region, the trigger efficiencies become constant over the sample. Their effect factorizes in the cut efficiencies and can be multiplied in after all other cuts.

The exact triggers that will be used in ATLAS for 14 TeV data taking have not been defined yet. Clearly, there will be triggers for single high-p_T muons and electrons. Combined triggers for lepton pairs may have lower p_T-thresholds for comparable rates. To stay on the conservative side, two single object triggers for muons an electrons respectively were examined. The muon trigger demands an isolated muon of 20 GeV/c, the electron trigger demands an isolated electron at 25 GeV/c.

In Figure 6.1, the trigger efficiency is plotted for events from the Whizard QCD sample with two leptons of $p_\mathrm{T}^\ell > p_\mathrm{T}^\mathrm{cut}$. The total trigger is considered passed, if one of the two individual triggers has fired.

$$\mathrm{eff} \equiv \frac{N(\text{trigger fired} \wedge {p_\mathrm{T}^{\ell,\mathrm{offline}}}_{1,2} \geq p_\mathrm{T}^\mathrm{cut})}{N({p_\mathrm{T}^{\ell,\mathrm{offline}}}_{1,2} \geq p_\mathrm{T}^\mathrm{cut})} \tag{6.1}$$

For cuts greater than 30 GeV/c, the trigger efficiency stays essentially constant at about 0.98. For the whole interval from 30 GeV/c to 100 GeV/c, the efficiency is 0.9816 ± 0.0014. In the following analysis a fiducial cut of 30 GeV/c^2 is included in

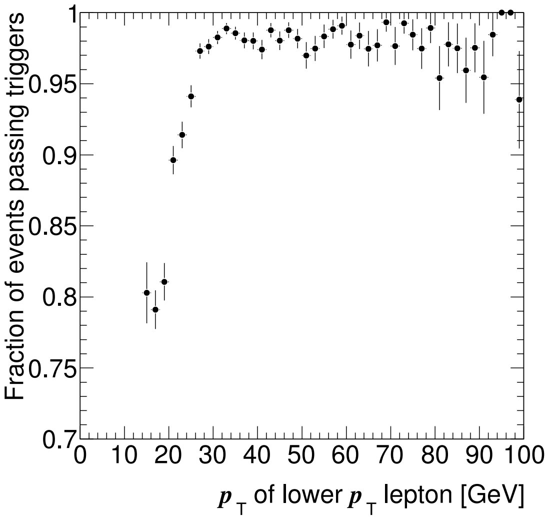


Figure 6.1: *Trigger turn-on curve for lepton triggers and fiducial lepton cut against low* p_T.

the precuts. For sensitivity and expected limit calculation, an efficiency of 0.98 is taken into account, though the effect is very small.

In the ATLAS expected performance paper [33] it was found that the isolated electron trigger efficiency deteriorates for electron momenta beyond $p_\mathrm{T} = 100\,\mathrm{GeV}/c$. This is not expected to have a large impact on this analysis. Furthermore, the trigger menus will change making it necessary to recalculate the efficiencies when the experiment will be performed.

By the time this analysis will be applied to data, trigger efficiencies will have been measured from data directly. The tag and probe method uses a neutral resonance that can decay into two leptons. In a sample where one lepton is found, one looks for a matching track of the other. From the efficiency with which the other lepton has caused the trigger to fire, one can deduce the actual trigger efficiency and becomes less dependent on Monte Carlo simulation.

6.2 Object Selection

When events have passed the trigger they are available for further analysis. In the various trigger levels, physics object reconstruction is more constrained for latency and bandwidth reasons than in off-line reconstruction. For all events, therefore, a completely separate off-line reconstruction is applied (cf. 3.3).

During data taking, most of the off-line object reconstruction will be done centrally before individual physics analyses are run. While reducing the chance to customize reconstruction on the analysis level, the advantages in terms of efficiency and overall consistency are considered more important. In particular it is not feasible to move all the raw unreconstructed data around the grid to consumers. This scheme still leaves a lot of flexibility in choosing reconstruction parameters by supplying alternatives in the reduced data. This analysis respects this by also using only the type of reconstructed object information from Monte Carlo that will be available to analyses on data.

6.2.1 Overlap Removal

The reconstruction output contains lists of reconstructed physics object candidates and supporting information. Among these are lists for electrons, muons, hadronically decaying taus, jets and a missing energy pseudo object. Physical objects may be reconstructed more than once, as the reconstruction algorithms each sift through a complete event independently. Consider a hypothetical real electron that has been correctly reconstructed as an electron candidate object in the list of reconstructed electron candidates in some event. At the same time, the real electron in the detector has often also caused the reconstruction as a jet candidate, since it leaves a localized amount of energy in the calorimeter. One and the same physics object may and often will have multiple reconstruction hypotheses. This “overlap” between different hypotheses is usually removed before event selection cuts are applied. An electron for example has the signature of a jet because it deposits energy in the calorimeter within a relatively small cone. In addition to that it tends to have a particular jet shape and lack of other energy entries in a certain radius around the main deposition. The latter is a somewhat more specific signature than that of a generic jet. Consequently, electron candidates that pass tight quality cuts are much more likely to be real electrons than real jets. Similarly, muons also have a rather clear signature if one demands that they be matched both in the inner tracker and muon spectrometer.

A rather robust method of overlap removal is to define a hierarchy of particle types which corresponds to the hierarchy of certainties with which they may be identified. Starting from the top, going down, the most reliable particle hypothesis for a particular object is accepted. For a practical implementation it is also important to define when two reconstructed object candidates are assumed to stem from the same real particle object. A simple and robust method is to consider candidate particle objects closer than some ΔR^{cut} from one another in $\eta \times \varphi$ space to overlap. For all sets of overlapping objects, the hypothesis highest in the hierarchy is accepted. Any reconstructed quantities, i.e. kinematics and charge are taken from the chosen reconstruction hypotheses, all others are completely rejected.

This prescription was followed in the present analysis, with a often used adjustment in the algorithm for efficiency reasons. Instead of first defining sets of reconstructed candidates that overlap and then choosing one of these candidates (a particular hypothesis), the order was turned around. For each list of reconstructed particles of a particular type, i.e. the list of electron candidates, the list of muon candidates and so on, all candidates passing quality cuts were accepted, unless a better candidate within ΔR^{cut} had been accepted before. Potentially overlapping object types of interest were electrons, muons and jets. Isolated electrons and muons hardly have any overlap with respect to one another. They do have more specific signatures than jets but are often falsely reconstructed as such. The concrete algorithm, therefore, started by immediately accepting all candidate electrons and muons passing quality cuts (cf. Section 6.2.2.) Then the list of candidate hadronic jets was examined. Those within a cone of $\Delta R^{\text{cut}} < 0.3$ of any previously accepted lepton candidate were considered to be falsely reconstructed as a hadronic jet and removed. All others were accepted if they passed the jet quality cuts.

6.2.2 Quality Cuts on Reconstructed Objects

Quality cuts on reconstructed object candidates may change the particle type assignment in the event. If for example the cuts on electrons are tightened, some reconstructed particle object candidates which would have been accepted as electrons are perhaps accepted as hadronic jets later on in the overlap removal chain. This results in a tight interdependence between object quality cuts and overlap removal. It is very involved to represent this in event level cuts, which is the main reason that quality cuts and overlap removal were separated out into a step before the main event selection.

Electrons For an electron candidate to be accepted, it has to be found within $|\eta| \leq 2.5$ and have a $p_{\text{T}} \geq 15\,\text{GeV}/c$. Additionally, electrons in a window of $1.35 < |\eta| < 1.57$ are rejected to avoid the gap between barrel and end cap of the electromagnetic calorimeter. Next to these kinematic cuts, isolation of the electron is demanded. Only electrons with less than $10\,\text{GeV}$ of transverse energy within a cone of radius $\Delta R < 0.2$ around the electron candidate are accepted. Energy deposits in the hadronic calorimeter are counted fully in the determination of the transverse energy. In the electromagnetic calorimeter a rectangular window of $\Delta\eta = 0.0625$ and $\Delta\phi = 0.0859$ is allowed as region of energy deposition for the electron candidate and not included in the limit calculation. Electrons also had to fulfill the ATLAS criteria for a tight electron identification. These consist of cuts on shower shape variables, such as longitudinal leakage and shape in the middle and first layer of the electromagnetic calorimeter. An appropriate track from the inner detector has to match to the clusters in the calorimeter. To reject photon conversions, a hit in the innermost pixel layer has to be found. Charged hadron contamination is further reduced by cuts on the ratio of high- to low-threshold hits in the TRT.

Muons Muons have to pass the following kinematic cuts: $|\eta| \leq 2.7$ and $p_{\text{T}} \geq 15\,\text{GeV}/c$. To ensure isolation, only $10\,\text{GeV}$ additional energy was allowed within $\Delta R < 0.2$ around the candidate. They have to have matching tracks from the inner detector and the muon system.

Hadronic Jets Hadronic jets are reconstructed using a cone algorithm with opening angle $\Delta R = 0.4$. Quality cuts are not as crucial for the jets as they are for electrons and muons since they are at the bottom of the overlap removal hierarchy. Only jets of $p_T \geq 15\,\mathrm{GeV}/c$ are accepted. Pseudorapidity is limited to $|\eta| \leq 4.2$. The latter ensures that generator level cuts on jets are safe.

6.3 Event Definition

Tagging Jets In addition to the physics objects that are centrally reconstructed, derived objects were defined on analysis level. The tagging jets which stem from the final state quarks in the $qq \to qq\ell\nu\ell\nu$ process are very characteristic due to their kinematics and correlation to other objects. From the list of jets it is a priori not clear, which of pair of reconstructed jets matches the quarks on parton level. One method suggested for vector boson fusion [85] was used in this analysis: For each hemisphere of the detector defined by the sign of η, the jet with the highest p_T is taken as a tagging jet candidate. Accordingly define:

$$\eta_1^{\mathrm{tag}} < 0.0 < \eta_2^{\mathrm{tag}} \tag{6.2}$$

Any further cuts are applied at event level.

Minijets The complement of tagging jets are minijets. In leading order only two jets are expected from the signal process. Due to the lack of color flow, additional jets are rarer between the tagging jets. With the tagging jet candidate definition given above, minijets are defined as those jets with pseudorapidity between the tagging jets. Formally, this includes all selected jets for which

$$\eta_1^{\mathrm{tag}} < \eta^{\mathrm{minijet}} < \eta_2^{\mathrm{tag}} \tag{6.3}$$

Leptons In the main selection cuts were applied to the leptons with the highest p_T, regardless of their flavor. Tau leptons are not considered as such. When decaying leptonically, their daughter leptons were subjected to the selection. The list of leptons is simply the p_T-ordered concatenation of the electron and muon lists.

***W* Bosons from Collinear Approximation** In the $Z \to \tau\tau \to \ell\nu\nu\ell\nu\nu$ process, the large mass difference between the Z boson and the tau leptons leads to large boosts for the latter. A subsequent leptonic decay of the taus will therefore have relatively little opening angle in the lab frame. One may assume that this angle approximately vanishes. Then the neutrinos from the reaction have the same initial direction as the daughter lepton. If the tau leptons are not themselves collinear, i.e. either back to back or highly boosted in the same direction, then the invariant mass of the Z boson may be reconstructed. Collinear approximation has proven very useful in this channel [86]. (For proposed use in searches, cf. [87, 88, 33].)

Letting the W boson take the role of the τ and a K-Matrix resonance that of the Z, one can apply the same method to the vector boson scattering channel. For a scalar isoscalar σ the subprocess would be $\sigma \to WW \to \ell\nu\ell\nu$. There are fewer neutrinos in the final state but the mass ratio is not as favorable.

$$\frac{m^W}{m^\sigma} \approx 5 \times \frac{m^\tau}{m^Z} \tag{6.4}$$

The validity of the collinear approximation had to be tested. For the $W \to \ell\nu$ vertex, let the fraction of W momentum $\vec{p}^\ell$ the daughter lepton is taking away be x.

$$x \equiv \frac{\vec{p}^\ell}{\vec{p}^W} \tag{6.5}$$

Using the lepton momenta and assuming the missing momentum p^{miss} being equivalent to the sum of the neutrino momenta, both momentum fractions can be calculated. The missing momentum p^{miss} is identified with the missing energy $E_{\text{T}}^{\text{miss}}$ as measured in the calorimeter.

$$x_1 = \frac{p^\ell_{1,x}p^\ell_{2,y} - p^\ell_{1,y}p^\ell_{2,x}}{p^\ell_{2,y}p^{\text{miss}}_x - p^\ell_{2,x}p^{\text{miss}}_y + p^\ell_{1,x}p^\ell_{2,y} - p^\ell_{1,y}p^\ell_{2,x}} \tag{6.6}$$

$$x_2 = \frac{p^\ell_{1,x}p^\ell_{2,y} - p^\ell_{1,y}p^\ell_{2,x}}{p^\ell_{1,x}p^{\text{miss}}_y - p^\ell_{1,y}p^{\text{miss}}_x + p^\ell_{1,x}p^\ell_{2,y} - p^\ell_{1,y}p^\ell_{2,x}} \tag{6.7}$$

Finally constraining the mass of the W bosons to be on-shell, the four-momentum of the hypothetical W boson pair may be reconstructed.

Lepton Centrality ζ A formal definition of ζ is given by Equation (2.19) in Section 2.5. The selected tagging jets and the two leptons with the highest p_{T} are used as inputs.

Lepton Azimuthal Angle Separation $\Delta\varphi^{\ell\ell}$ The angular correlations between the scattered bosons are dependent on the type of any additional resonances. In the case of W bosons, these correlations are propagated to the charged leptons in further decays. The low mass and therefore almost well-defined helicity of the neutrinos ensures strong angular correlations at the $W \to \ell\nu$ vertex. One useful angular correlation variable is $\Delta\varphi^{\ell\ell}$. Since the resonances in question are probed at 850 GeV/c^2 and 1150 GeV/c^2, they tend to be produced with a low transverse momentum. The scattered vector bosons are then on usually emitted in opposite directions in the lab frame and highly boosted. The boost is conferred to the daughter lepton and neutrino, smearing angular correlation from resonance spin. For all of the tested resonances $\Delta\varphi^{\ell\ell}$ tends to values close to π.

Transverse Momentum Balance $p_{\text{T}}^{\text{bal}}$ In the signal process and the irreducible backgrounds, two jets, missing energy and two leptons are in the final state of the hard process. These objects are approximately balanced against each other in transverse momentum $p_{\mathrm{T}}^{\mathrm{bal}}$.

$$p_{\mathrm{T,x}}^{\mathrm{bal}} = p_{1,x}^{\mathrm{tagjet}} + p_{2,x}^{\mathrm{tagjet}} + p_{1,x}^{\ell} + p_{2,x}^{\ell} + p_{x}^{\mathrm{miss}} \tag{6.8}$$

$$p_{\mathrm{T,y}}^{\mathrm{bal}} = p_{1,y}^{\mathrm{tagjet}} + p_{2,y}^{\mathrm{tagjet}} + p_{1,y}^{\ell} + p_{2,y}^{\ell} + p_{y}^{\mathrm{miss}} \tag{6.9}$$

$$p_{\mathrm{T}}^{\mathrm{bal}} = \sqrt{(p_{\mathrm{T,x}}^{\mathrm{bal}})^2 + (p_{\mathrm{T,y}}^{\mathrm{bal}})^2} \tag{6.10}$$

Transverse Mass $m_{\mathbf{T}}^{\mathbf{VBS}}$ In the fully leptonic channel, the invariant mass of the boson pair and hence the resonance are not directly accessible. Identifying p^{miss} with the neutrinos in the reaction as above, the transverse momentum of all final state particle candidates are available. Simply ignoring the component along the beam axis where p^{miss} is unknown, a transverse mass is defined as follows:

$$\begin{aligned} m_{\mathrm{T}} &\equiv |p_{1,\mathrm{T}}^{\ell} + p_{2,\mathrm{T}}^{\ell} + p_{\mathrm{T}}^{\mathrm{miss}}| \\ &= \sqrt{2\vec{p}_{\mathrm{T}}^{\ell\ell}\vec{p}_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\varphi(\vec{p}_{\mathrm{T}}^{\ell\ell}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}))} \end{aligned} \tag{6.11}$$

6.4 Fiducial Precuts

As a first step in the event selection, fiducial cuts were defined. Their purpose is to make sure that any effect that generator cuts might have is removed at analysis level.

For the reducible backgrounds, there is a cut on tagging jet candidates (see Section 5.2.) In order be be far from the thresholds of $\Delta\eta_{\mathrm{truth}}^{jj} > 2.0$ and $m_{\mathrm{truth}}^{jj} > 300\,\mathrm{GeV}/c^2$, after detector simulation cuts on $\Delta\eta^{jj}$ and m^{jj} were imposed.

$$\Delta\eta^{jj} \geq 3.0 \tag{6.12}$$

$$m^{jj} \geq 400\,\mathrm{GeV}/c^2 \tag{6.13}$$

Jets is already selected on object level using a transverse momentum cut, but that cut is tightened here to:

$$p_{\mathrm{T}1,2}^{j} \geq 20\,\mathrm{GeV}/c \tag{6.14}$$

The irreducible background samples and some reducible background samples also have cuts on leptons. To take these into account, two high-p_{T} leptons were demanded.

$$p_{\mathrm{T}1,2}^{\ell} \geq 30\,\mathrm{GeV}/c \tag{6.15}$$

These cuts have been introduced in Section 6.1 to ensure that the trigger efficiency is in the plateau region. To reduce $Z+\mathrm{jets}$ background from the start also a cut on the invariant mass of the lepton pair $m^{\ell\ell}$ was imposed.

$$m_{\ell\ell} \geq 150\,\mathrm{GeV}/c^2 \tag{6.16}$$

The cut was not further tightened to not remove all events in the $Z+\mathrm{jets}$ sample before the main selection (cf. Section 5.2.3.)

6.5 Multivariate Analyzer

6.5.1 Method

Having ensured via the fiducial precuts that the selected part of the phase space is reasonably described by the simulation, the next step in the analysis is to define the main event selection procedure. Its purpose is to help selecting a part of the remaining phase space where the change in observables due to the studied alternative models is large. This should in turn lead to a large sensitivity as determined by any subsequent statistical analysis. Ideally one would like to optimize all free parameters of any cut procedure for the largest sensitivity. This is not practical for involved statistical procedures. Instead, a substitute figure of merit is often used for optimization. In the case of an additive signal as in this thesis, possible choices include the signal fraction s/b or the classical approximation of significance $s/\sqrt{b}$. The correct calculation of the significances and hence sensitivity is done afterwards.

A classical approach to event selection accepts an event if it passes a list of cuts on event observables. In the phase space spanned by these observables, a cut selection defines a hypercuboid. By comparison, the unattainable theoretical optimum selection is defined by the true probability densities of signal and background $\mathrm{PDF}^{\mathrm{signal}}(\vec{o})$ and $\mathrm{PDF}^{\mathrm{background}}(\vec{o})$. In general the optimum accepted region in this phase space will not be a hypercuboid as selected by rectangular cuts. Instead it may be a more complicated structure. It does not have to be simply connected or even connected at all. In other terms, rectangular cuts ignore correlations between individual observables.

Actually estimating the true probability density functions from Monte Carlo simulation in any but the lowest dimensional phase spaces is not usually possible due to the lack of computing power to attain decent statistical uncertainties. An increasing fraction of phase space elements will simply not be filled. In order to find a compromise between the optimal selection and feasibility, many multivariate discriminators have been developed. As their name suggests they consider more than one variable at one time to model correlations. Discriminators based on probability density are among those but do not always perform optimally.

In the present analysis, implementations of multiple estimators were taken from the Toolkit for Multivariate Analysis (TMVA) [89]. In this framework, the whole work flow of training and initial testing is integrated for multiple discriminator types. Configurable parameters for each estimator and the training as a whole leave a lot of freedom for the user. A generic training consists of multiple phases, starting with the input of known signal and background samples. The samples need to be split up into training and testing samples to insure statistical independence of these two steps. For this analysis, the splitting was done automatically. Next is the actual training. The configured discriminators are each independently tested against the training data set. Since the correct classification is known, it is possible to use backpropagation techniques to change its output such that it is closer to the desired, known results. The exact backpropagation procedure is specific to each discriminator. After each training step, the remaining input data is used to evaluate a figure of merit. Training is repeated until satisfactory results are attained.

The reducible backgrounds immediately fit into the framework's concept of independent signal and background. As the signal definition for WHIZARD samples introduced in Section 5.4.2 is available on histogram level but TMVA only accepts individual events, a sample-level definition had to be substituted in this step. As signal, a sample with a scalar isoscalar resonance σ of mass $m = 850\,\mathrm{GeV}/c^2$ was chosen. The QCD and electroweak $qq \to qq\ell\nu\ell\nu$ sample was considered background. Overlap between those two samples is estimated using the electroweak $qq \to qq\ell\nu\ell\nu$ sample which is added with a weight of -1.0 to the background. However, it is only present in the TMVA tests, not the training, as negative weights are ignored for technical reasons. The $t\bar{t}$ sample has the same issue. Due to this changed definition, the signal and background definitions within TMVA are not the same as in the rest of the analysis. In particular the two figures of merit, signal efficiency ε_s and background efficiency ε_b, differ slightly. In terms of the statistical interpretation, this is unproblematic, as the test was redone outside of TMVA for this thesis.

Only one method was selected at the end for all possible resonances. While it is feasible that there might be some improvement by choosing individual optimizations, preliminary tests have not shown a large improvement. Furthermore, not varying the analysis with respect to model parameters has statistical advantages by avoiding the look elsewhere effect. This will be further detailed in the discussion of sensitivity in Section 7.3.3.

6.5.2 Boosted Decision Tree

Among the discriminators trained for this analysis were a multilayer perceptron, a probability density estimator and a boosted decision tree (BDT). A Fisher discriminator was trained as a robust baseline for quality comparison. Of the more sophisticated methods, the boosted decision tree implementation consistently had better performance and convergence behavior than the others. It was chosen accordingly as the method to be used for the main event selection.

A single decision tree is a set of branching cuts arranged in a hierarchical graph. Each inner node corresponds to a simple cut on any variable. Depending on either passing or not passing the cut, events are moved down the tree into one of the two next nodes. If this is still an inner node, then the procedure is repeated until a leaf node is reached. Each leaf node corresponds to either a signal or background assignment. This procedure uses rectangular cuts and can therefore only select hypercuboids in phase space as traditional cuts. However, it may select multiple hypercuboids in parallel, allowing for a much more flexible separation of background.

Trees are grown starting with one root node. The cut in this node is determined by finding the input observable and cut value with the best separation power for the whole sample. The two subsequent nodes repeat the same prescription for the corresponding subsample of events and so on. When the number of events in a node is below a predefined minimum, it becomes a leaf node. The majority of event weights decide whether it is a signal or background node. As figure of merit for the separation, the Gini index $p(1-p)$ was used, where p is the purity of the sample. Smaller values are better, indicating either pure signal or pure background. Cut direction, i.e. which is pass or fail does not matter, as the interpretation is decided at the end.

Decision trees have a very clear cut interpretation, but are vulnerable to statistical fluctuations. The iterative nature of the growing algorithm leads to a strong dependence of the nodes on slight changes in the upper nodes. Statistical fluctuations may lead to different observables to be optimal for different samples in one particular node, changing all nodes below this one. Often this also leads to fluctuations in total separation power of decision trees. This limitation can be ameliorated or overcome by using forests, instead of individual decision trees. Instead of relying solely on one tree, multiple trees are grown consecutively. For each new tree, events that have been misclassified in the previous tree are given a larger weight (are boosted). Each tree is given a weight β_i according to its classification performance. Each individual tree decision $h_i(\vec{o})$ for a given event $\vec{o}$ is mapped to $+1$ for signal decisions and -1 for background decisions.

$$h_i(\vec{o}) = \begin{cases} +1 & \text{for signal} \\ -1 & \text{for background} \end{cases} \tag{6.17}$$

The weighted average of these decisions is the response $r(\vec{o})$ of the whole forest to a particular event:

$$r(\vec{o}) \equiv \frac{1}{N} \sum_{i=1}^{N} \beta_i h_i(\vec{o}) \tag{6.18}$$

Typically, both the stability and performance of a forest is greatly improved by boosting. This can be combined with bagging for additional stabilization. For each new tree in the forest, a subsample is randomly selected from the training sample. This simulates statistical fluctuations in the inputs and makes the forest more robust against similar fluctuations in the test samples and ultimately data.

6.5.3 Input Variables

One advantage of boosted decision trees is their ability to cope with a relatively large number of input variables. Other discriminators are often limited by training procedures that exhibit expensive scaling with the number of inputs. For each individual decision tree node, the number of distributions to test is equivalent to that number. Hence, the algorithm scales roughly linear at that point. There is no expensive and in some cases not well-behaved minimization procedure as for neural nets. A potential disadvantage is the risk of overtraining. Boosting and bagging help to mitigate these effects.

From the automatic choice of variable importance in each decision tree, a ranking of variable importance may be derived for a trained tree [90]. For each use in a node, the squared Gini index gained by the node times the number of events in that node is counted towards a total weight. The same definition also works for forests of a

Rank	Input Variable	
1	Largest Flavor Tag Weight	b_1
2	Largest Lepton Transverse Momentum	$p_{\mathrm{T},1}^{\ell}$
3	Invariant Tagging Jet Mass	m^{tagjet}
4	Second To Largest Lepton Transverse Momentum	$p_{\mathrm{T},2}^{\ell}$
5	Tagging Jet Pseudorapidity Separation	$\Delta\eta^{\mathrm{tagjets}}$
6	Transverse Mass	$m_{\mathrm{T}}^{\mathrm{VBS}}$
7	Second To Largest Flavor Tag Weight	b_2
8	Tagging Jet Transverse Momentum ($\eta < 0.0$)	$p_{\mathrm{T},2}^{\mathrm{tagjet}}$
9	Missing Transverse Momentum	$p_{\mathrm{T}}^{\mathrm{miss}}$
10	Lepton Centrality	ζ
11	Tagging Jet Transverse Momentum ($\eta \geq 0.0$)	$p_{\mathrm{T},1}^{\mathrm{tagjet}}$
12	Transverse Momentum Balance	$p_{\mathrm{T}}^{\mathrm{bal}}$
13	Largest Minijet Transverse Momentum	$p_{\mathrm{T},1}^{\mathrm{minijet}}$
14	Second To Largest Minijet Transverse Momentum	$p_{\mathrm{T},2}^{\mathrm{minijet}}$
15	Lepton Momentum Fraction 1	x_1
16	Lepton Momentum Fraction 2	x_2

Table 6.1: *Table of the input variables used in the BDT. Ranked in order of importance from most import to least important by a heuristic figure of merit. The order is not deterministic and may change with retraining. E.g. $p_{\mathrm{T},1}^{\mathrm{tagjet}}$ and $p_{\mathrm{T},2}^{\mathrm{tagjet}}$ should be equivalent but their ranks are not consecutive.*

boosted decision tree. While this ranking has to be considered heuristic, it gives a rough idea of what variables are important overall. A ranked list of input variables can be found in Table 6.1.

The training used very similar but less strict precuts as the later analysis. Main purpose was to ensure that the main parameters were defined for passing events. To achieve this, the presence of two tagging jets and two leptons was demanded. Transverse momenta of tagging jets and leptons had to be larger than 20 GeV/c. Pseudorapidity separation $\Delta\eta^{\mathrm{tagjets}}$ of the two tagging jets needed to be equal or exceed 3.0. Distributions after precuts are shown in Figures 6.2 to 6.5.

Flavor Tag Weight The output of the b-tagging algorithms is a continuous variable rather than a yes/no decision. In cut analyses this allows to choose a working point of given b-tag efficiency and corresponding light-jet rejection. For a multivariate discriminator this is also very beneficial as the tree can effectively choose its working point during training. Several flavor tagging algorithms are available in ATLAS. The standard algorithm for 14 TeV is the combined IP3D+SV1 tagger (cf. 3.3.5) which was used here. The two highest flavor tag weights b_1 and b_2 were passed to the BDT. Undefined values may occur when the algorithms were not successful to find at least two jets passing the jet precuts. In this case, a default value of -10.0 was passed into the tree both in the training and in the later tests. Data would have to

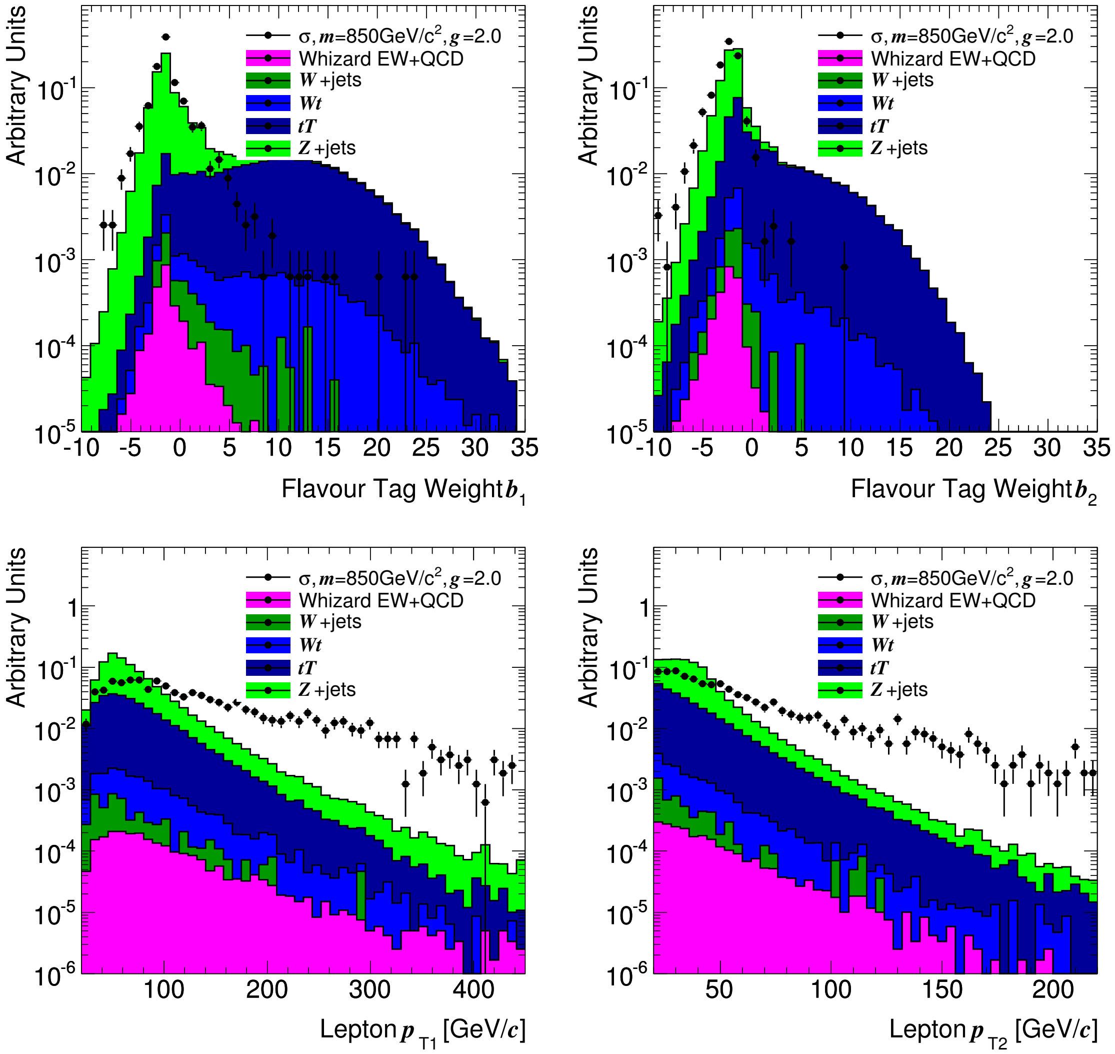


Figure 6.2: *Distributions of BDT input variables for training. BDT training precuts are applied. Signal and sum of all backgrounds are independently normalized to unity. Relative scaling within the backgrounds is preserved.* $W+$jets *and* $Z+$jets *were not part of the training.*

be treated identically. The value was chosen close to the lower edge of the range of possible flavor-tag weights to indicate an absence of bottom quarks.

As the signal does not usually contain real bottom quark jets, the BDT is expected to veto events with high weights. Both background processes including top quarks have bottom quarks in the tree level final states and should be particularly well separable. Accordingly, the two weight variables took place one and seven in the BDT variable ranking.

Lepton Transverse Momenta $p^{\ell}_{\mathrm{T},1}$, $p^{\ell}_{\mathrm{T},2}$ An additional massive resonance leads to highly boosted decay products. Accordingly, the resulting leptons tend to have a larger transverse momentum p_{T} than most backgrounds. The two charged leptons with the highest transverse momenta $p^{\ell}_{\mathrm{T},1}$ and $p^{\ell}_{\mathrm{T},1}$ were fed into the BDT for training. The precuts ensured that both variables were defined for any event passed into the BDT. Their variable ranks are two and four, respectively.

Tagging Jet Invariant Mass m^{tagjet} Tagging jets from vector boson scattering tend to be rather energetic as they are produced directly from valence quarks. More importantly, they have on average a large separation in η. This increases the invariant dijet mass compared to background giving the variable a rank of three. Presence of tagging jets is ensured by precuts.

Tagging Jet Pseudorapidity Separation $\Delta\eta^{\mathrm{tagjets}}$ This variable, nother characteristic property of tagging jets from vector boson scattering, is highly correlated to the invariant dijet mass above. At rank five it is also one of the most important ones, despite the correlation.

Transverse Mass $m_{\mathrm{T}}^{\mathrm{VBS}}$ For the signal, the transverse mass of the leptons and $p_{\mathrm{T}}^{\mathrm{miss}}$ is related to the invariant mass of the new resonance. While is is smeared out so much that no edge clearly related to the mass is visible, there is an enhancement for large masses. The remaining $Z+$jets background shows the opposite behavior, producing naturally neither a large $E_{\mathrm{T}}^{\mathrm{miss}}$ nor a large $m^{\ell\ell}$.

Tagging Jet Transverse Momenta $p^{\mathrm{tagjet}}_{\mathrm{T},1}$, $p^{\mathrm{tagjet}}_{\mathrm{T},2}$ In the signal, the tagging jets need to balance the vector bosons merging into a high-mass resonance. On average, this leads to high transverse momenta for the tagging jets $p^{\mathrm{tagjet}}_{\mathrm{T},1}$ and $p^{\mathrm{tagjet}}_{\mathrm{T},2}$. There is no ordering between the tagging jets, as they have each the highest p_{T} in their respective detector hemisphere. Therefore the distribution of the two variables is the same within statistical fluctuations. They occupy rank eight and eleven.

Missing Transverse Momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ Both signal and background processes fed into the BDT have neutrinos in the final state. These increase the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ in the detector. For signal events the value is slightly higher, analogous to the higher lepton transverse momenta. The variable has a medium rank of nine but with

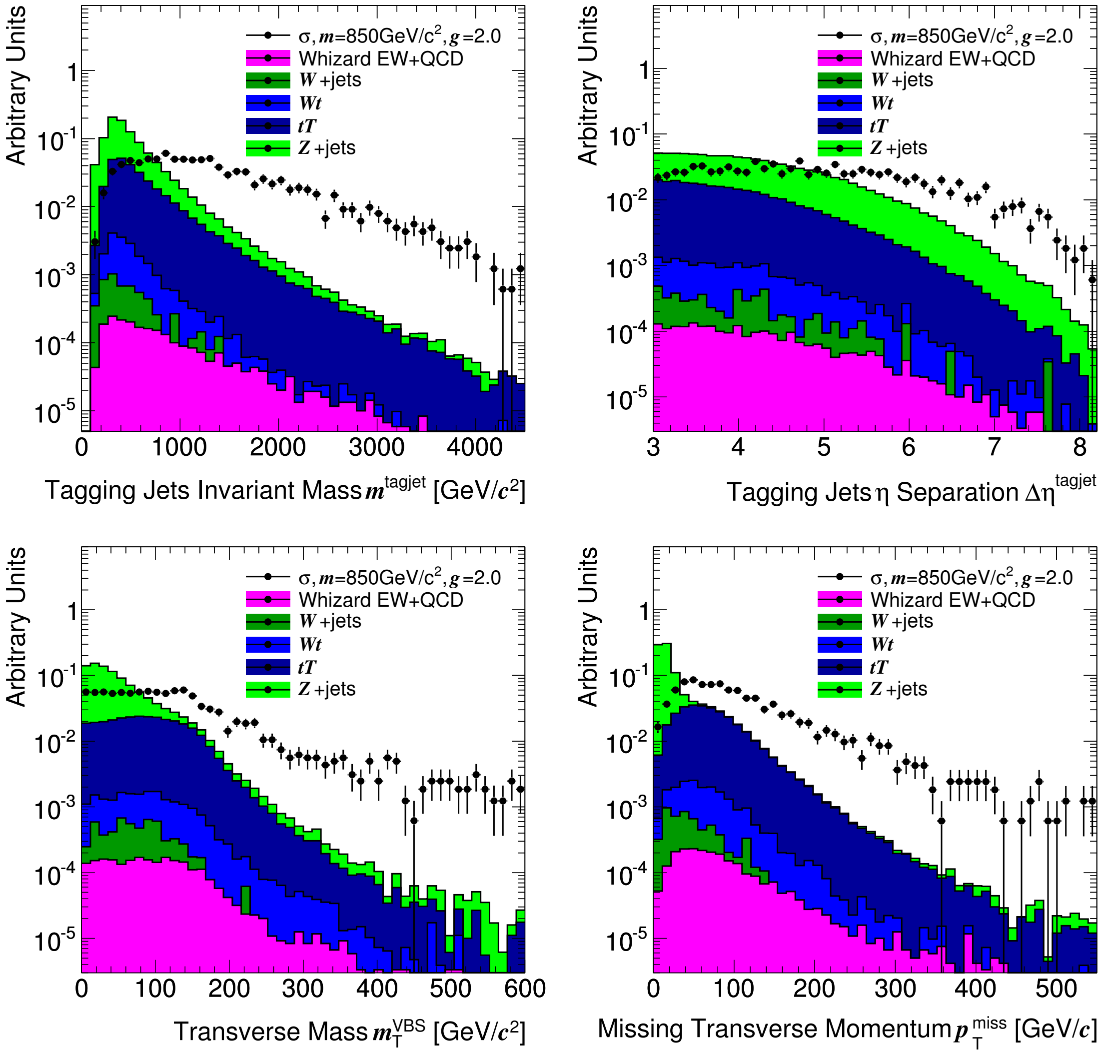


Figure 6.3: *Distributions of BDT input variables for training. BDT training precuts are applied. Signal and sum of all backgrounds are independently normalized to unity. Relative scaling within the backgrounds is preserved.* $W+\text{jets}$ *and* $Z+\text{jets}$ *were not part of the training.*

$Z+$jets included it becomes more important, as this channel lacks neutrinos on tree level to increase $p_{\mathrm{T}}^{\mathrm{miss}}$.

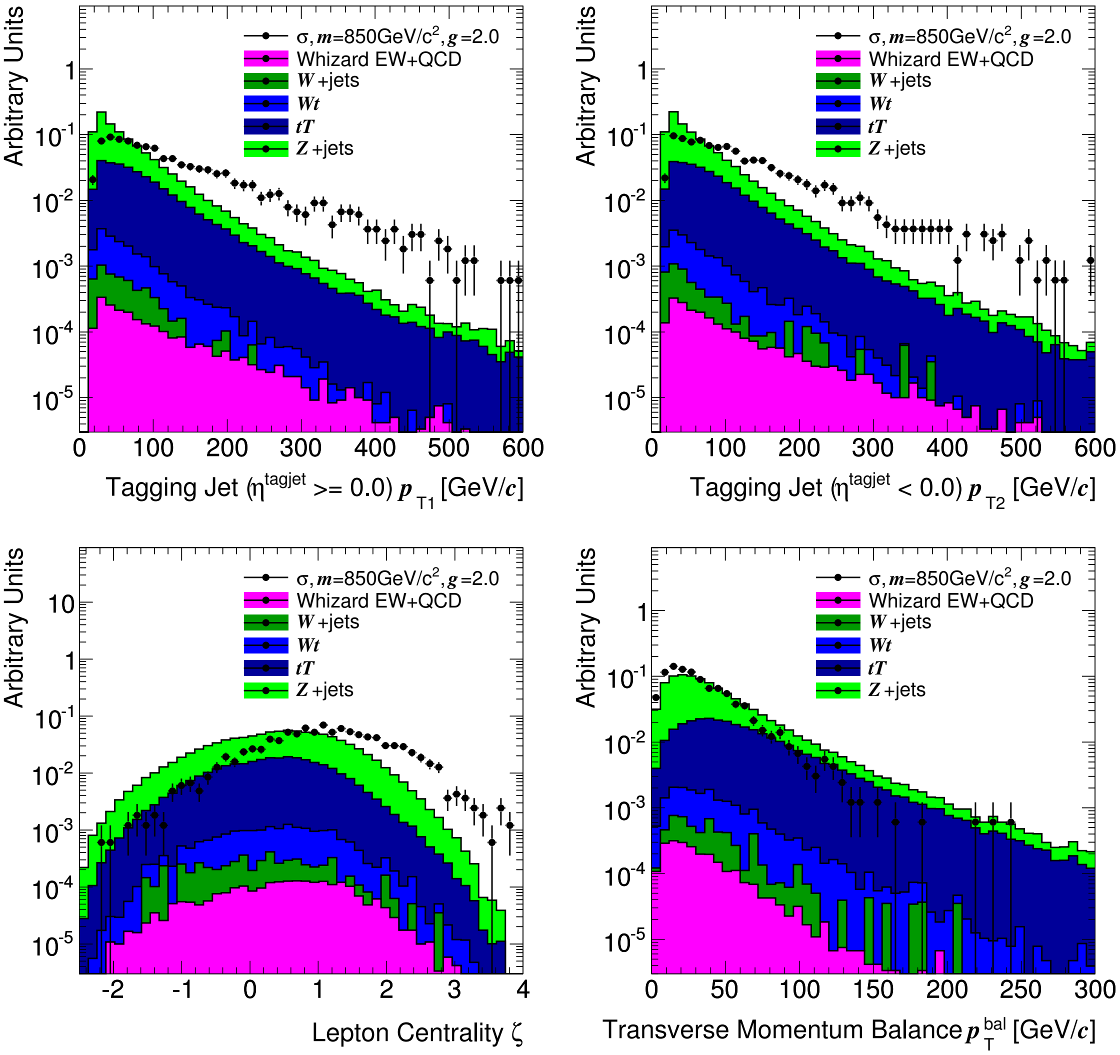


Figure 6.4: *Distributions of BDT input variables for training. BDT training precuts are applied. Signal and sum of all backgrounds are independently normalized to unity. Relative scaling within the backgrounds is preserved.* $W+$jets *and* $Z+$jets *were not part of the training.*

Lepton Centrality ζ Lepton centrality is only ranked at ten, despite the clear difference between signal and background. Strong correlations to tagging jet variables which have similar separation power play a role. If most of the separation power of this variable is already included in one which is slightly better, the other variable will be used preferentially. Reduced BDT with only one of each set of strongly correlated variables performed consistently worse, however, so the larger selection of input variables was retained.

Transverse Momentum Balance $p_{\mathrm{T}}^{\mathrm{bal}}$ Next in order of importance is the transverse momentum balance at position twelve. While by construction a variable that is specifically tailored to the signal process, the actual separation is not the strongest of the variables. All reducible backgrounds have longer tails than the signal as expected.

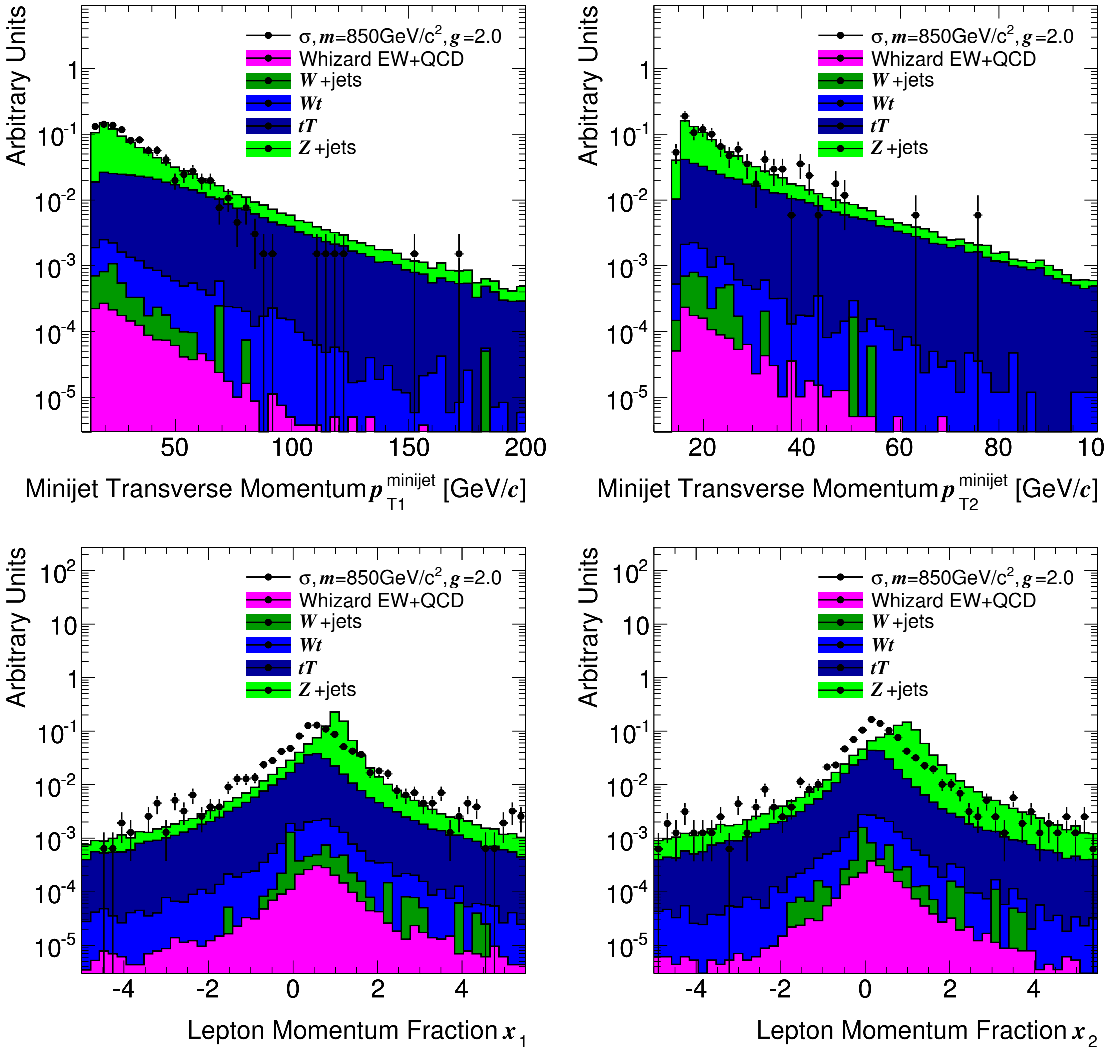


Figure 6.5: *Distributions of BDT input variables for training. BDT training precuts are applied. Signal and sum of all backgrounds are independently normalized to unity. Relative scaling within the backgrounds is preserved.* $W+\text{jets}$ *and* $Z+\text{jets}$ *were not part of the training.*

Minijet Transverse Momentum $p_{T,1}^{\text{minijet}}$, $p_{T,2}^{\text{minijet}}$ A minijet veto is usually realized via a cut on the largest transverse momentum of any minijet in the event. The highest and second highest momenta, $p_{\mathrm{T},1}^{\text{minijet}}$ and $p_{\mathrm{T},2}^{\text{minijet}}$ were given as inputs to the BDT. When not defined, a default of $-1.0\,\text{MeV}$ was substituted. While this value still contains the information that the minijet was not present at all, it is not far enough from the distribution to induce binning problems during training. While the individual distributions show separation power, the overall importance is not very large. As there are still large theory uncertainties associated with this variable, this fact is expected to make the overall result more robust.

Lepton Momentum Fraction x Using the collinear approximation the lepton momentum fraction may be defined according to Equation (6.5). If the approximation is valid, the two fractions are limited by the initial boson momenta.

$$0.0 < x_{1,2} < 1.0 \tag{6.19}$$

In the case of background, the W bosons are not decay products of one resonance and are thus not limited. While this approach has proven to be very useful in the decay of a Z boson to tau leptons, for a K-Matrix resonance to vector bosons the collinear approximation is not valid. A plot of the reconstructed invariant mass distribution of the hypothesized W pair m^{WW} is shown in Figure 6.6. No resonance is visible. This is also reflected in the $x_{1,2}$ distributions, where there is no shape difference between signal and background. A notable exception in both cases is the $Z+\text{jets}$ background. As the Z bosons decay directly into leptons, and at the same time, the missing transverse momentum is small, $x_{1,2}$ peak at 1.0. The reconstructed W pair mass from collinear approximation tends to approach $m^{\ell\ell}$ for small $p_{\mathrm{T}}^{\text{miss}}$ but is limited by the minimum of the two assumed W boson masses. Hence, $Z+\text{jet}$ events show a peak close to the limit. The lepton momentum fractions are thus hardly used by the BDT.

6.5.4 Performance

With the input variables given in the preceeding section, a BDT was trained using a subset of the Monte Carlo samples introduced in Section 5. The signal was represented by a scalar isoscalar resonance σ of mass $m = 850\,\text{GeV}/c^2$ at the original sample coupling of $g = 2.0$. As background samples, $t\bar{t}$, Wt and the WHIZARD QCD and electroweak samples were included in the training, $W+\text{jets}$ and $Z+\text{jets}$ were not. $W+\text{jets}$ lacks a second lepton from the hard process and is already strongly suppressed by the precuts. $Z+\text{jets}$ has the right number of leptons, but from the Z resonance of much lower mass than the K-Matrix resonances. Since the sample was also produced with a problematic $m^{\ell\ell}$ cut, it was excluded from the training to avoid unphysical cuts in the BDT. TMVA was configured to randomly sample half of the events for training. The other half was used internally for performance testing and to determine when to stop the training iterations.

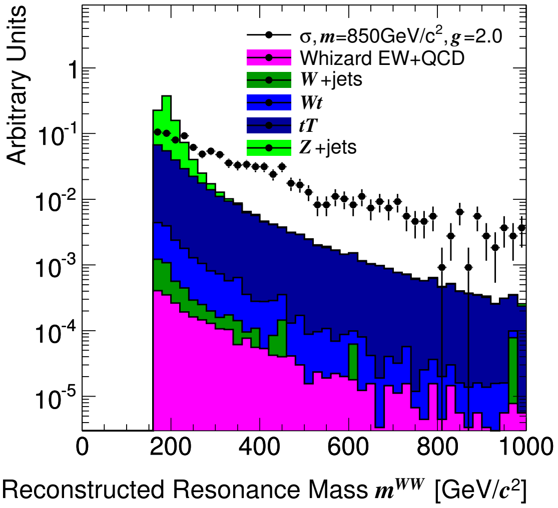


Figure 6.6: *Distribution of the reconstructed invariant WW mass from collinear approximation for comparison. BDT training precuts are applied. Signal and sum of all backgrounds are independently normalized to unity. Relative scaling within the backgrounds is preserved.*

A strict separation of these datasets is crucial to avoid overtraining. If the same data were used, a discriminator with enough free parameters would be able to ultimately learn the properties of individual events and reach perfect separation of the training sample. Performance on real data would deteriorate in this case. An independent test sample makes this very unlikely. The BDT response after training is plotted in Figure 6.7 separately for signal and background training and test samples. A statistically implausible difference between training and test sample results would be an indication of overtraining. For the background samples the agreement is excellent. A Kolmogorov-Smirnov test yields a result close to one. In the signal case the agreement is still good, but the Kolmogorov-Smirnov test result is slightly above 5%. Still, this is not exceedingly unlikely to be an effect due to statistical fluctuations.

The overtraining test was done at TMVA level. Sensitivities were then determined using the trained BDT but the analysis was performed completely outside of TMVA. The signal definition from Section 5.4 was used with the appropriate fiducial precuts. All background samples including $W+$jets and $Z+$jets and all available simulated events were used. This way, training events from the $t\bar{t}$, Wt, WHIZARD QCD+EW and one signal sample were reused, since otherwise the available number events would drop to unacceptable levels. If there was overtraining for the signal in the BDT, the results for this one signal sample would be too optimistic. For all other samples, including the scalar isoscalar sample at mass $m = 1150\,\mathrm{GeV}/c^2$, the results would still be statistically valid. While performance for the scalar isoscalar resonance is best, that is true for both masses. Any overtraining effect would therefore have to be small.

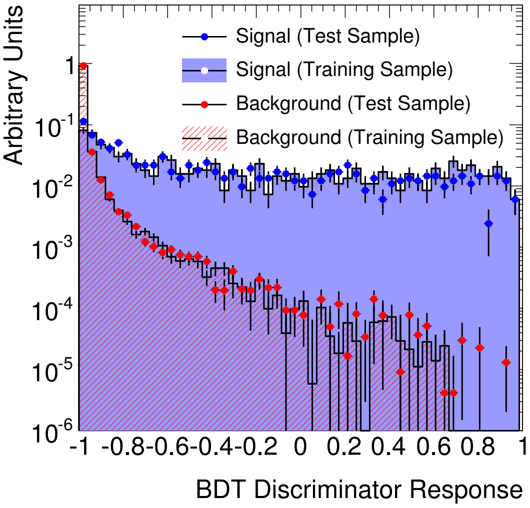


Figure 6.7: *Distribution of BDT discriminator responses for the signal samples and the background. Signal definition according to the BDT inputs. Histograms represent the training samples, points the test samples.*

For the analysis, events are selected which have a BDT response r greater or equal to a cut value r^{cut}. In Figure 6.8 the expected number of events after all cuts for $\mathscr{L} = 100\,\text{fb}^{-1}$ integrated luminosity is shown for each process versus the cut on the BDT response. This is analogous to a cutflow diagram for multiple individual cuts. Most prominent feature is the fact that from a cut value of $r \geq 0.6$ onward, the reducible backgrounds have all vanished. Only irreducible background remains. Of the reducible backgrounds, $W +$ jets quickly disappears and is irrelevant for further analysis. $Z +$ jets survives longer but is susceptible to harder cuts on $m^{\ell\ell}$ and $E_{\text{T}}^{\text{miss}}$ and will also be neglected. For the electroweak component of $Z +$ jets no samples were available, but its lower cross section should allow it to be neglected as well. The two top quark containing processes $t\bar{t}$ and Wt warrant further investigation. While they do vanish completely in simulation, the amount of simulated events is limited. A lack of selected events might thus be due to a statistical fluctuation downwards. In the analysis of sensitivity the statistical uncertainty on the Monte Carlo simulation was taken into account (cf. Section 7.1.)

6.5.5 Lepton Azimuthal Angle Separation $\Delta\varphi^{\ell\ell}$

After the main selection cut on the BDT response, there is no reducible background left. A likelihood fit on data to improve the estimation of the signal fraction using shape information after all cuts is only feasible for remaining background, i.e. irreducible WHIZARD QCD+EW. The lepton azimuthal angle separation $\Delta\varphi^{\ell\ell}$ is a candidate for such a likelihood fit. Figure 6.9a shows the $\Delta\varphi^{\ell\ell}$ distribution before cuts. Other backgrounds are neglected in this plot to avoid having to show a shape-

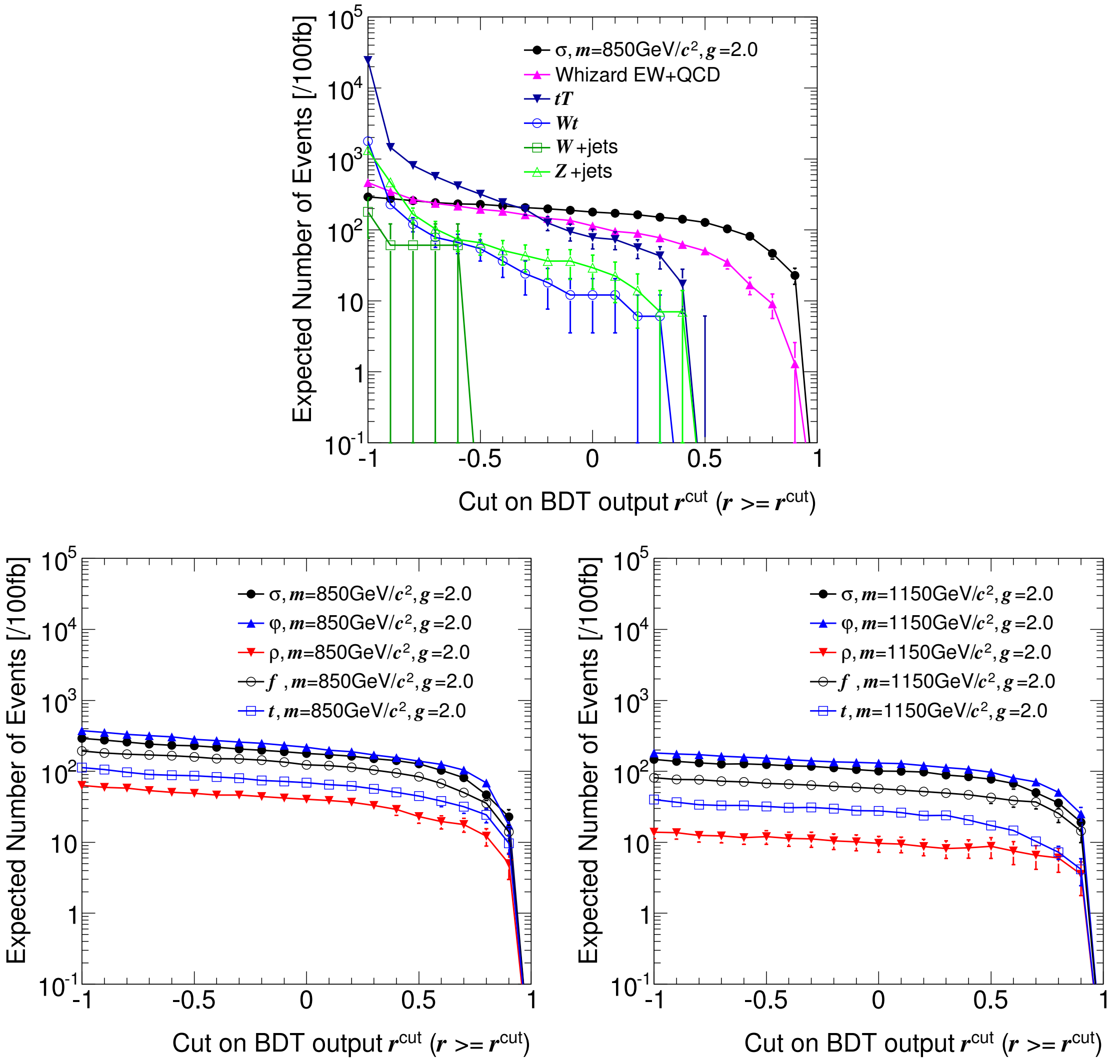
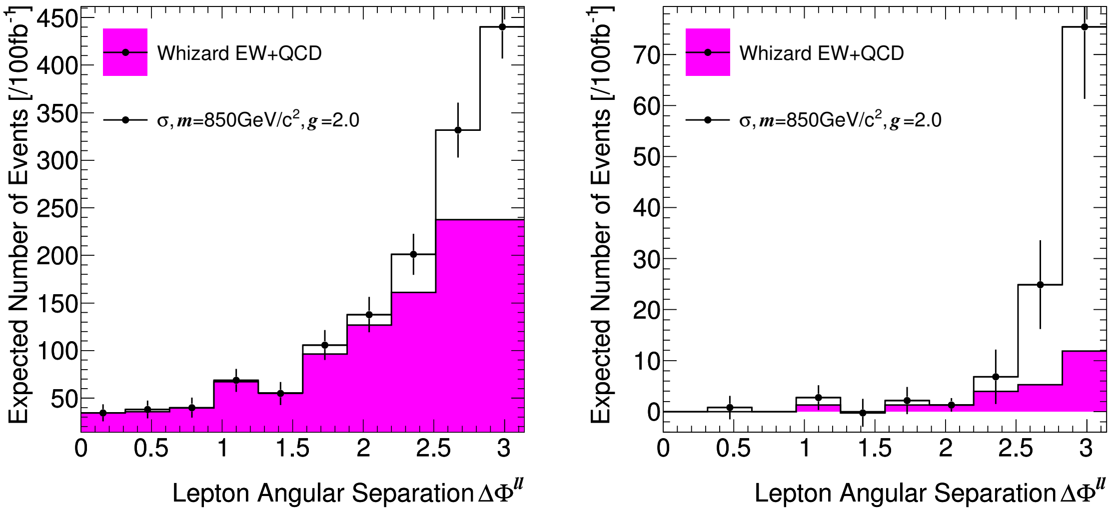


Figure 6.8: *Expected number of events after all precuts and an additional cut on $\Delta\varphi^{\ell\ell} \geq 0.8\pi$ versus the BDT cut value r^{cut}. In the upper plot, backgrounds and scalar isoscalar resonance σ at $m = 850\,\mathrm{GeV}/c^2$ are shown. Most backgrounds quickly disappear, only irreducible background is left for hard cut values. On the lower left and right, the different signals are compared to one another at $m = 850\,\mathrm{GeV}/c^2$ and $m = 1150\,\mathrm{GeV}/c^2$, respectively. Signal definition according to Section 5.4.*

distorting log scale.

In order to keep the distribution mostly undisturbed, $\Delta\varphi^{\ell\ell}$ has not been used for any precuts or the BDT. However, after event selection, the shape difference between signal and background becomes a lot smaller as the background now also peaks for high $\Delta\varphi^{\ell\ell}$. Distributions after a BDT cut of $r \geq 0.7$ are shown in Figure 6.9b. The BDT seems to have learned to use a lot of the information contained in $\Delta\varphi^{\ell\ell}$ via correlations from other variables.



(a) Before BDT Cut

(b) After BDT Cut $r \geq 0.7$

Figure 6.9: *Expected number of events after all precuts versus $\Delta\varphi^{\ell\ell}$. The scalar isoscalar resonance σ at $m = 850\,\mathrm{GeV}/c^2$ and the* WHIZARD *QCD and electroweak background are stacked. Other backgrounds not shown. Left: No BDT cut has been applied. The irreducible background has a significantly different shape than the signal. For low values of $\Delta\varphi^{\ell\ell}$ a control region for background estimation seems to be present. Right: The same situation after BDT cut. The irreducible background peaks as well, the control region has little background and a comparable amount of signal left.*

The potential control region for low $\Delta\varphi^{\ell\ell}$ contains little background after cuts and at the same time a comparable amount of signal. A background estimation or a fit would therefore yield little more information than a cut-and-count experiment in the high $\Delta\varphi^{\ell\ell}$ region. The statistical shape uncertainties in the fit would on par with the absolute statistical uncertainties. Instead of a fit in the conserved variable $\Delta\varphi^{\ell\ell}$, a additional simple cut at $\Delta\varphi^{\ell\ell} \geq 0.8\pi$ was performed. The number of events expected after all cuts for the different processes was then used to setup the statistical analysis.

7 Sensitivity and Limits

7.1 Statistical Model

7.1.1 Hypothesis Test Setup

Let the hypothesis that a signal is present with a certain coupling be called $\mathrm{H}_1(g)$ and the complementary SM-like hypothesis H_0. By construction, $\mathrm{H}_1(g = 0)$ corresponds to H_0. Ten different types of H_1 correspond to the ten considered resonance parameter sets. Unless there is risk of confusion these are not differentiated explicitly. The event yield in each signal hypothesis is composed partly by interactions such as $t\bar{t}$ and irreducible SM-like vector boson scattering. Together they make up the background. Its true expectation value in the experiment is denoted as $b = \sum_i b_i$. Any additional, increased yield from vector boson scattering when $g \neq 0$ is caused by the signal in a more narrow sense of the word and is correspondingly called s as in 5.4. Per construction, the total yield expectation value is $s + b$.

Implicitly, destructive interference leading to a negative value of s is ignored. This is typically done in many analyses. In fact, for processes that are generated in different samples there is no way to know the interference. In this thesis, it was actually possible to see the interference within the $qq \to qq\ell\nu\ell\nu$ channel which was generated all in one. While negative s have been seen for some resonances and small g, the effect was too small to be discovered on its own and could thus be neglected. There are higher order interference effects linking vector boson scattering to QCD processes, which are not included in the tree-level matrix elements [91].

In order to make a discovery, one has to try to reject H_0, typically with a confidence of $1 - \mathrm{CL} \approx 2.82 \times 10^{-7}$. This corresponds to a one-sided Gaussian tail at 5σ. A detailed overview of how the confidence intervals are calculated is given in the following sections. Before the experiment has collected data the result is unknown. To quote the discovery sensitivity, i.e. the expected significance of a discovery, one has to assume $\mathrm{H}_1(g)$ to be realized in nature for some g. From this follow expectation values for the experimental results while the actual outcome may fluctuate. One could simulate these in possible fluctuations in detail, learning something about the fluctuation of the significance given a particular signal hypotheses. The most interesting piece of information, though, is the median expected significance. This can be estimated by using the expectation values for the experiment as a test sample to calculate the significance. Such test data is called an Asimov [33] data set.

Analogously one can define upper limits. Should H_0 be true, which is the lowest g still compatible with the expected data? By convention one asks for a confidence of

95% for this upper limit.

7.1.2 Likelihood Function

As seen in Section 6.5.5, there is currently little to be gained from setting up a multi bin fit to determine a signal fraction. Instead, the current analysis uses a counting method, i.e. single bin fit. The former is a generalization of the latter from one bin to many. Most ideas of the following statistical approach apply to either setup, but for clarity only one the one-bin version is explained. The generalization to multiple bins is straight forward (cf. [33].)

The LHC's huge luminosity, while a boon in the long run, poses new statistical challenges. Since the necessary background rejection factors are so large or conversely efficiencies so small, it is not anymore possible to generate as many Monte Carlo events as one would like. For backgrounds in particular this leads to sizable statistical uncertainties on the background estimates up and above other uncertainties. These can no longer be ignored but have to be somehow incorporated in the statistical treatment.

Here, the approach from [92, 33] is adopted. It uses a likelihood method to attain frequentist limits. Using the definitions from Section 7.1, the likelihood of observing n events in the experiment is simply:

$$L(n; s, b) = \frac{(s+b)^n e^{-(s+b)}}{n!} \tag{7.1}$$

Following the procedure from [92, 33], one can extend the experiment conceptually to encompass the Monte Carlo generation as well. From a frequentist point of view, a repetition of the experiment is not just taking the same amount of data again, but also rerunning all of the Monte Carlo simulations. Considering unweighted for the time being, the likelihood can simply be extended to cover the Monte Carlo samples. For some number of individually generated background samples, let the simulated yields be m_i. Each of them is a draw from a Poisson process. Their expectation values are not simply b_i because the samples are generated for a different luminosity than the experiment is expected to use. Instead, the expectation value becomes $\tau_i b_i$, where $\tau_i \equiv L_i^{MC}/L_i^{exp}$ is assumed to be known.

$$L(n, m_i; s, b_i) = \frac{(s+b)^n e^{-(s+b)}}{n!} \Pi_i \frac{(\tau_i b_i)^{m_i} e^{-(\tau_i b_i)}}{m_i!} \tag{7.2}$$

The Monte Carlo estimates m_i of b_i act as constraints in the fit, allowing the handling of the b_i as nuisance parameters. This setup allows to put limits on s, however, one is really interested in limits on g. While it is conceivable to convert from s to g after the fit, this ignores the statistical uncertainty in $s(g)$ which ought to be included just as the uncertainties on the backgrounds. For discovery, only $g = 0$ has to excluded, which has no uncertainty on s, as $s(g = 0) = 0$ by definition. In Higgs boson searches within ATLAS this technique had only been used for discovery sensitivity estimation. In this work, to include $s(g)$ uncertainties for upper limits on g, the likelihood needed to be extended. There is a slight catch here, that prevents the analogous use of a Poisson distribution. The events for signal samples are reweighted to their respective values of g, which would necessitate using a large number of Poissonians binned by weights to model correctly. Instead, the Gaussian approximation derived in 5.4.2 was used. With $s(g)$ being the true expectation value for the signal in the experiment, let $s^{\mathrm{MC}}(g)$ be the signal estimator from Monte Carlo simulation.

$$\begin{aligned} L(n, m_i; g, b) = &\frac{(s(g)+b)^n e^{-(s(g)+b)}}{n!} \Pi_i \frac{(\tau_i b_i)^{m_i} e^{-(\tau_i b_i)}}{m_i!} \\ &\times \frac{1}{\sigma_s(g)\sqrt{2\pi}} e^{-\frac{(s^{\mathrm{MC}}(g)-s(g))^2}{2\sigma_s(g)^2}} \end{aligned} \tag{7.3}$$

$t\bar{t}$ background has weights -1 and 1 and would therefore warrant special treatment in principle. Since the admixture of negative weights is only $\approx 15\%$, they are not subdivided. All expectation values take the weights correctly into account.

Systematic uncertainties further reduce the power of the test and need to be accounted for. There are potentially many ways to include them in the likelihood. Here, a generic linear transfer function from the underlying uncertainties to each sample's yield is assumed. $s(g)$ and b_i retain their meaning as expectation values of the experiment. The terms of the likelihood pertaining to Monte Carlo simulation are replaced by new expectation values b_i' and $s'(g)$.

$$\begin{aligned} b_i' &= b_i - \sum_j c_{ij} e_j \\ s'(g) &= s(g) - \sum_j c_j^{\mathrm{s}} e_j \end{aligned} \tag{7.4}$$

c_{ij} and c_j^{s} are linear transfer constants and e_j the deviations from the expected values for some systematic variable numbered j.

How to constrain the e_j? Typically, they are taken to follow a Gaussian distribution. To be more formally frequentist one ought to say, that the measurements from which they have been estimated follow a Gaussian distribution. In general this is an approximation. Since a Gaussian distribution is symmetric under exchange of mean and variable, there is no quantitative difference between the two points of view as long as the variance is independent of the mean. Therefore, a Gaussian distribution of a measurement e_j^{obs} around an unknown true parameter e_j is included for each systematic uncertainty.

$$\begin{aligned} L(n, m_i; g, b_i, e_j) = &\frac{(s(g)+b)^n e^{-(s(g)+b)}}{n!} \Pi_i \frac{(\tau_i b_i')^{m_i} e^{-(\tau_i b_i')}}{m_i!} \\ &\times \frac{1}{\sigma^{\mathrm{s}}(g)\sqrt{2\pi}} e^{-\frac{(s^{\mathrm{MC}}(g)-s'(g))^2}{2\sigma^{\mathrm{s}}(g)^2}} \\ &\times \Pi_j \frac{1}{\sigma_j\sqrt{2\pi}} e^{-\frac{(e_j^{\mathrm{obs}}-e_j)^2}{2\sigma_j^2}} \end{aligned} \tag{7.5}$$

For the determination of the limits, the e_j^{obs} are taken to vanish – the expected value is calibrated away. They are only needed for ensemble tests (see Section 7.1.4).

7.1.3 Confidence Interval Construction

To calculate expected limits and sensitivities, a confidence interval construction [93] was used in this thesis. The definition of the confidence interval demands a certain coverage i.e. confidence level (CL) of the interval, but does not constrain the choice of points to include any further. Feldman and Cousins suggest a method [94] that uses a likelihood ratio to make that choice.

$$\lambda(n, m_i; g, b_i, e_j) = \frac{L(n, m_i; g, b_i, e_j)}{L(n, m_i; \hat{g}, \hat{b}_i, \hat{e}_j)} \tag{7.6}$$

Quantities with a hat ("$\hat{b}_i$") are simultaneous maximum likelihood estimators. Technically this boils down to defining a figure of merit for each parameter point and picking from highest to lowest until the correct coverage is reached. One could interpret the ratio as a measure for how well a particular parameter point does in comparison to the optimal fit value. The result has some very desirable properties. It guarantees a correct coverage by construction and avoids empty intervals in corner cases. For the analysis at hand, there are two weaknesses: Doing the necessary Monte Carlo integration for each parameter point is computationally very expensive. Also, one is not interested in the b_i and e_j and would really just want to have information about g. The first problem can be overcome using Wilks' theorem [95], the second using an extension usually called profile likelihood method.

Following [92, 33] analogously, the likelihood ratio is modified to also fit the nuisance parameters in the nominator. Quantities with a single hat are defined as before, the double hat ("$\hat{\hat{b}}_i$") denotes maximum likelihood estimators under the condition that one parameter (g) is fixed.

$$\lambda(n, m_i; g) = \frac{L(n, m_i; g, \hat{\hat{b}}_i, \hat{\hat{e}}_j)}{L(n, m_i; \hat{g}, \hat{b}_i, \hat{e}_j)} \tag{7.7}$$

Assume the sampling distribution of λ, in other words its PDF, is known for a particular g. Then one can immediately tell from its value, if g is included in the confidence interval, by identifying the integral of the sampling distribution with a p-value:

$$p = \int_0^{\lambda^{\mathrm{obs}}} \mathrm{PDF}(\lambda; g) d\lambda \tag{7.8}$$

By construction, if $p \leq$1-CL, it is less likely to have observed λ^{obs} than 1-CL, which corresponds to the correct coverage. A priory it is not obvious that the sampling distribution does not depend on the nuisance parameters. Define, however, the variable q,

$$q(n, m_i; g) \equiv -2 \ln \lambda(n, m_i; g). \tag{7.9}$$

Given a set of regularity conditions, Wilks' theorem states that in the asymptotic limit, q is distributed according to a χ^2 distribution for the correct value of g. The number of degrees of freedom equals the number of parameters of interest – in this case one. Since the distribution of q is known and independent of the nuisance paremeters, it can immediately be used to determine the significance for a given n, m_i and g.

For convenience, q is transformed once more to a new variable u:

$$u(n, m_i; g) \equiv \sqrt{q(n, m_i; g)} = \sqrt{-2 \ln \lambda(n, m_i; g)} \tag{7.10}$$

Since q is distributed according to a χ^2 distribution of one degree of freedom, u is distributed according to a standard Gaussian of mean 0 and standard deviation 1. This way, values of u directly correspond to the typically specified "number of sigmas" that are quoted for some confidence levels.

Having established these tools, they can be applied to sensitivity and expected limit determination. In the former case, $u(n, m_i; g^{\text{test}} = 0)$ has to be larger than 5 for rejection of the hypothesis H_0. n is taken from the Asimov data set, i.e. the expectation value for the experiment. The m_i are available from Monte Carlo simulation. It has no physical interpretation to accept negative values of g, so these are not allowed in the fit. They correspond to downward fluctuations in the backgrounds, which are mapped to $\lambda = 0$, ensuring correct normalization of the upward tail at the same time. Otherwise, one would be sensitive to both tails of the Gauss distribution instead of just one.

For limit setting, the relevant expected significance is given by $u(n, m_i; g^{\text{test}})$ for any non-vanishing g^{test}. Now, the SM-like H_0 is assumed, which leads to a lower expected n. As for discovery, without constraint, the tested hypothesis $\mathrm{H}_1(g^{\text{test}}$ would be rejected for both tails of the Gaussian. Either real $g \gg g^{\text{test}}$ or for large statistical upward fluctuations of the yield n in the experiment correspond to the fit suggesting an estimated value of g larger than g^{test}, which is irrelevant for an upper limit. To avoid these unwanted rejections, the estimator $\hat{g}$ of g is limited to $\hat{g} \leq g$.

Both for the upper limit and the discovery sensitivity one cannot simply use the value of u to distinguish the tails, as u itself is constrained to be positive. The tails are superimposed on one another, so one-sidedness has to be introduced beforehand. The procedure for choosing n from the assumed model's expectation value make this problem very unlikely for Monte Carlo studies, but it is important to keep in mind for ensemble tests.

To implement the fitting procedures, the RooFit [96] package was used.

7.1.4 Ensemble Test of Test Statistic Distribution

For a large data set, one can usually assume the asymptotic limit in Wilks' theorem is reached. In the case of comparatively small data sets it is prudent to verify the distribution of u. The strategy for setting up such a test is to choose some reasonable true parameters and then generate pseudodata for the experiment. The coupling g, the background expectation values b_i and the true systematic parameter deviations e_j are all model parameters and thus cannot be simulated. Reasonable examples have to be picked instead.

It is important to note that all other variables, not only n have to be randomly drawn from their relevant distribution. To be completely frequentist, this includes the observed systematic deviations. Conceptually, the notion of repeating the experiment includes not only waiting for more data to be produced at the same facility. It expands to cover doing the Monte Carlo simulations of the background again and to repeat all the measurements necessary to determine the parameters whose systematic uncertainties are included in the likelihood.

The whole point of applying Wilks' theorem is, that doing these kinds of simulations is computationally expensive. An ensemble test can only cover a couple of examples or the point becomes moot. To focus on individual features, not all of the complexity has been included in the simulation from the start. First, consider Equation (7.2), an early step in the likelihood function derivation that includes statistical MC background uncertainties but no systematic uncertainties.

In Figure 7.1a the sampling distribution for 1600 pseudo experiments is shown. There is one background simulated, with $\tau = 1$ and $b = 40$, which is on the order of the values for the final cuts. By definition, $s = 0$ for determining the sampling distribution in case of a SM-like hypothesis. The blue curve is a one-tailed Gaussian distribution of mean $\mu = 0.0$ and standard deviation $\sigma = 1.0$, normalized to the number of events in the test. It clearly matches the distribution of the simulation. All events in which a downward fluctuation would have lead to a negative $\hat{s}$ are contained in the bin at zero. This is the other tail of the Gaussian which is mapped to zero because of the constraints needed for limit setting introduced above. In this simplified model they are applied to the fit parameter s which is used directly rather than $s(g)$.

For clarity, the same plot is shown again in Figure 7.2, with the first bin including the peak is ignored in the fit and not shown. This is done in all of the following plots.

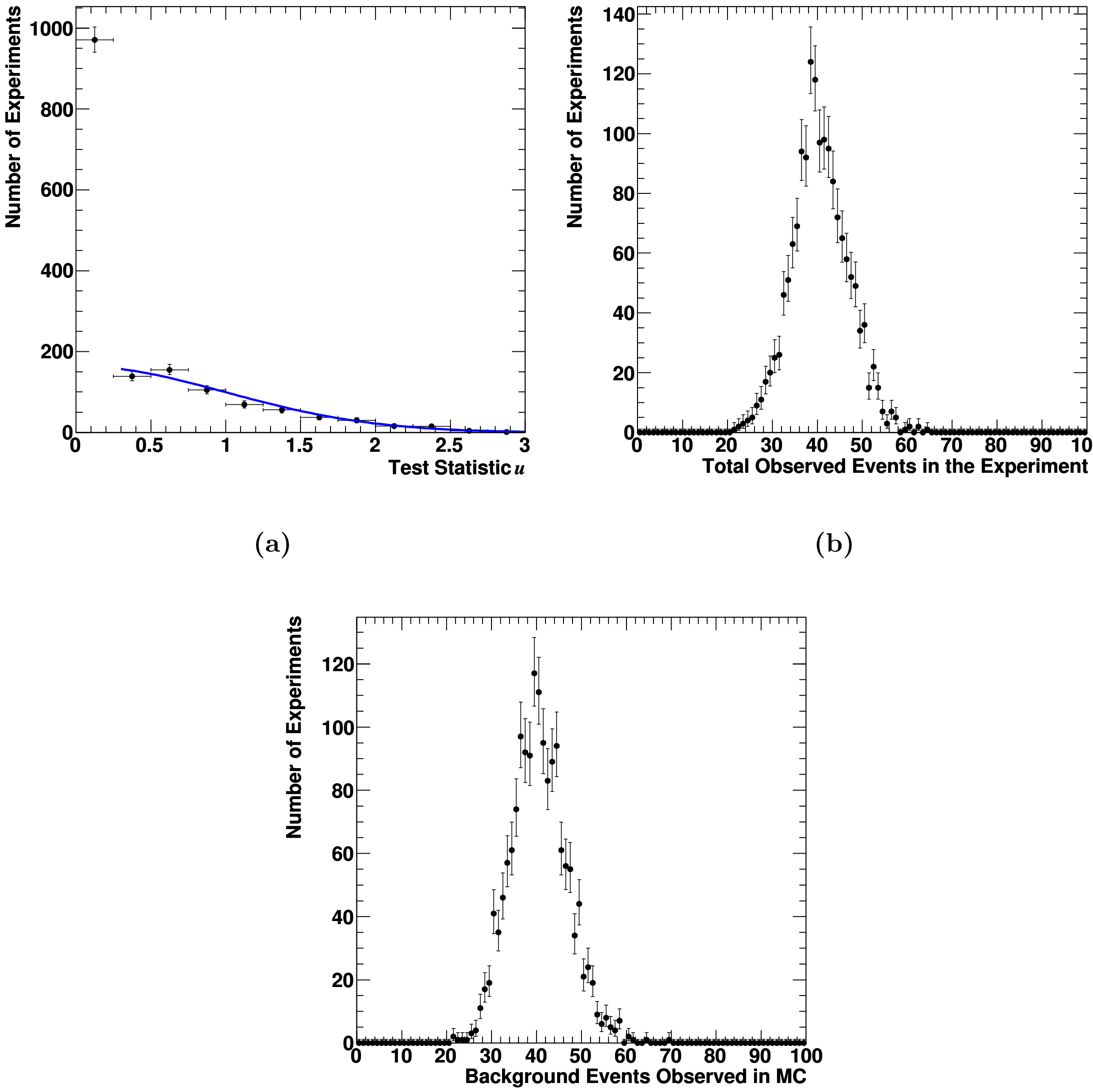
To check how strong the dependence on the exact value of b is, the simulation and fit is repeated for a factor of two in either direction (cf. Figure 7.3). A standard Gaussian ($\mu = 0.0, \sigma = 1.0$) is superimposed and still fits nicely.

If the tested model is not describing the underlying true distributions correctly, i.e. if $s > 0$, one would expect to see the Gaussian's mean move to higher values of u. An example for $s = 10$ is shown in Figure 7.4 which noticeably deviates from the superimposed standard Gaussian. From such a plot one could extract expectation bands for the sensitivity.

Normalized Gaussian distributions are customarily used to denote confidence levels. Explicitly, a confidence level is identified with the cumulative Gaussian distribution for a particular significance value Z defined as follows:

$$\mathrm{CL} = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} \tag{7.11}$$

From now on confidence level CL and significance Z will be used interchangeably,



(a) **(b)**

(c)

Figure 7.1: *Top left: Sampling distribution of the u statistic for $b = 40$ and $s = 0$. A standard Gaussian ($\mu = 0$, $\sigma = 1$) is overlaid. No systematic effects. Top right: Distribution of n drawn from $\frac{(s+b)^n e^{-(s+b)}}{n!}$. Bottom: Distribution of m drawn form $\frac{(\tau b)^m e^{-(\tau b)}}{m!}$ with $\tau = 1$.*

according to typical usage.

Discovery can only be claimed for a confidence of $Z = 5\sigma$, so it is important to also check the fit at larger values of u. Generating millions of pseudo experiments and then performing the fit will do the trick but it is wasteful. For low values of u

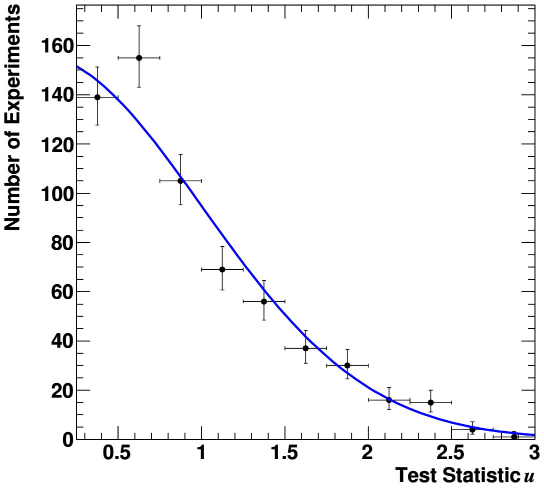
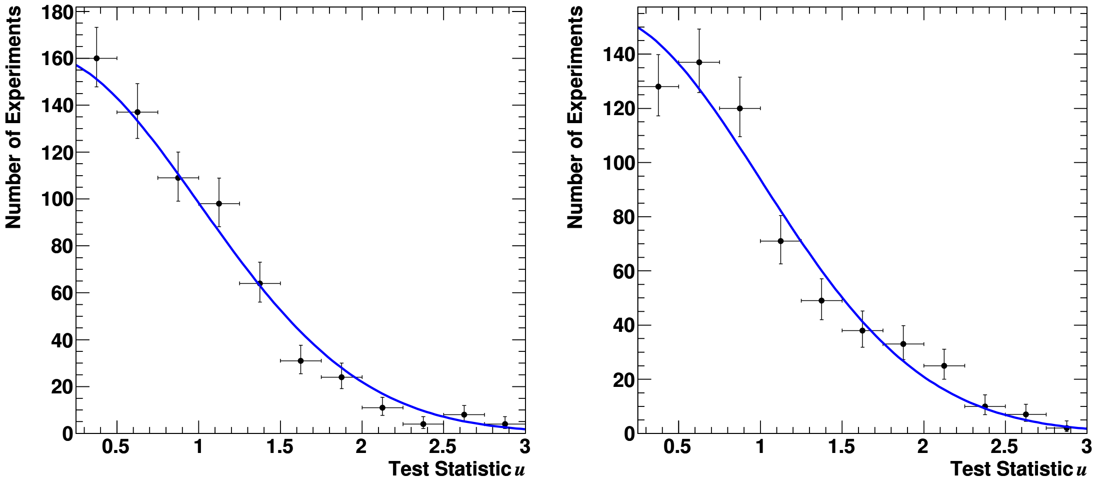


Figure 7.2: *Sampling distribution for u: $b = 40$, $s = 0$. A standard Gaussian is overlaid which matches the data points. In order to cut out the large peak at $u \approx 0$, the horizontal axis range starts at 0.25.*



(a) $b = 20$

(b) $b = 80$

Figure 7.3: *Sampling distribution of u for varied background expectation values. $s = 0$. A standard Gaussian still describes the pseudo data points, despite the change in underlying true values. Peak at $u \approx 0$ removed, horizontal axis range starts at 0.25.*

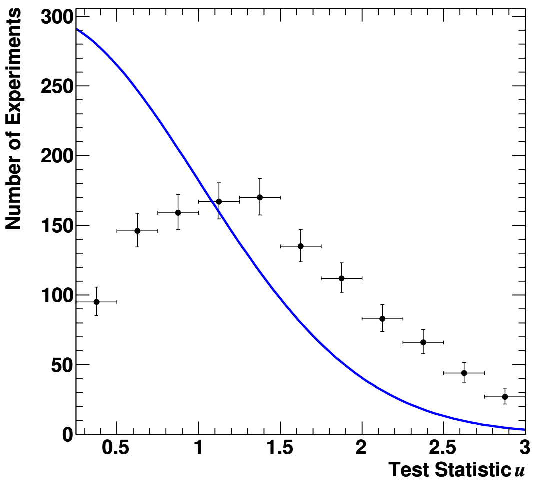


Figure 7.4: *Sampling distribution of* $u(s^{\text{test}} = 0)$ *with* $b = 40$ *and* $s = 10$. *The overlaid standard Gaussian does not match the pseudo data anymore. A test would frequently reject the hypothesis* $s^{\text{test}} = 0$ *as it should. The (smaller) peak at* $u \approx 0$ *is not shown.*

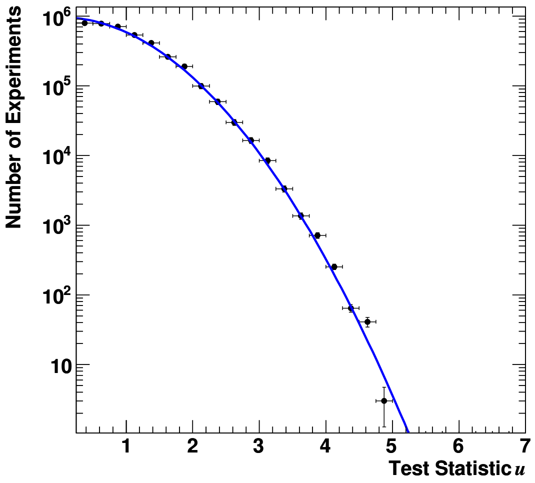


Figure 7.5: *Fit to high values of* u. *Vertical axis drawn logarithmically to emphasize small values. In blue a standard Gaussian is overlaid which matches the pseudo data over the simulated range. Peak at* $u \approx 0$ *not shown.*

the uncertainty will be more than adequate, long before it has reached acceptable levels in the high u region.

For this analysis an importance sampling scheme was developed to make this more efficient. The fit is the numerically expensive bit in each experiment. Generating suitable values for n and m_i is cheap by comparison. Instead of calculating u for every point, one would want to calculate it relatively more often for high values than for low ones to obtain small uncertainties across the board. At first sight it seems that one might need to know the result to get at the result more quickly. If one knew u to begin with, one would not need to calculate it any more. However, while a simple "$Z_{\text{simple}} = s/\sqrt{b}$" criterion is not exact enough for being quoted as a result, it is correlated with the significance Z and hence with u. For the high statistics fit, it is used to decide, whether or not a particular set of n and m_i is interesting enough to have its corresponding u calculated.

More formally, a heuristic, monotonously increasing function $p(Z_{\text{simple}}) \in [0, 1]$ was defined to give the probability to retain a particular pseudo experiments. In order to make the distribution more uniform, the PDF of the surrogate statistic Z_{simple} was divided out.

$$G(Z_{\text{simple}}) \equiv \frac{1}{\sqrt{2\pi}} e^{\frac{Z^2_{\text{simple}}}{2}} \tag{7.12}$$

$$p(Z_{\text{simple}}) \equiv \begin{cases} \frac{G(4)}{G(Z_{\text{simple}})} & \text{for } Z_{\text{simple}} \leq 4 \\ 1 & \text{for } Z_{\text{simple}} > 4 \end{cases} \tag{7.13}$$

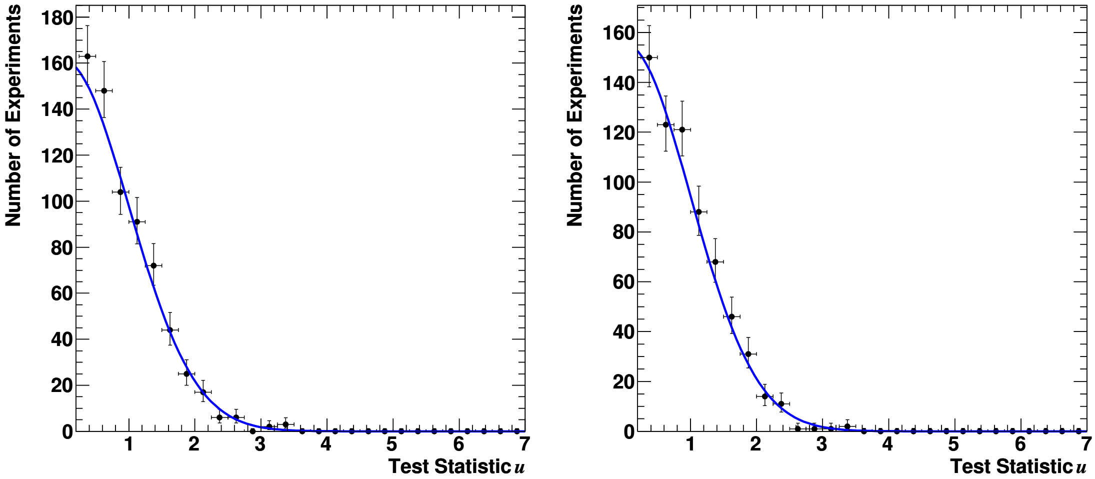
The values of u that were kept entered the histogram with a weight of $w = \frac{1}{p}$ to compensate. The results are shown in Figure 7.5 on a logarithmic scale. Note that the exact nature of the function p is arbitrary as far as the result is concerned, it only has a bearing on the amount of computation necessary to achieve reasonably low statistical uncertainties. As the superimposed Gaussian shows, the simulation conforms to expectations up to large u.

Adding Gaussian systematic errors to the likelihood in these tests, one returns to the more complete definition used for sensitivity determination (Equation (7.5)). Figures 7.6 and 7.7 show the distribution of u for different values of one uncertainty. As these approach the order of the backgrounds expecations vanishing and negative yields become an issue. Here, as in the evaluation of sensitivities, negative expectations are truncated[1]. For large uncertainties, the method leads to some over coverage and is therefore conservative. This can be seen in Figure 7.7b where the probability (p value) to generate a value greater than about 1.0 is always lower than predicted by the overlaid Gaussian of standard deviation 1.0: The blue curve overestimates the data points in the high-Z region.

As a final step, also the true systematic errors e_{obs} are given non-vanishing values in the simulation. Results can be seen in Figure 7.8. For negative values there is considerable overcoverage due to the zero bound. This leads to a potentially

---

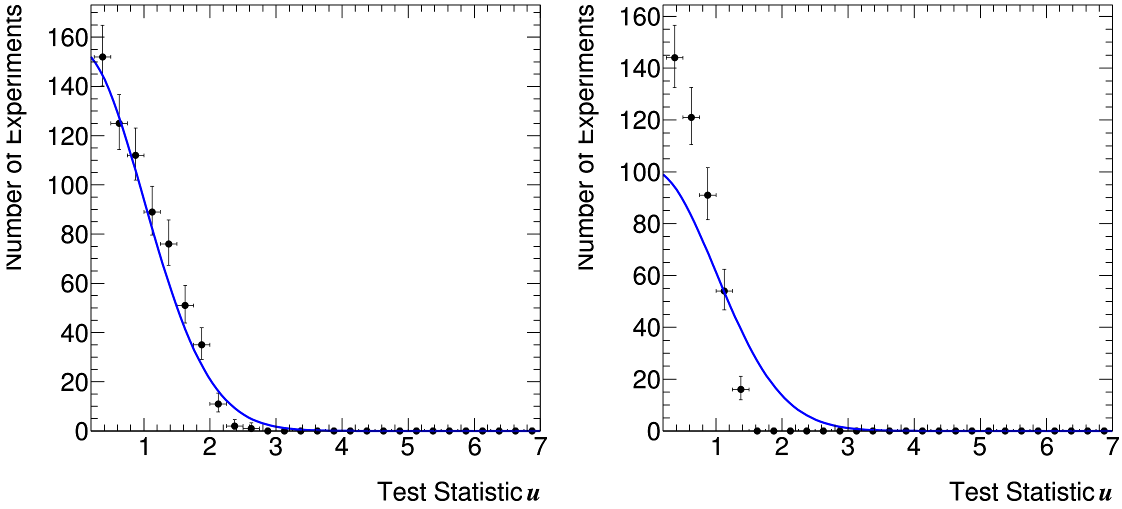
[1] One could alternatively use a $\Gamma(x; k, \vartheta)$ distribution which naturally only defined on $[0, \infty)$.



(a) $c_{00} = 4$

(b) $c_{00} = 8$

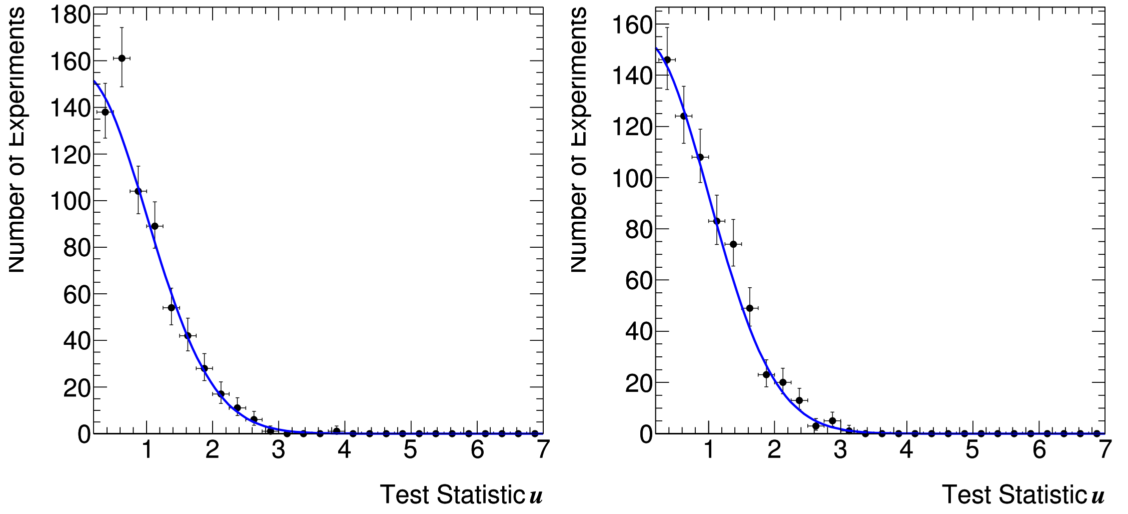
Figure 7.6: *Sampling distribution of* u *for several systematic background uncertainty magnitudes* c_{00}*.* $b = 40$, $s = 0$, $\sigma_{\mathrm{unc}} = 1$. *Overlaid are standard Gaussians which match the pseudo data for values of* $c_{00} \leq 16$. *Peaks at* $u \approx 0$ *removed, horizontal axis starting at* 0.2. *(Figure continues)*



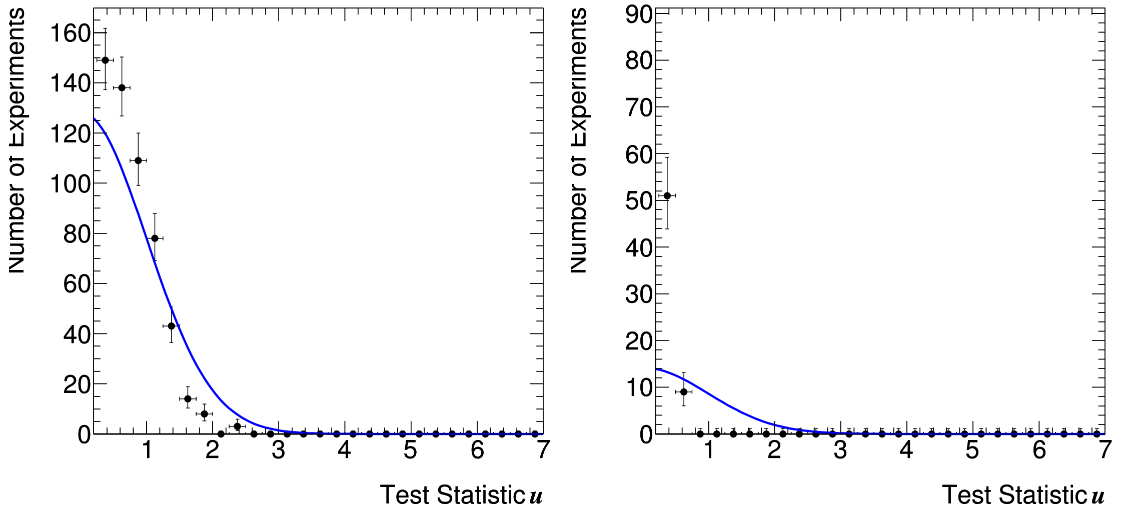
(a) $c_{00} = 16$

(b) $c_{00} = 32$

Figure 7.7: *Continued from Figure 7.6. For* $c_{00} = 32$*, the distributions differ, such that* $p^{\mathrm{exp}}(u) < p^{\mathrm{Gauss}}(u)$ *when* $u > 1.2$*.*



(a) $\Delta\sigma = 3.0$ **(b)** $\Delta\sigma = -3.0$



(c) $\Delta\sigma = 5.0$ **(d)** $\Delta\sigma = -4.5$

Figure 7.8: *Sampling distribution of* u *for several systematic background uncertainty offsets* $\Delta\sigma$. $b = 40.0$, $s = 0.0$, $\sigma_{\mathrm{unc}} = 1.0$, $c = 8.0$. $\Delta\sigma = -5.0$ *is skipped, because it would have a vanishing true yield. Overlaid standard Gaussians fit the pseudo data, except for the largest offsets. There,* $p^{\mathrm{exp}}(u) < p^{\mathrm{Gauss}}(u)$ *for large enough* u. *Peaks at* $u \approx 0$ *removed, horizontal axis starts at* 0.2.

conservative method. In the other direction, there is no noticeable undercoverage. This would have been surprising given earlier results, since the offset only changes the effective value of b.

Values of $s^{\text{test}} \neq 0$ have not been covered explicitly. Since the Monte Carlo estimation uncertainties of s are Gaussian, they are believed to be covered by the simulations of systematic errors.

Large overcoverage is confined to certain areas of parameter space. While still undesirable, it is at least clear if a particular configuration is problematic or not.

7.2 Systematic Uncertainties

7.2.1 Overview

Systematic uncertainties have been accounted for in the likelihood of section 7.1. They have not been properly introduced, though. Uncertainties which are not directly dependent on counting statistics are termed systematic. These include experimental effects such as uncertainties on calibration constants or reconstruction efficiencies but also theoretical issues like sample normalization uncertainties. All these have in common, that they do not diminish automatically with accumulation of events. It immediately follows, that their importance relative to statistical uncertainties, which do follow counting statistics, will change over time.

There is some blurring of boundaries. Monte Carlo statistical uncertainties for example do follow counting statistics but may become systematic uncertainties in the context of a measurement. These are explicitly taken into account by terms in Equation (7.5) and not considered here any more. Similarly, knowledge of calibration constants typically improves over the course of an experiment and may depend on the amount of data available in some channel. Still, these effects are considered to be systematic.

Modeling Absolute Uncertainties Modeling these errors has some degrees of freedom and can usually not be done without some arbitrary methodical choices. Here, a strictly linear Gaussian approach is used. Each sample yield y, which may be either signal s^{MC} or any background $\frac{m_i}{\tau_i}$, is assumed to depend linearly on the uncertain parameter. y refers to the yield scaled to $\mathscr{L} = 100\,\text{fb}^{-1}$ and is thus an estimator of $s'(g)$ or one of the b'_i respectively.

$$y(e_j) = y(0) + ce_j, \tag{7.14}$$

For each background i and parameter j the proportionality constant c corresponds to one of c_{ij} and c^{s}_j as introduced in Equation (7.4).

The uncertainties themselves were taken from published sources or experimental performance group results. To determine the scale factors c_{ij} and c^{s}_j, the complete analysis chain was run repeatedly for modified values of one error parameter e_j at a time for integer multiples of its uncertainty σ_j. From the ensuing $y(e_j)$ plots, the slope was extracted and used as estimate of the corresponding error scale factor c.

Figure 7.9 contains plots for the muon energy scale which will be introduced in more detail in the next section. On the horizontal axis, the deviation e from the energy scale implemented in the Monte Carlo simulation is given in units of its standard deviation σ. The parameters e_j had already been introduced in equation (7.4). When not used in the context of a formula with multiple error parameters, the subscript is dropped in the notation. The left plot shows the total yields $y_i(e)$ for each sample i. On the right side, the change from the central value $\Delta y_i(e) = y_i(e) - y_i(0)$ is shown. Both are prepared for a BDT cut of > 0.0. A roughly linear relationship between yield y and error parameter e can be seen. Both the magnitude and the sign of the slope differs for different samples. For a relatively weak BDT cut of ≥ 0.0, there are some events left for all samples of interest, including some $t\bar{t}$ and very few Wt events. In this cut region, the systematic uncertainties play a large role. For stricter cuts as will be shown in the next sections, the statistical uncertainty on the Monte Carlo is more important for the few remaining events.

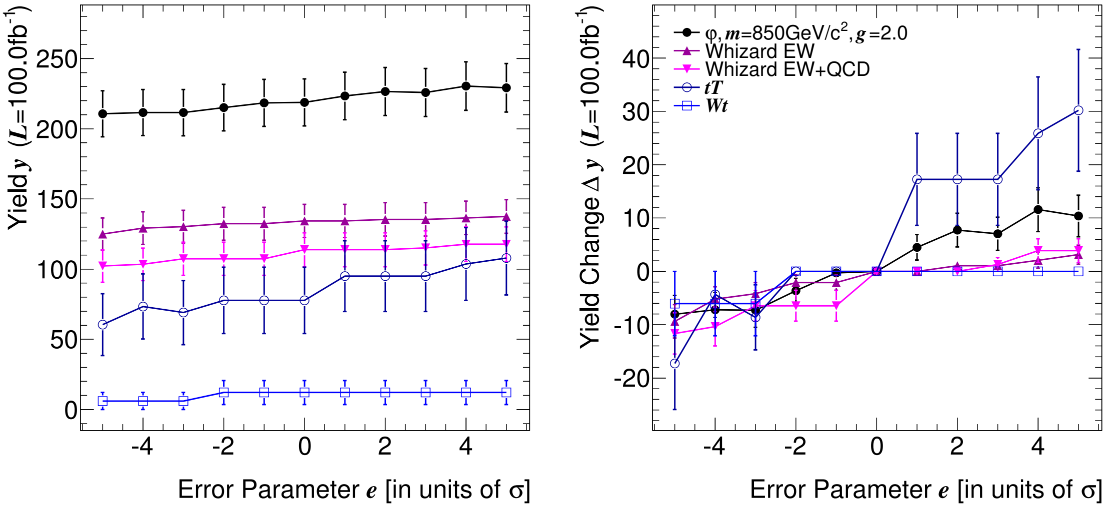


Figure 7.9: *Left: Expected event yield y. Right: Change in event yield Δy after cuts. Both versus error parameter e for muon energy scale. Values are shown for a BDT cut of > 0.0, and an integrated luminosity of $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. The signal sample φ is drawn with the SM VBS component subtracted. More details on the error bars can be found in the text.*

Due to the finite Monte Carlo statistics, however, there are again uncertainties on the yields. The common advice here is, to produce more Monte Carlo events until the error is negligible [97]. For the LHC this is not usually possible as is the case here. Instead, the uncertainties need to be estimated. For the plots of $y(e)$, the uncertainty on each point individually is given by Poisson statistics. Error bars are given by the Gaussian approximation to Poisson statistics appropriately scaled to $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. In the general case of events with weights w_i and a Monte Carlo simulation to experimental luminosity ratio $\tau = \frac{L_{\mathrm{MC}}}{L_{\mathrm{exp}}}$ this results in:

$$\sigma(y(e)) \approx \frac{1}{\tau}\sqrt{\sum w_i^2} \tag{7.15}$$

Since the individual data points are highly correlated, these uncertainties do not describe the uncertainty of any linear fit. Most of the correlation can be removed, by looking at the uncertainty of the yield change $\Delta y(e)$ plotted on the right side instead. In a simple cut analysis, scaling a variable will result in some values migrating into the acceptance region or some migrating out, but never both at the same time. Without loss of generality, let $\Delta y(e)$ be positive, i.e. some additional events have been accepted. Since all of the other events are shared, the variance of the number of newcomer events is the variance of the difference. An estimator for this is the square root of the sum of their weights squared.

$$\sigma(\Delta y(e)) \approx \frac{1}{\tau}\sqrt{\left|\sum w_i^2(0) - w_i^2(e)\right|} \tag{7.16}$$

Due to the absolute value being taken this holds true also for negative Δy. In the case of the used BDT, event migration may happen in both directions simultaneously, therefore the uncertainty may be larger. As approximation error bars are drawn according to the simpler case just derived.

Either way, to actually take these uncertainties into account, they would have to be subtracted from the systematic uncertainties [97] This is usually not considered to be acceptable, therefore the only choice left is to ignore the statistical uncertainties, staying on the conservative side.

The quoted scaling constants c_{ij} and c_j^{s} were determined by choosing the greatest deviations in either direction, typically $\pm 5\sigma$ and calculating the yield slopes $c^{\pm} = \Delta y^{\pm}/\pm 5\sigma$. The one with the greater magnitude is chosen, keeping its sign. This biases the result to higher slopes, but is able to give a conservative answer in the case where the signs on both sides differ or the slopes are clearly different beyond what is suggested by statistical uncertainty on opposite sides. Before taking the slope of the signal samples including irreducible background, it was transformed to correspond the signal definition in Section 5.4.

Alternative Approach to Absolute Uncertainties While not used in this thesis, a possible different approach that has been conceived in this context is the following. Instead of estimating the yield vs. error parameter relationship $y(e)$ with a linear model the estimates of $y(e)$ could be used directly. Each simulated point already corresponds to a simulation of m_i or s^{MC} at an alternative error parameter point. Assuming as above that the yield changes Δy for different error parameters add up, one may write

$$y(e_0, e_1, \ldots, e_j, \ldots) = y(0) + \Delta y(e_0) + \Delta y(e_1) + \ldots \tag{7.17}$$

$y(e_j)$ is understood to mean $y(0, 0, \dots, e_j, \dots)$. Transforming back to the raw Monte Carlo estimates without luminosity scale factors $\tau_i = m_i/y_i$ and τ^{s} analogously

$$\begin{aligned} m'_i &= m_i(0) + \Delta m_i(e_0) + \Delta m_i(e_1) + \dots \\ s^{\mathrm{MC}\prime} &= s^{\mathrm{MC}}(0) + \Delta s^{\mathrm{MC}}(e_0) + \Delta s^{\mathrm{MC}}(e_1) + \dots \end{aligned} \tag{7.18}$$

which replaces (7.4). The whole likelihood then looks almost as before in equation (7.5) except that now m_i and $s^{\mathrm{MC}}(g)$ are modified, rather than b'_i and $s(g)$.

$$\begin{aligned} L(n, m_i; g, b_i, e_j) = {} & \frac{(s(g)+b)^n e^{-(s(g)+b)}}{n!} \Pi_i \frac{(\tau_i b_i)^{m'_i} e^{-(\tau_i b_i)}}{m'_i!} \\ & \times \frac{1}{\sigma^{\mathrm{s}}(g)\sqrt{2\pi}} e^{-\frac{(s^{\mathrm{MC}\prime}(g)-s(g))^2}{2\sigma^{\mathrm{s}}(g)^2}} \\ & \times \Pi_j \frac{1}{\sigma_j\sqrt{2\pi}} e^{-\frac{(e_j^{\mathrm{obs}}-e_j)^2}{2\sigma_j^2}} \end{aligned} \tag{7.19}$$

One advantage of this is, that there is no need to determine the slope parameters c. Another advantage is that the statistical uncertainties of the simulations of m'_i and $s^{\mathrm{MC}\prime}(g)$ can be treated just as in the absence of systematic effects. They are already included in the likelihood from the start.

Currently, however, the more traditional approach outlined above is used.

Relative and Absolute Uncertainties Some uncertainties are not treated on an event by event basis. These can be treated analytically or have been estimated from another large sample of Monte Carlo events. The result is a relative uncertainty that is mostly independent of additional cuts, e.g. the luminosity uncertainty. Conversion into the coefficients c_{ij} is trivial and in particular not dependent on the BDT cut or the value of g. For the uncertainties estimated from rerunning the chain with changed error parameters, this was found not to be the case. At least, the relative uncertainty could not be easily be estimated due to the low number of events left for interesting cuts. Instead, the absolute deviation is simulated and given. To distinguish between the two cases, the former are called relative uncertainties, the latter absolute uncertainties. There is no very fundamental difference between the two, their only treated different in practice.

7.2.2 Experimental Uncertainties

This thesis' treatment of systematic uncertainties follows closely the recommendations for the ATLAS paper on expected performance [33].

At high particle energies $E >> m$, energy and momentum are approximately the same values in natural units. Conventionally the terms are often used interchangeably despite the fact that the quantities are measured by different subdetectors. Unless otherwise noted, any scaling is applied to both energy E and momentum p as well.

The transversal quantities are changed accordingly. For jets, the scale is kept the same for E and p. For leptons, p is changed with E such that the mass is conserved instead.

Luminosity At the time this measurement can be made, luminosity is expected to be known at around the 3% level. This makes luminosity a minor effect. One can immediately give its effect on yields, since they are all proportional. There is no need to redo the analysis chain. Correlation between samples is 100%. 3% uncertainty is assumed.

Muon Energy Scale Customarily the energy scale is given, although muons are measured via their momentum. Since they are minimally ionizing, high-p_T muons traverse both calorimeters leaving only some of their energy. This deposit is only weakly depending on their total energy. Inner detector and muon system tracks are used for reconstruction and measurement of p_T.

Highly energetic muons are an important ingredient for event selection. Their calibration has an immediate effect on yields. In fact, this is one of the most important effects. A deviation will affect all muon momenta in the same way, as they are all measured against the same deviating scale. Despite the fact that the measurement is of p_T, conventionally the parametrization is done in energy. For some error parameter e it follows for all muons:

$$E'_\mu = (1 + e)E_\mu \tag{7.20}$$

An uncertainty of $\sigma = 1\%$ of the scale is expected.

Muon Reconstruction Resolution Beyond the total scale deviation, there is also some uncertainty associated with the reconstruction resolution. It is given by the following smearing function in terms of reverse p_T.

$$\sigma\left(\frac{1}{p_\mathrm{T}}\right) = \frac{0.011}{p_\mathrm{T}} \oplus \frac{0.00017}{\mathrm{GeV}/c} \tag{7.21}$$

The first term enhances the coulomb smearing. The second term takes into account alignment uncertainties affecting TeV-scale muons. In contrast to a scale variation this smearing needs to be applied on an event by event basis, with a new p_T deviation randomly chosen each time from a Gaussian of standard deviation as given above. The resolution expectation from Monte Carlo simulation is 3-5% of $1/p_\mathrm{T}$ [51], depending on η and φ. 3-5σ deviation would bring the added effect on par to the one already in Monte Carlo (neglecting the constant term which dominates at high p_T.)

The corresponding error parameter e scales the constants in σ, where e has itself a standard deviation of $\sigma_e = 1$.

$$\sigma\left(\frac{1}{p_\mathrm{T}}\right) = e\left(\frac{0.011}{p_\mathrm{T}} \oplus \frac{0.00017}{\mathrm{GeV}/c}\right) \tag{7.22}$$

Per run of the analysis chain, e is chosen once, but the change in $\frac{1}{p_\mathrm{T}}$ for the event number k, $\Delta\frac{1}{p_\mathrm{T}}$, is chosen each time.

$$E'_\mu = \left(\frac{1}{1 + p_\mathrm{T}\Delta\frac{1}{p_\mathrm{T}}}\right) E_\mu \tag{7.23}$$

Electron and Muon Reconstruction Efficiency In addition to uncertainties on the parameters of reconstructed objects, the raw efficiency of reconstruction itself is uncertain. For muons, the relative uncertainty is taken to be 1%, for electrons 0.2%. Simulating a loss of objects is straight forward. One can run the analysis again with random objects removed and record new yields. It is more problematic to simulate increases in efficiency. New objects are needed, but are not available. An alternative is to make an analytical approximation.

For each event, two leptons are demanded by the cuts. Any additional ones are accepted but ignored. Making the assumption that the relevant samples are dominated by events containing two leptons, the change in efficiency can be estimated by reweighting each event. The weight is scaled with the relative change in efficiency. At the 1σ-level:

$$w_i^\ell = (1 \pm \Delta\varepsilon^e)^{n_i^e} \, (1 \pm \Delta\varepsilon^\mu)^{n_i^\mu} \tag{7.24}$$

In this context, the number of electrons n_i^e and the number of muons n_i^μ per event are assumed to add up to two. The total relative change $\Delta\varepsilon^\mathrm{total}$ can now be estimated from the average. In the general case of events that are already weighted with weights w_i^ev this becomes an event-weighted average of the reconstruction efficiency weights.

$$\Delta\varepsilon^\mathrm{total} = \frac{\sum_i w_i^\ell w_i^\mathrm{ev}}{\sum_i w_i^\mathrm{ev}} - 1 \tag{7.25}$$

For all samples and tested cuts, the uncertainty comes out to be $\leq 1.4\%$.

This ignores the possibility that an event with more than two leptons of which one is rejected to reduce efficiency may still pass the cuts because of an additional lepton. Similarly for increased efficiency. This systematic uncertainty is already quite small compared to other effects, however, so this additional correction can be ignored safely.

Electron Energy Scale The electron energy scale uncertainty is similarly important for the sensitivity determination as the muon energy scale. While electron p_T can be measured in the inner tracker, a much better measurement is available via the electromagnetic calorimeter where electrons loose the bulk of their energy. The scale uncertainty is expected to be 0.5%. It is simulated just as the muon energy scale by choosing one e for multiple runs of the analysis chain each. Individual events are modified as in (7.20).

$$E'_e = (1 + e)E_e \tag{7.26}$$

Electron Energy Resolution Similar to the behavior of muons, electron resolution is not known exactly. In contrast to the behavior of a track based measurement in the case of muons, the energy resolution of the calorimeter does not deteriorate for high-energy electrons. The 1σ-level uncertainty is therefore only simulated with an enhancement of the constant term.

$$\sigma\left(E_{\mathrm{T}}\right) = e * 0.0073 E_{\mathrm{T}} \tag{7.27}$$

e is picked once per analysis run. The shift ΔE_{T} is drawn from a Gaussian distribution of $\mu = 0$ and σ as defined above for each electron.

$$E_e' = \frac{E_{\mathrm{T}} + \Delta E_{\mathrm{T}}}{E_{\mathrm{T}}} E_e \tag{7.28}$$

Expected resolutions from Monte Carlo are between 3-1% of E_{T} for $E_e > 30\,\mathrm{GeV}$ before the enhancement due to uncertainty [51].

Jet Energy Scale The jet energy scale is known less well than the lepton energy scales. Both the inherent theoretical uncertainties in their definition and experimental uncertainties of the reconstruction algorithms play a role here. Out of cone losses are a typical effect. The ATLAS calorimeter is not compensating, leading to different hadronic and electromagnet scales. Jets have to be recalibrated from the electromagnetic to the hadronic scale.

While leptons can reliably be identified only if they leave a track in the inner detector, jets are detected in large η regions as well. Since detector properties change in the forward direction, the uncertainty is parametrized in two distinct regions. One is the central region where $|\eta| \leq 3.5$ and the other is the forward region where $|\eta| > 3.5$. It is expected that the scale uncertainty is 3.5% in the former and 7.5% in the latter region. These values were halved from the ones given in the reference as they are expected to become much better for higher integrated luminosities. As above, a chosen value of e scales the energies of all jets in one analysis run with the same factor.

$$E_{\mathrm{jet}}' = (1 + e) E_{\mathrm{jet}} \tag{7.29}$$

Jet Energy Resolution To simulate uncertainty of the resolutions, the stochastic term is enhanced in the simulation. For the inner part of the detector, $|\eta| \leq 3.5$:

$$\sigma\left(E\right) = 0.45 \sqrt{E \times 1\,\mathrm{GeV}} \tag{7.30}$$

In the outer region, $|\eta| > 3.5$, a larger uncertainty is assumed:

$$\sigma\left(E\right) = 0.63 \sqrt{E \times 1\,\mathrm{GeV}} \tag{7.31}$$

e is fixed for the analysis run, ΔE is drawn from a Gaussian defined by σ per jet individually.

$$E' = E + \Delta E \tag{7.32}$$

The expected stochastic term in Monte Carlo simulation for jets with $2.4 < |\eta| < 2.6$ approximately 1.8 before adding the additional uncertainty [51].

Missing Transverse Momentum Systematic uncertainty in $E_{\mathrm{T}}^{\mathrm{miss}}$ is closely related to the measurement of physics objects like leptons. While it is based on energy deposits in the calorimeter, which need not be combined into jets before the measurement, uncertainties affecting the latter will also affect $E_{\mathrm{T}}^{\mathrm{miss}}$. Similarly, corrections from muon measurements will be off, when these are mismeasured. Instead of defining the uncertainty individually for $E_{\mathrm{T}}^{\mathrm{miss}}$, deviations in the energy scales of object measurements are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. This procedure models correlations between objects and $E_{\mathrm{T}}^{\mathrm{miss}}$ better than an individual new smearing. For both electrons and muons, the deviation is propagated fully, for jets 5% is taken into account. Taus and photons would have to be processed like jets and electrons respectively. Since they are not in the signal final state and thus quite suppressed, their effect is ignored. With ΔE_e for the electrons, Δp_μ for the muons and ΔE_{jet} for the jets:

$$E_{\mathrm{T}}^{\mathrm{miss}\prime} = E_{\mathrm{T}}^{\mathrm{miss}} + \sum_{\text{electrons}} \Delta E_e + \sum_{\text{muons}} \Delta p_\mu \times c + 0.05 \sum_{\text{jets}} \Delta E_{\mathrm{jet}} \tag{7.33}$$

In this vector equation only the x- and y-components are evaluated. Note that for the muons the change in momentum is used instead of energy. The speed of light c enters only to conserve units.

***b*-tag Efficiency and Rejection** Since two important background processes produce bottom quarks via top quarks, i.e. $t\bar{t}$ and single top (Wt), understanding the uncertainty in the b-tag variable rejecting them is important. A current, somewhat conservative estimate places the efficiency uncertainty at 5%, and the uncertainty of the rejection at 10%. These values are dependent on the exact working point of a single cut. In this analysis the boosted decision tree is using multiple cuts on the variable. To simulate the effect, a reasonable transformation of the input variables has to be performed. To be conservative, a large relative value of 10% was assumed for the 1σ deviation. With an error parameter e chosen for each analysis run, each bottom tag weight b_i of each jet is changed.

$$b_i' = (1 + e) b_i \tag{7.34}$$

7.2.3 Theoretical Uncertainties

In addition to experimental unknowns, the theoretical predictions are not completely exact. Some important parameters like the PDFs are really experimental uncertainties albeit from other experiments. The distinction is not completely clear. Both types of uncertainties are handled the same way in the setup of the likelihood. Some of the uncertainties can be expected to become smaller over the course of the experiment, as measurements of the background processes are being made.

Statistical Fluctuations in Monte Carlo Simulations These have already been included in the very first step of setting up the likelihood. They are mentioned here, because they really are uncertainties on the predicted cross sections, making them systematic uncertainties of sort. Their special status stems from the fact that they can be more accurately simulated in the manner given above than the other systematic effects.

Parton Density Function Uncertainties PDFs are only known to a certain precision from experiment. One could estimate the effect of changes by reweighting the Monte Carlo simulations. This is not done here.

Normalization Scale Uncertainty Generation of samples is performed using perturbation theory. Intrinsically, these calculations need to be renormalized, which introduces an additional scale. At infinite order in perturbation theory, results should not depend anymore on this scale, rendering this somewhat unphysical parameter inconsequential. For finite order, the uncertainty on the result is estimated by varying the scale around some reasonable value, typically a scale in the process.

For the reducible backgrounds, numbers are given in the literature (cf. Equation (5.2).) Since typically asymmetric errors are given, but are not simulated in the likelihood, the larger value was used. By the time this analysis is performed on data, it is believed that a well measured value of the $t\bar{t}$ and Wt cross sections will become available. At this point the uncertainty will be limited by luminosity uncertainty, which has already been considered. Since, on the other hand, there is unclear whether this particular bit of phase space will be completely well-known from simulation, the current literature value for the uncertainty was considered (cf. Table 7.1).

For the WHIZARD samples, the effect was estimated by varying the scale in the simpler, and consequentially less numerically intensive process $ud \to WW$. No resonance was included. The Monte Carlo samples used are generated at the Z mass taken to be $m_Z = 91.1876$. The scale was varied up to 850 GeV which resulted in a 16% reduction in cross section. This was taken to be the normalization uncertainty for all WHIZARD samples. In the long run, theoretical predictions will likely become better, but the current value is used in order to be conservative.

7.2.4 Results For Relative Uncertainties

Luminosity, sample normalization and lepton reconstruction efficiency have relative uncertainties. Their values are given in Table 7.1.

7.2.5 Results for Absolute Uncertainties

For nine systematic effects, the yield change had to be estimated from Monte Carlo simulation on an event-per-event basis as introduced in section 7.2.1. This has been done here for each choice of the BDT cut, and in the case of the samples containing resonances also for each mass point and coupling that is needed in the following sections for the determination of sensitivities and limits. Results are shown in Figures

Source of Uncertainty	Relative Value
Luminosity	3%
WHIZARD Normalization	16%
$t\bar{t}$ Normalization	6.2%
Wt Normalization	3%
Lepton Reconstruction Efficiency	1.4%

Table 7.1: *Relative uncertainties for different sources*

7.10, 7.11 and 7.12 for a BDT cut of $r \geq 0.7$. The included signal sample simulates a scalar isotensor φ at $m = 850\,\text{GeV}/c^2$. As before, on the left side the expected event yields $y(e)$ are plotted and on the right side the deviation $\Delta y(e)$ from $e = 0$. Values correspond to an integrated luminosity of $\mathscr{L} = 100\,\text{fb}^{-1}$ after all cuts.

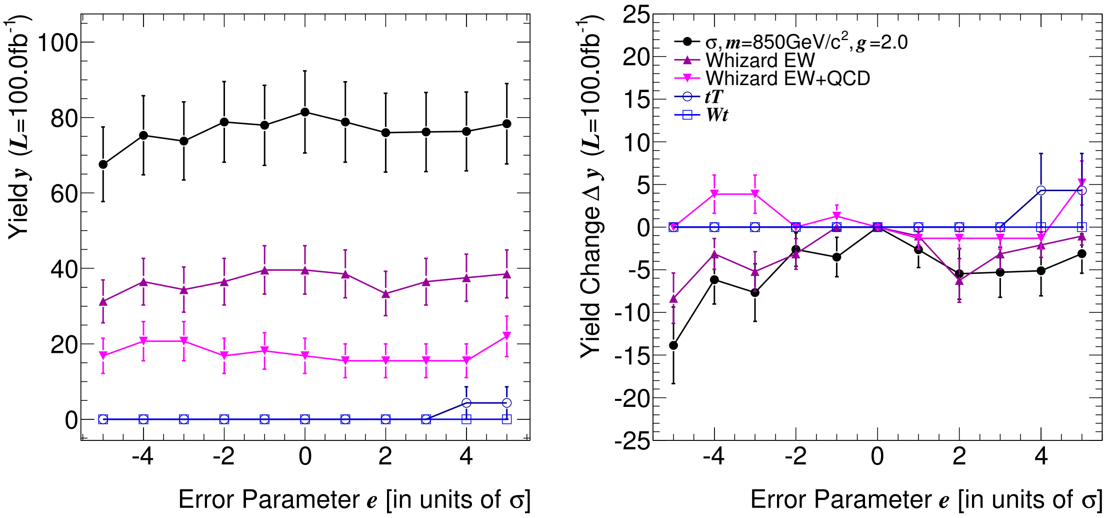
The harder, but more realistic BDT cut compared to Figure 7.9 leads to a much lower number of events. Consequently statistical fluctuations play a larger role. Extracting the slope is not in all cases straight-forward. The jet energy scale dependence looks quite linear for both signs of the deviation, but the slopes on positive and negative sides do not match up. Since always the larger absolute slope is chosen, this leads to a conservative estimate. Alternatively one might have used the slope from the maximum positive deviation. For discovery this is a good approach, as only upward fluctuations of the backgrounds matter. However, this is problematic in the case of limit setting, where one is interested in the downward fluctuations of all samples.

Table 7.2 contains the extracted slopes c_{ij} for backgrounds, Tables 7.3a and 7.3b contain the c^{s}_j for signals at $m = 850\,\text{GeV}/c^2$ and $m = 1150\,\text{GeV}/c^2$ respectively. The parameters are chosen as in the plots, the coupling of all signal resonanced is set to $g = 2.0$. Slopes are given with respect to an error parameter of $e_j = 1\sigma_j$. From the tables one can see that both the jet and lepton energy scale uncertainties play a large role. Of these objects, the electrons are best measured and their scale uncertainty has the least effect. Jet and electron resolution effects are almost negligible. The muon resolution is again an important quantity, as it is strongly deteriorated for highly boosted muons typical for this channel. Uncertainty on the flavor weights is among the group of important effects.

For $t\bar{t}$ and Wt backgrounds there are not always events left when using strong cuts. Correspondingly, the dependence on the systematic parameters cannot be extracted in all cases. However, the uncertainty due to the small Monte Carlo luminosity is explicitly included in the likelihood (7.5). Any hidden systematic uncertainty, would have to be small or even insignificant compared to the larger uncertainty from Monte Carlo statistics that is being taken into account.

Systematic Effect	WHIZARD EW	WHIZARD EW+QCD	$t\bar{t}$	Wt
Muon Energy Scale	+1.7 (4.2%)	+0.5 (3.1%)	+0.9 (—%)	+0.0 (—%)
Jet Energy Scale $\lvert\eta\rvert \leq 3.2$	+1.0 (2.6%)	+0.8 (4.6%)	+0.0 (—%)	+0.0 (—%)
Jet Energy Scale $3.2 < \lvert\eta\rvert$	+1.3 (3.2%)	+0.0 (0.0%)	+0.9 (—%)	+0.0 (—%)
Electron Energy Scale	-0.4 (1.1%)	-0.8 (4.6%)	+0.0 (—%)	+0.0 (—%)
Muon Reconstruction Resolution	-2.1 (5.3%)	-0.5 (3.1%)	+0.0 (—%)	+0.0 (—%)
Electron Energy Resolution	+0.2 (0.5%)	+0.5 (3.1%)	+0.0 (—%)	+0.0 (—%)
Jet Energy Resolution $\lvert\eta\rvert \leq 3.2$	+0.0 (0.0%)	+0.0 (0.0%)	+0.0 (—%)	+0.0 (—%)
Jet Energy Resolution $3.2 < \lvert\eta\rvert$	+0.0 (0.0%)	+0.0 (0.0%)	+0.0 (—%)	+0.0 (—%)
Flavour Weight Scale	+1.7 (4.2%)	+1.0 (6.2%)	+0.9 (—%)	+0.0 (—%)

Table 7.2: *Dependence estimators $c = c_{ij}$ of the expected yield $y(e_j) = y(0) + c_{ij}e_j$ on the error parameter e_j. Given in deviation per 1σ: $\Delta y(e_j = 1\sigma_j)$ for a BDT cut $r \geq 0.7$. Relative deviations are given in parentheses except where the expected yield $y(0)$ vanishes.*



(a) Flavor Weight Scale

Figure 7.10: *Expected event yield y on the left and change in event yield Δy after cuts on the right against error parameters e_j for multiple systematic uncertainties. Values are shown for a BDT cut of $r \geq 0.7$, and an integrated luminosity of $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. SM VBS is subtracted from the signal (cf. Section 5.4.) More details on the error bars can be found in the text.*

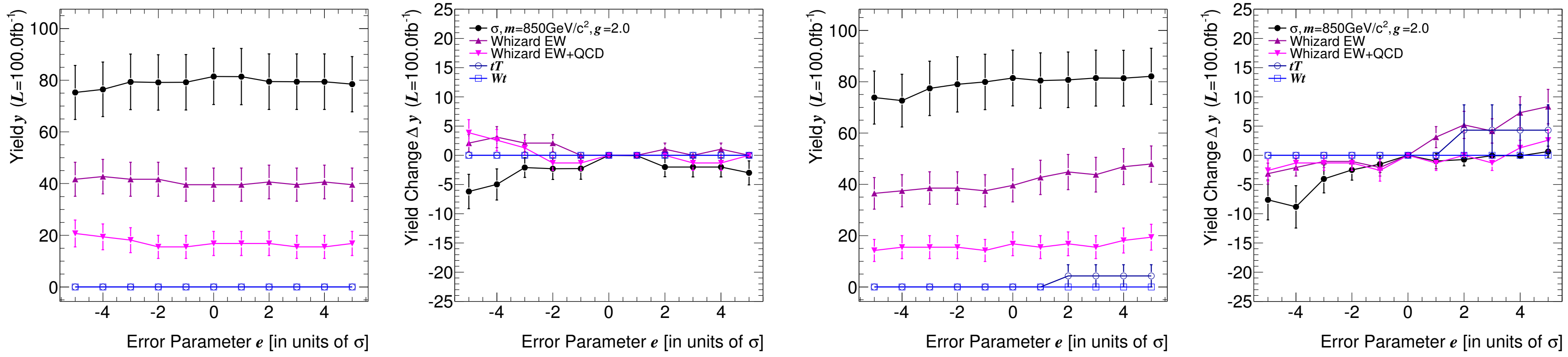
Systematic Effect	σ	φ	ρ	f	t
Muon Energy Scale	+1.5 (1.9%)	+0.9 (0.8%)	-0.0 (0.2%)	+1.0 (2.1%)	+0.2 (0.5%)
Jet Energy Scale $\lvert\eta\rvert \leq 3.2$	+3.5 (4.3%)	+2.2 (2.1%)	+0.6 (3.6%)	+0.9 (1.8%)	+0.7 (2.1%)
Jet Energy Scale $3.2 < \lvert\eta\rvert$	+3.7 (4.5%)	+7.3 (7.0%)	+0.9 (5.1%)	+2.9 (5.7%)	+1.7 (5.2%)
Electron Energy Scale	+1.2 (1.5%)	+0.7 (0.7%)	-0.0 (0.3%)	+0.2 (0.4%)	+0.4 (1.2%)
Muon Reconstruction Resolution	-3.5 (4.3%)	-6.7 (6.4%)	-1.2 (6.9%)	-1.3 (2.6%)	-1.6 (5.0%)
Electron Energy Resolution	+0.1 (0.1%)	-0.5 (0.4%)	+0.1 (0.5%)	+0.2 (0.3%)	-0.2 (0.6%)
Jet Energy Resolution $\lvert\eta\rvert \leq 3.2$	+0.0 (0.0%)	-0.2 (0.2%)	-0.0 (0.2%)	+0.0 (0.0%)	-0.1 (0.3%)
Jet Energy Resolution $3.2 < \lvert\eta\rvert$	-0.2 (0.2%)	+0.0 (0.0%)	+0.2 (1.3%)	+0.0 (0.0%)	+0.1 (0.3%)
Flavour Weight Scale	+2.8 (3.4%)	+2.6 (2.4%)	+0.5 (2.6%)	+1.5 (2.9%)	+0.5 (1.5%)

(a) Signal Samples $m = 850\,\mathrm{GeV}/c^2$

Systematic Effect	σ	φ	ρ	f	t
Muon Energy Scale	+1.4 (2.8%)	+0.8 (1.2%)	+0.2 (2.7%)	-0.5 (1.3%)	+0.2 (2.3%)
Jet Energy Scale $0.0 < \lvert\eta\rvert \leq 3.2$	+0.8 (1.7%)	+0.5 (0.6%)	-0.1 (0.9%)	+0.8 (2.3%)	+0.3 (2.8%)
Jet Energy Scale $3.2 < \lvert\eta\rvert$	+2.2 (4.4%)	+2.5 (3.6%)	+0.3 (4.0%)	+1.9 (5.1%)	+0.3 (3.1%)
Electron Energy Scale	+0.9 (1.7%)	+0.3 (0.5%)	+0.1 (1.7%)	-0.4 (1.2%)	+0.3 (2.9%)
Muon Reconstruction Resolution	-2.7 (5.3%)	-3.5 (4.9%)	-0.1 (1.0%)	-1.5 (4.2%)	-0.6 (5.9%)
Electron Energy Resolution	-0.0 (0.0%)	-0.2 (0.3%)	+0.0 (0.3%)	-0.5 (1.4%)	+0.3 (3.1%)
Jet Energy Resolution $0.0 < \lvert\eta\rvert \leq 3.2$	+0.0 (0.0%)	+0.0 (0.0%)	+0.0 (0.0%)	-0.1 (0.2%)	+0.0 (0.0%)
Jet Energy Resolution $3.2 < \lvert\eta\rvert$	-0.2 (0.4%)	+0.0 (0.0%)	+0.0 (0.0%)	+0.0 (0.0%)	+0.0 (0.0%)
Flavour Weight Scale	+1.7 (3.4%)	+1.2 (1.6%)	+0.2 (2.6%)	+0.6 (1.5%)	+0.2 (1.7%)

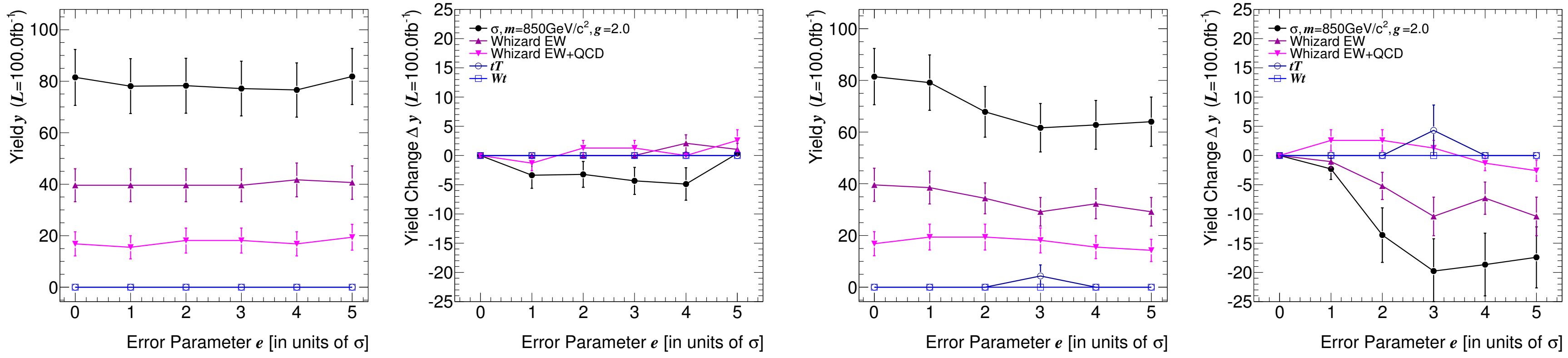
(b) Signal Samples $m = 1150\,\mathrm{GeV}/c^2$

Table 7.3: *Dependence estimators $c = c^{\mathrm{s}}_j$ of the expected yield $y(e_j) = y(0) + c^{\mathrm{s}}_j e_j$ on the error parameter e_j. Given in deviation per 1σ: $\Delta y(e_j = 1\sigma_j)$ for a BDT cut $r \geq 0.7$. Relative deviations are given in parentheses except where the expected yield $y(0)$ vanishes. The signals are all reweighted to $g = 2.0$ with SM VBS components subtracted (cf. Section 5.4.)*



(a) Electron Energy Scale

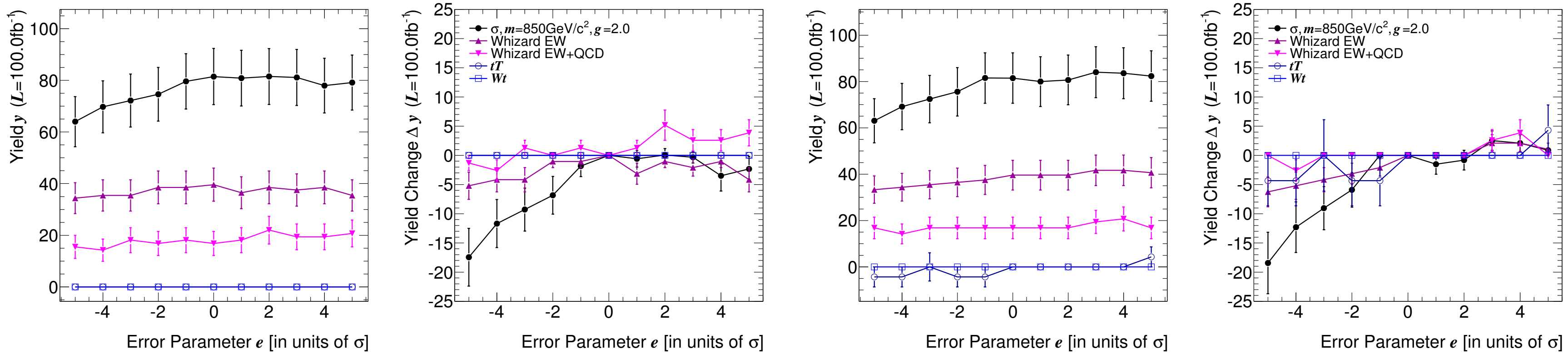
(b) Muon Energy Scale



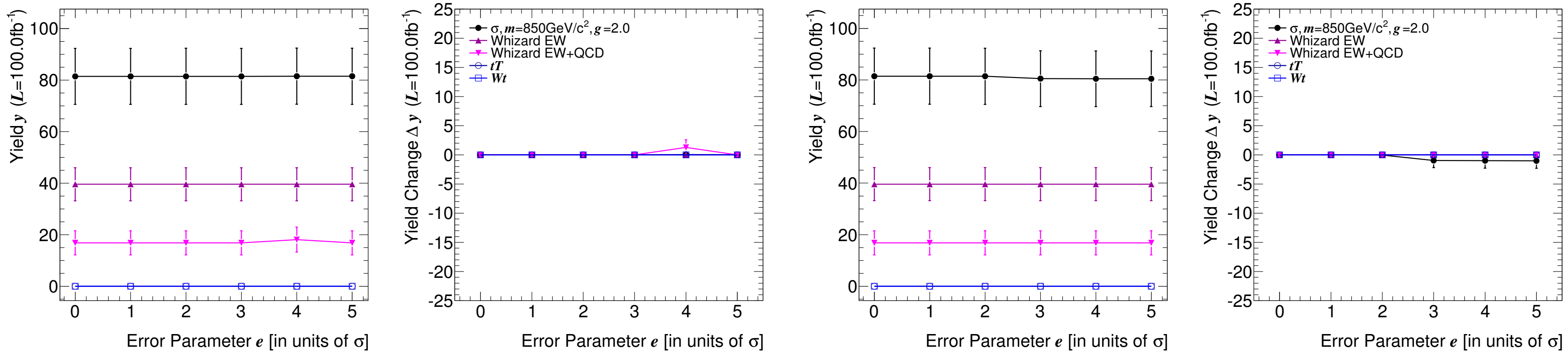
(c) Electron Energy Resolution

(d) Muon Reconstruction Resolution

Figure 7.11: *Expected event yield y on the left and change in event yield Δy after cuts on the right against error parameters e_j for multiple systematic uncertainties. Values are shown for a BDT cut of $r \geq 0.7$, and an integrated luminosity of $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. SM VBS is subtracted from the signal (cf. Section 5.4.) More details on the error bars can be found in the text.*



(a) Jet Energy Scale $0.0 < \eta \leq 3.2$ **(b)** Jet Energy Scale $\eta > 3.2$



(c) Jet Energy Resolution $0.0 < \eta \leq 3.2$ **(d)** Jet Energy Resolution $\eta > 3.2$

Figure 7.12: *Expected event yield y on the left and change in event yield Δy after cuts on the right against error parameters e_j for multiple systematic uncertainties. Values are shown for a BDT cut of $r \geq 0.7$, and an integrated luminosity of $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. SM VBS is subtracted from the signal (cf. Section 5.4.) More details on the error bars can be found in the text.*

7.3 Optimization of Boosted Decision Tree Cut

7.3.1 Method

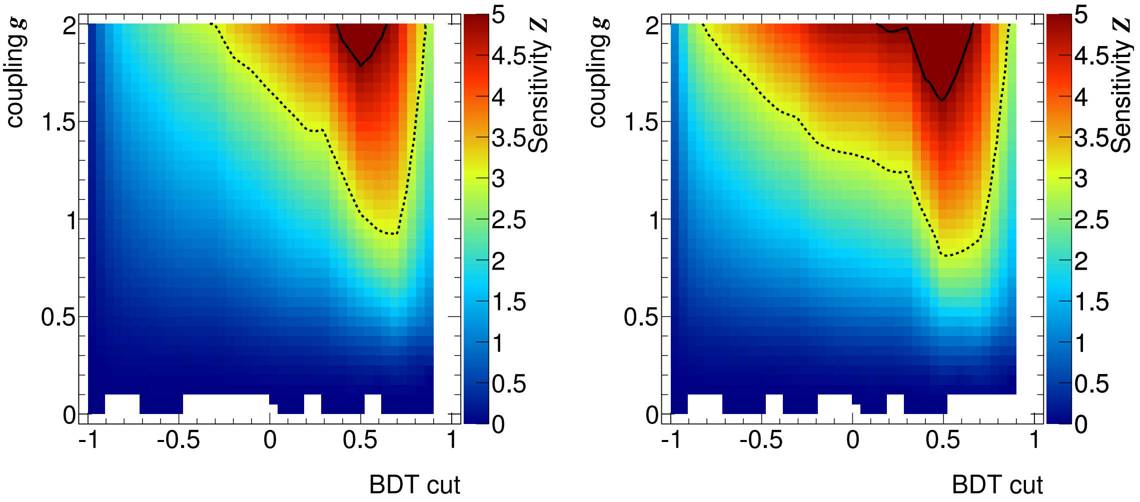
In the preceding sections all instruments for the determination of confidence belts have been laid out. The test statistic λ allows to calculate significances for hypothetical results of the experiment, the Asimov dataset (Section 7.1.) Fiducial cuts have been introduced to steer clear of limits at Monte Carlo simulation level and the trigger thresholds (Section 6.1.) A boosted decision tree has been trained to enrich the final sample in signal-like events (Section 6.5.) With the hypothesis definitions $H_1(g)$ for each resonance type and H_0 for the absence of any resonances, both discovery sensitivities (assuming H_1 is true), or expected limits (given H_0 is true) can be determined.

The analysis has at this point effectively one free parameter left: A cut value $r^{\text{cut}} \in [-1.0, 1.0)$ on the BDT response r, such that events with $r \geq r^{\text{cut}}$ will be accepted. The five resonances and two mass points under consideration together make up ten different scenarios. Consider the sensitivity determination for one such scenario of a chosen resonance and mass. Still undetermined are the coupling g which is a natural constant and the BDT cut value r^{cut}, which is to be optimized. For each point in the g vs. r^{cut} plane, one can estimate the event yield for H_0 and H_1 from Monte Carlo. Using these numbers, the statistical procedure just derived results in an expected significance, i.e. the sensitivity. This has been done for the scalar isoscalar σ of mass $850\,\text{GeV}/c^2$ in Figure 7.13a. The grid has a resolution of 0.1 in both directions, points in-between are interpolated.

The discovery and limit setting potential of the experiment is to be optimized. Both are determined by the ability to exclude either H_0 or H_1 with a given confidence. Traditionally, the confidence at which a hypothesis must be excluded for discovery corresponds to a 5σ one-sided Gaussian tail. For limits, the canonical confidence is 95%. For the purposes of this discussion, the optimal cut minimizes the coupling strength g, at which a 5σ discovery can be expected. Analogously, for limit setting, the optimal cut yields the lowest expected exclusion limit g. At the given granularity of 0.1 this can be read off directly from Figure 7.13a. The optimum is at $r^{\text{cut}} = 0.5$. For comparison, the same plot is shown without systematic uncertainties in Figure 7.13b.

7.3.2 Monte Carlo Equivalent Luminosity

Monte Carlo event abundance limits the statistical terms in Equation 7.5. However, when the experiment will have taken the amount of data supposed in this analysis, a lot more Monte Carlo events are expected to be available. All in all, only 25% of the experiment's luminosity are to be matched by Monte Carlo simulation [98]. This limit is predominantly determined by a few processes with relatively high cross section. Double leptonic $t\bar{t}$, which dominates the computational resource requirements of this analysis, is expected to be available at a higher relative equivalent luminos-

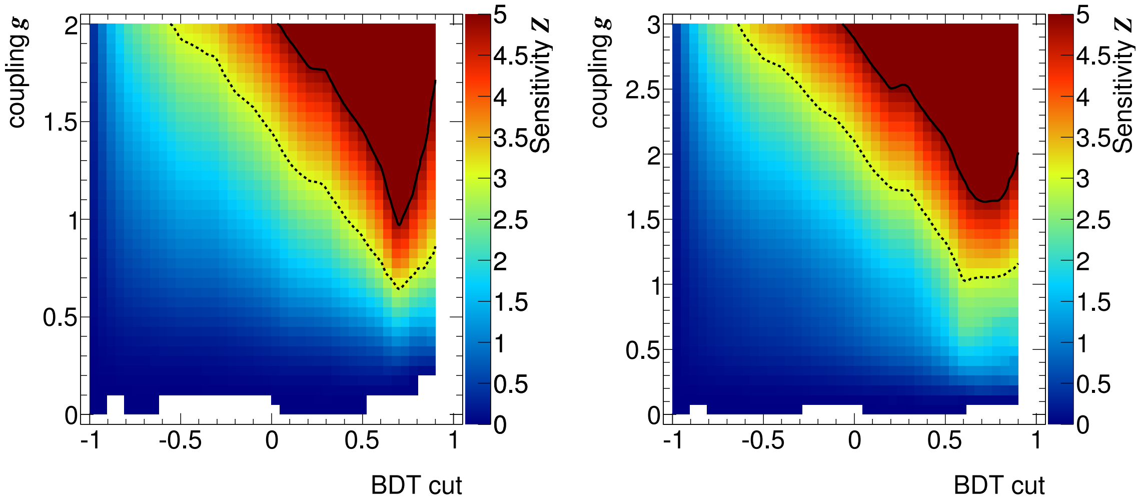


(a) Systematic Uncertainties Included **(b)** Systematic Uncertainties Ignored

Figure 7.13: *Expected significance Z (sensitivity) to exclude H_0 (SM) if $H_1(g)$ is realized for a scalar, isoscalar σ resonance of mass $m = 850\,\mathrm{GeV}/c^2$. Z is plotted against (reweighted) coupling g and BDT response cut r^{cut}. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line. A minimum coupling g, i.e. best sensitivity is achieved for a cut of $r \geq 0.5 = r^{\mathrm{cut}}$. White areas correspond to no sensitivity. $r^{\mathrm{cut}} = 1.0$ is shown for completeness, but this region is empty as all events are rejected. Left: Systematic uncertainties taken into account. Right: Systematic uncertainties are ignored. The optimum value for the cut is not affected, but the 5σ confidence line is lower.*

ity. Similarly for the much rarer and therefore easier to simulate WHIZARD signals and backgrounds. For this analysis, it is assumed that an equivalent luminosity of $\mathscr{L} = 10\,\mathrm{fb}^{-1}$ will be available for $t\bar{t}$ and Wt, corresponding to ≈ 9 million simulated events. The WHIZARD samples have a much lower cross section and $100\,\mathrm{fb}^{-1}$ are assumed to be available. For all samples combined, this corresponds to a bit less than a million events.

In order to quote sensitivities and limit expectations for the final experiment, these larger equivalent Monte Carlo luminosities have been assumed. Since this amount of simulation is not yet available, the yields from current samples have been scaled appropriately to arrive at the corresponding values for the Monte Carlo yields m_i in Equation (7.5). Figure 7.14a shows the resulting new BDT optimization plot taking both the higher assumed luminosity and systematic uncertainties into account. The optimal value for the BDT cut now moves to $r \geq r^{\mathrm{cut}} = 0.7$, which is used from now on unless otherwise noted. A comparison of the sensitivities achieved using the expected and current Monte Carlo luminosities is plotted in Figure 7.15. This new working point for the BDT cut has the desirable property to have a lower dependence



(a) Scalar, isoscalar σ, $m = 850\,\mathrm{GeV}/c^2$ **(b)** Scalar, isoscalar σ, $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.14: *Optimization plots for the BDT cut. Discovery sensitivity vs. BDT cut and g. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line. Note that the range of g differs between the types of resonances.*

on systematic effects than the old one as shown in Figure 7.16.

7.3.3 Generalization to Multiple Resonances

For the other resonances and mass points, corresponding coupling g versus BDT cut r^{cut} plots have been made. These can be found in figures 7.17 and 7.18 for the sensitivities and in 7.19, 7.20 and 7.21 for expected limits. The optimum is still not far from cut of $r^{\mathrm{cut}} = 0.7$. A further differentiation is not strictly necessary. Thus, the common cut value of 0.7 was used for all following results.

This has an important advantage in terms of statistical interpretation. As long as the same analysis is applied to the same data, no “look elsewhere effect” (cf. [99]) needs to be taken into account. This is true, even when interpreting the results multiple times in the context of different theoretical models. If the analysis may change, i.e. is optimized for multiple models individually, this is no longer strictly true. The difference between these cases is the possibility of individual fluctuations between the analysis. Multiple analyses, even when examining the same data, may select different event subsets. This would certainly be the case if different BDT cuts were chosen. Ignoring their correlation, one would be a subset of the other, there are now two chances to see a large upward fluctuation. If one may claim discovery in either case, the coverage of the confidence bands would then be incorrect. That is similar to the effect that a sliding mass window has in other analyses. Interpreting the same analysis multiple times does not give rise to multiple chances of fluctuation.

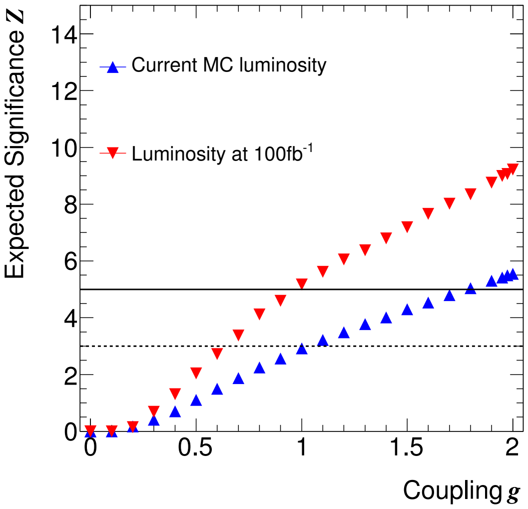
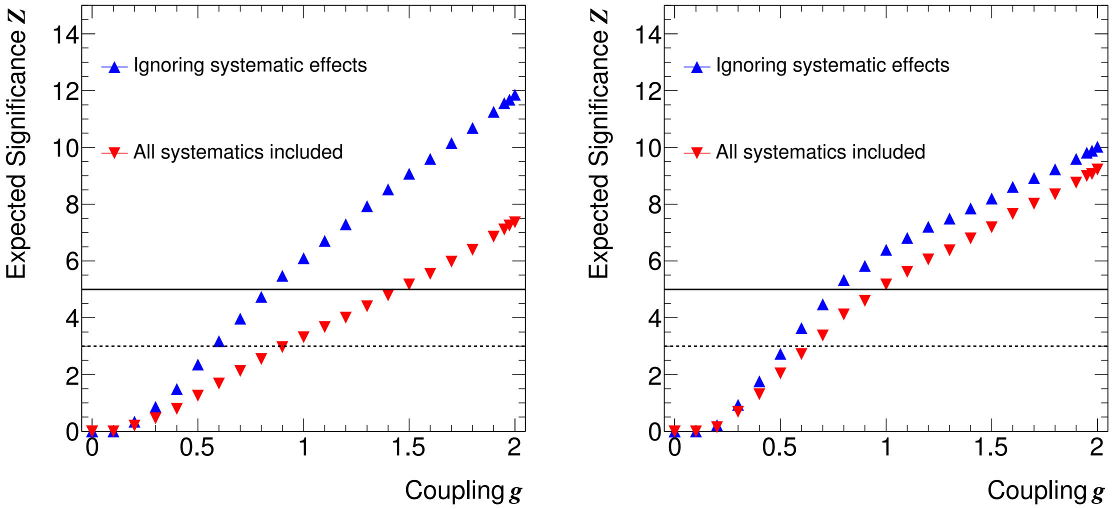
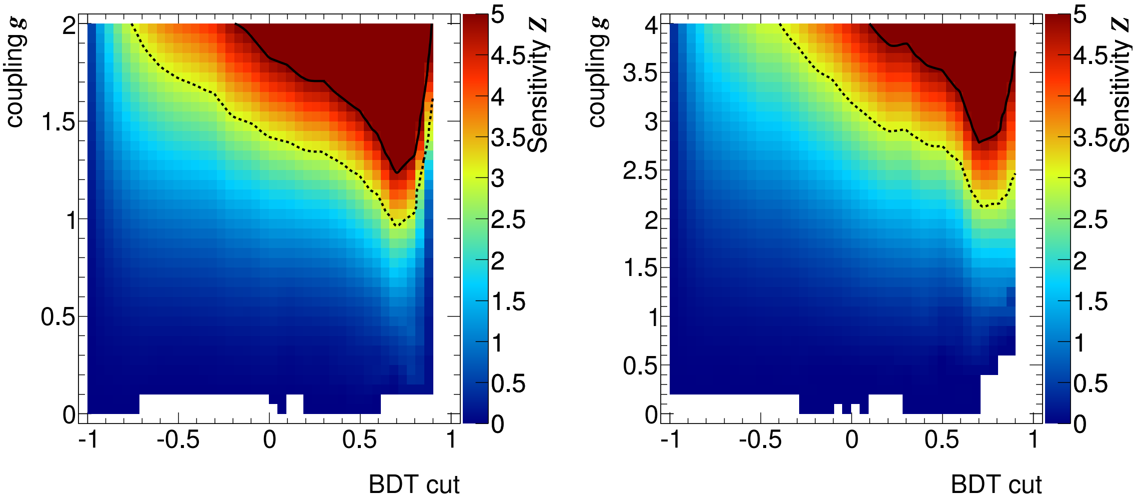


Figure 7.15: *Discovery sensitivity vs. coupling g for a scalar isoscalar resonance σ of mass $m = 850\,\mathrm{GeV}/c^2$. Plotted are sensitivities for the current Monte Carlo luminosity and the luminosity expected to be available at the time of the experiment. For each sample, the optimal BDT cut was used. Current MC luminosity: $r^{\mathrm{cut}} = 0.5$. Expected MC luminosity: $r^{\mathrm{cut}} = 0.7$. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line.*

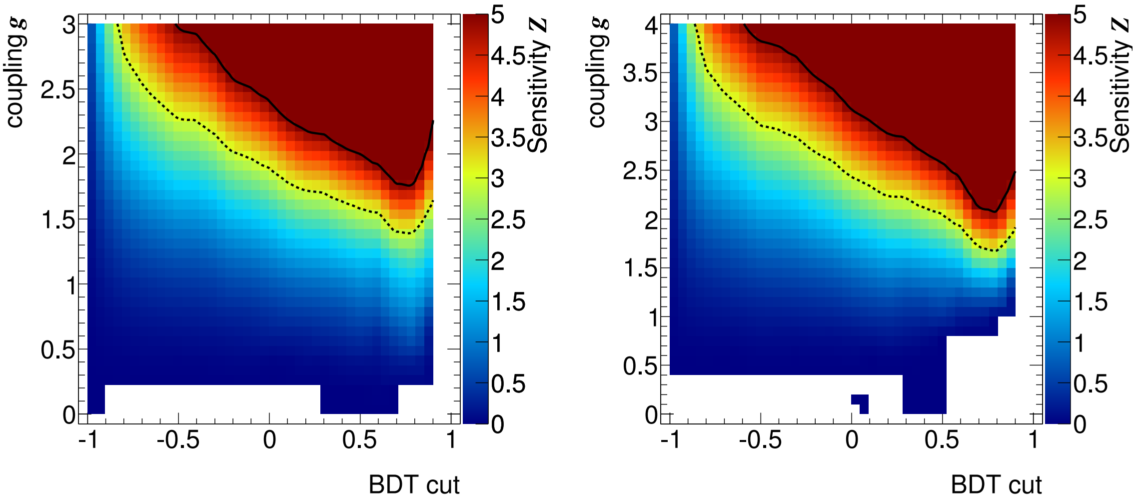


(a) BDT cut $r \geq 0.5$ **(b)** BDT cut $r \geq 0.7$

Figure 7.16: *Discovery sensitivity vs. coupling g for a scalar isoscalar resonance σ of mass $m = 850\,\mathrm{GeV}/c^2$ with the expected Monte Carlo luminosity at $\mathscr{L} = 100\,\mathrm{fb}^{-1}$. Plotted are sensitivities including and ignoring systematic uncertainties. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line.*

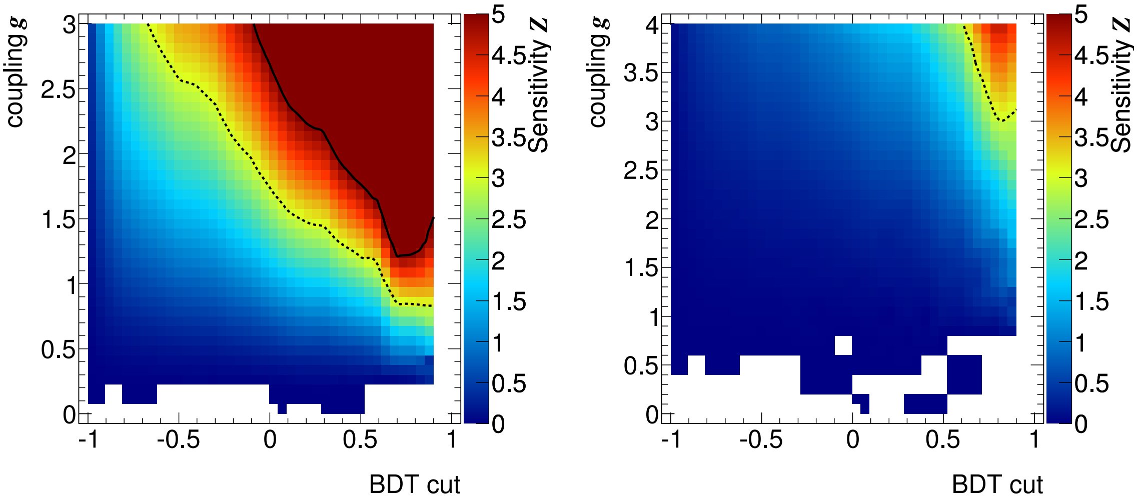


(a) Scalar, isotensor ϕ, $m = 850\,\mathrm{GeV}/c^2$ **(b)** Vector, isovector ρ, $m = 850\,\mathrm{GeV}/c^2$

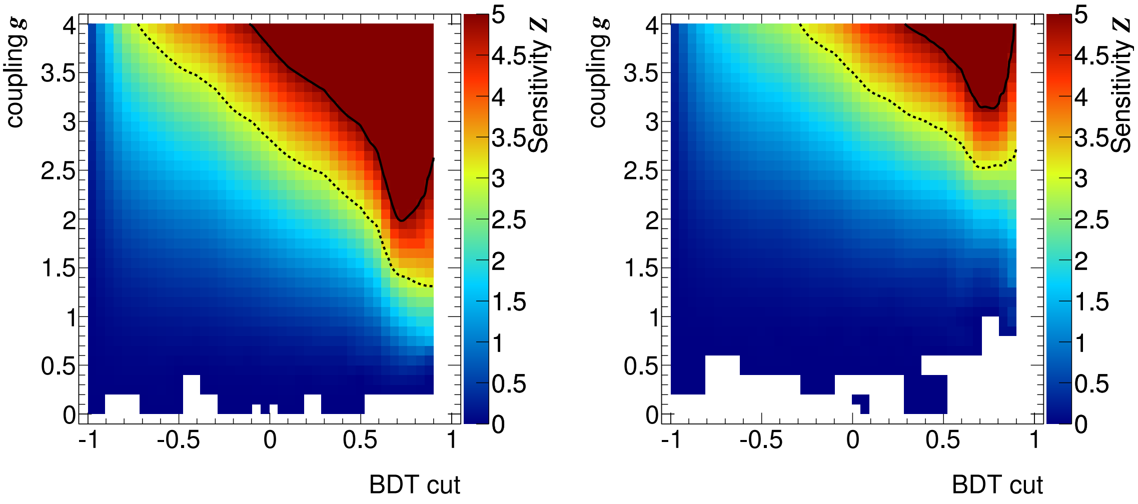


(c) Tensor, isoscalar f, $m = 850\,\mathrm{GeV}/c^2$ **(d)** Tensor, isotensor t, $m = 850\,\mathrm{GeV}/c^2$

Figure 7.17: *Optimization plots for the BDT cut. Discovery sensitivity vs. BDT cut and g. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line. Note that the range of g differs between the types of resonances.*

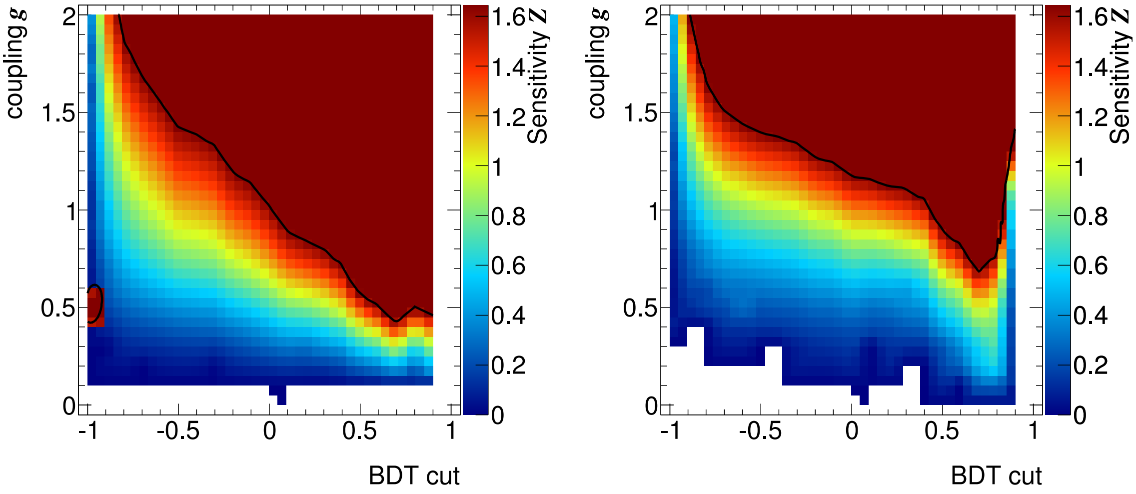


(a) Scalar, isotensor ϕ, $m = 1150\,\mathrm{GeV}/c^2$ **(b)** Vector, isovector ρ, $m = 1150\,\mathrm{GeV}/c^2$

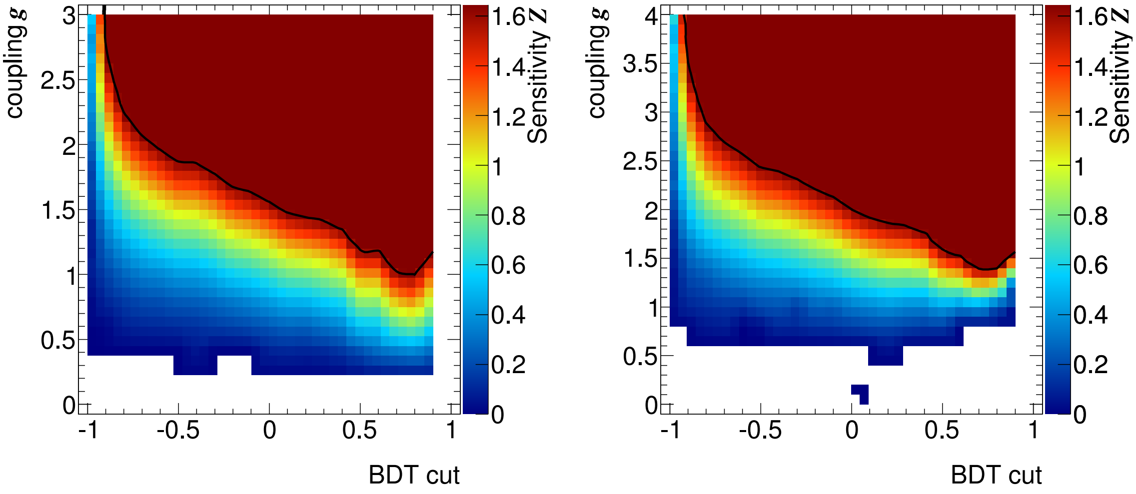


(c) Tensor, isoscalar f, $m = 1150\,\mathrm{GeV}/c^2$ **(d)** Tensor, isotensor t, $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.18: *Optimization plots for the BDT cut. Discovery sensitivity vs. BDT cut and g. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line. Note that the range of g differs between the types of resonances.*

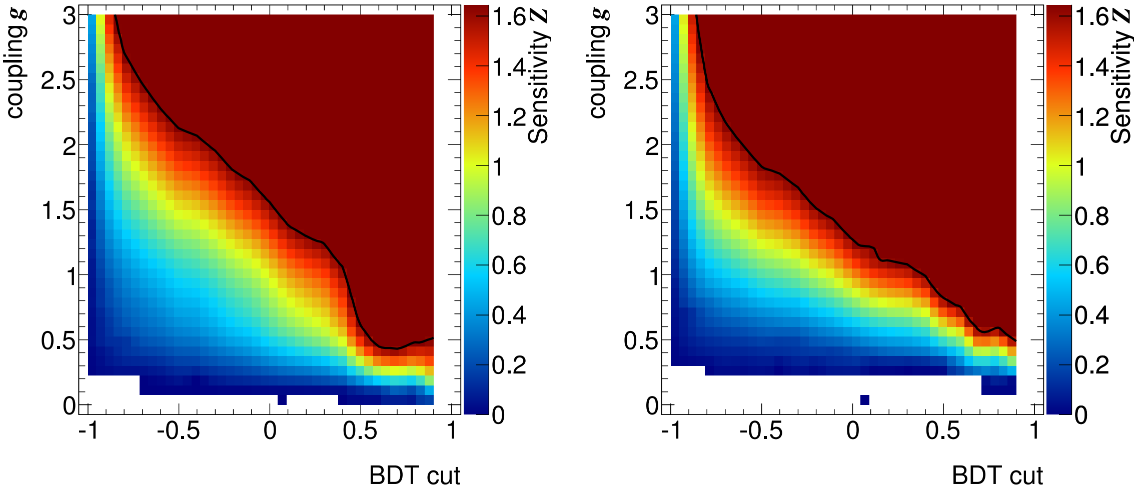


(a) Scalar, isoscalar σ, $m = 850\,\mathrm{GeV}/c^2$ **(b)** Scalar, isotensor ϕ, $m = 850\,\mathrm{GeV}/c^2$



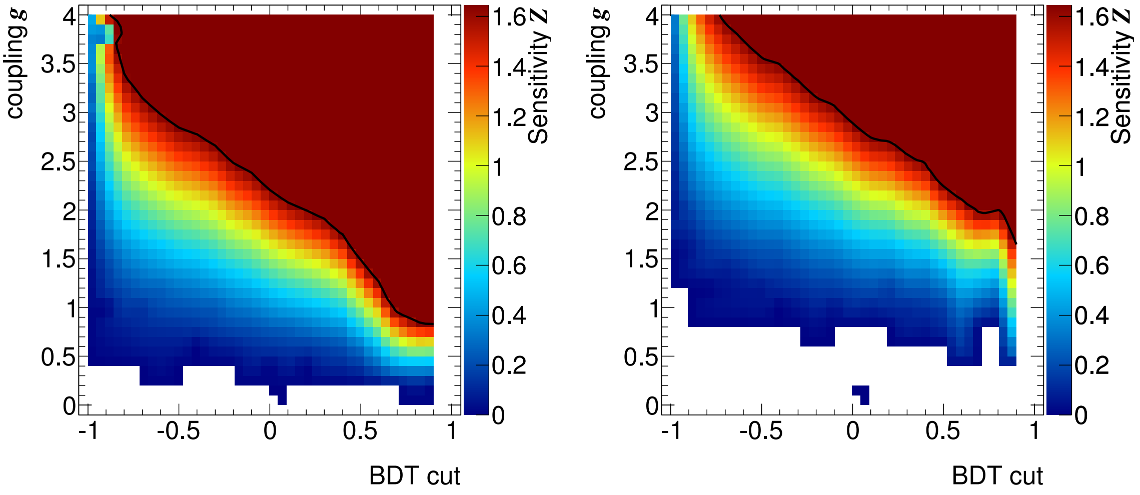
(c) Tensor, isoscalar f, $m = 850\,\mathrm{GeV}/c^2$ **(d)** Tensor, isotensor t, $m = 850\,\mathrm{GeV}/c^2$

Figure 7.19: *Optimization plots for the BDT cut. Expected limit vs. BDT cut and g. The 1.64σ (95%) CL contour is marked with a solid line. Note that the range of g differs between the types of resonances. The high significance for the scalar isoscalar σ for $r^{\mathrm{cut}} = -1.0$ and $g = 0.5$ is a numerical instability. It is ignored in the analysis.*



(a) Scalar, isoscalar σ, $m = 1150\,\mathrm{GeV}/c^2$

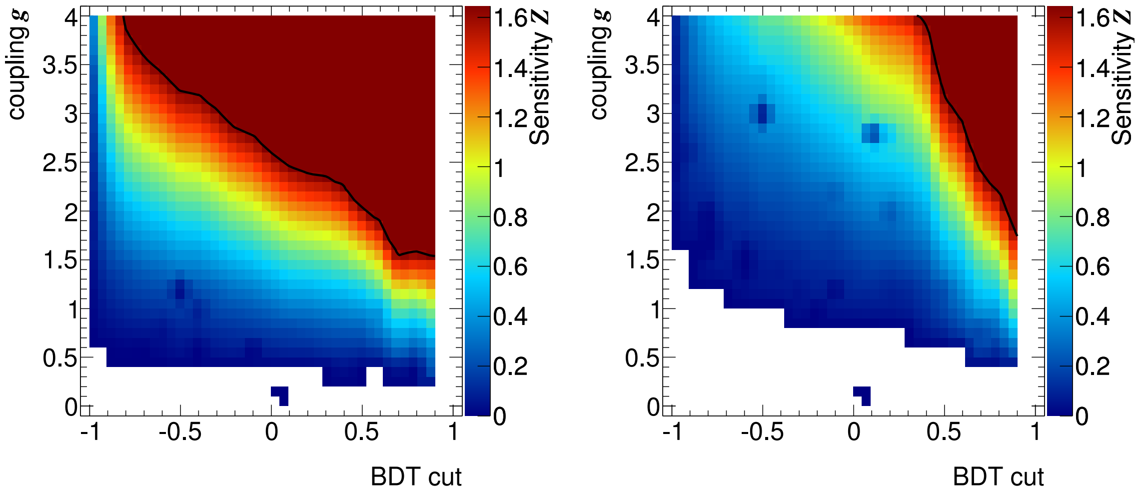
(b) Scalar, isotensor ϕ, $m = 1150\,\mathrm{GeV}/c^2$



(c) Tensor, isoscalar f, $m = 1150\,\mathrm{GeV}/c^2$

(d) Tensor, isotensor t, $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.20: *Optimization plots for the BDT cut. Expected limit vs. BDT cut and* g*. The* 1.64σ *(95%) CL contour is marked with a solid line. Note that the range of* g *differs between the types of resonances.*



(a) Vector, isovector ρ, $m = 850\,\mathrm{GeV}/c^2$ **(b)** Vector, isovector ρ, $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.21: *Optimization plots for the BDT cut. Expected limit vs. BDT cut and g. The 1.64σ (95%) CL contour is marked with a solid line. Note that the range of g differs between the types of resonances. Numerical instabilities are visible for a few points in regions of low significance. They do not affect the interpretation.*

A small caveat are the fluctuations of the signal Monte Carlo samples, as they are not shared. These are the cheapest to produce for large equivalent luminosities, however, so it is believed that this problem can be overcome.

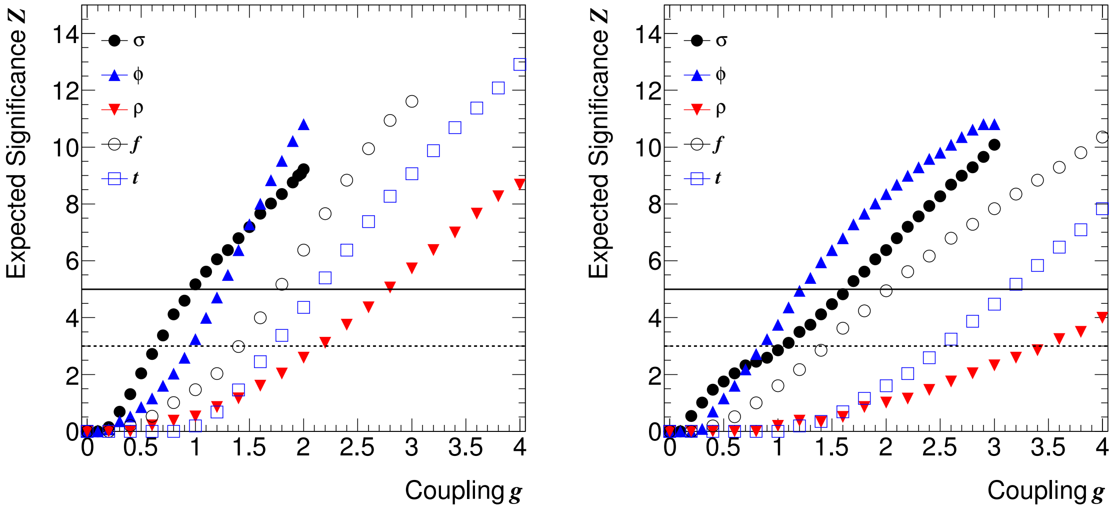
There is a flip side to this: While there is no look elsewhere effect with one analysis, it is not directly possible to distinguish between different resonances either.

7.4 Results

With last section's result for an optimal BDT cut, it was possible to derive the coupling strengths needed to expect discovery. As a visual guide and to compare the different resonances, Figure 7.22 contains plots of the expected discovery significance versus coupling g for all examined resonances and mass points. For 3σ and 5σ expected significance, the corresponding coupling is given in Table 7.4. For comparison, a heavy Higgs boson corresponds to a scalar isoscalar resonance with coupling $g = 1$. If present, it is expected to be (re-)discovered in this channel if it has a mass of $m = 850\,\mathrm{GeV}/c^2$. At $m = 1150\,\mathrm{GeV}/c^2$, only evidence at 3σ level is expected.

Theories with strong electroweak symmetry breaking may exhibit couplings that are noticeably larger than one. From a theoretical prejudice one would expect the term $\frac{g^2}{2\pi}$ not to exceed one. That would allow couplings of up to $g \approx 2.5$ [46]. Except for the vector isovector ρ, for all of the resonances, if present, there is some discovery reach within that limit at $m = 850\,\mathrm{GeV}/c^2$. At $m = 1150\,\mathrm{GeV}/c^2$, the tensor isotensor t becomes too elusive as well. Evidence at the 3σ level only is expected for t for $g \approx 2.5$.

A preceding study [33] within ATLAS has investigated Padé unitarization models. For some parameters, a scalar isoscalar is predicted, which corresponds to the σ particle investigated here with a coupling constant of $g = \sqrt{2/3} \approx 0.81$ [34]. The study examines amongst amongst others the semileptonic final state $\ell\nu jj$, which allows for WW and WZ scattering. The latter is not included in this analysis. While it estimates a discovery significance of 5σ to be reached with as little as $\mathscr{L} = 60\,\mathrm{fb}^{-1}$ of integrated luminosity, this does not include systematic uncertainties. It is advantaged using the semileptonic channel. Results from Table 7.4 show that a lot of the disadvantages from using the leptonic channel can be recovered with a more sophisticated cut method using an MVA.



(a) $m = 850\,\mathrm{GeV}/c^2$ **(b)** $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.22: *Expected significance plots against coupling g for all examined resonances and the two mass points $m = 850\,\mathrm{GeV}/c^2$ and $m = 1150\,\mathrm{GeV}/c^2$. For a particular demanded significance, one can read off the coupling needed by finding the appropriate point along a horizontal line. Significance contours are marked for 5σ CL with a solid line and for 3σ CL with a dashed line.*

Analogous results for the expected upper limits can be found in Figure 7.23 and Table 7.5. While the tested hypotheses stay the same (H_0 against $H_1(g)$) the problem is not symmetrical. In particular, it tends to be easier to fluctuate down from a high expectation value to a low one. Thus, it is more difficult to exclude $H_1(g)$ when H_0 is given, than the other way around, giving a lower expected significance. Since the confidence level for exclusion limits is customarily only $95\% \approx 1.64\sigma$ instead of the 5σ one-sided Gaussian tail demanded for discovery, the numerical value for the

Resonance	Mass [GeV/c^2]	3σ	5σ
σ	850	0.64	0.97
ϕ	850	0.96	1.24
ρ	850	2.16	2.78
f	850	1.40	1.77
t	850	1.72	2.12
σ	1150	1.06	1.64
ϕ	1150	0.85	1.21
ρ	1150	3.47	4.40
f	1150	1.44	2.02
t	1150	2.53	3.15

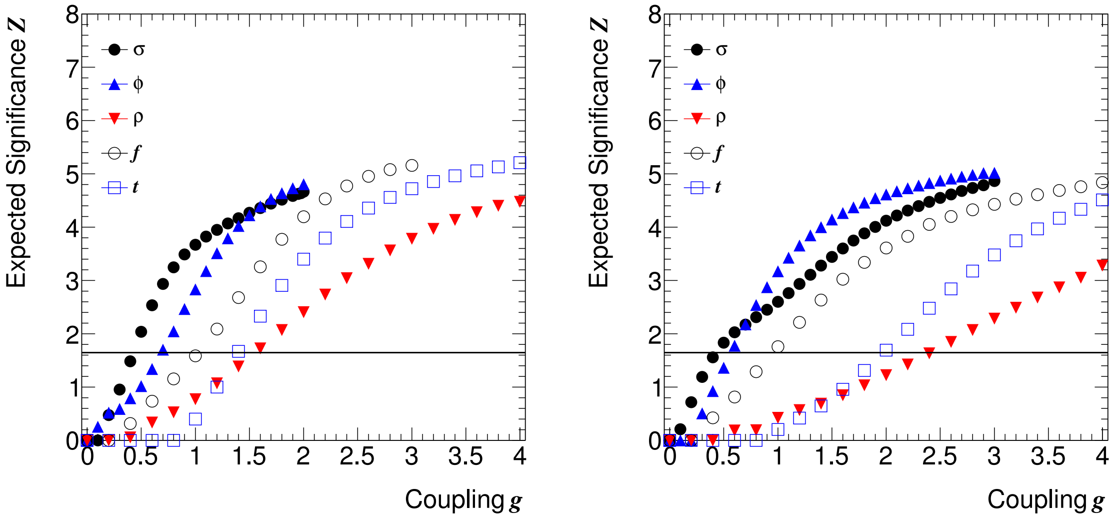
Table 7.4: *Coupling g needed such that the expected discovery significance reaches 3σ or 5σ, respectively.*

Resonance	Mass [GeV/c^2]	95% CL
σ	850	0.43
ϕ	850	0.68
ρ	850	1.55
f	850	1.02
t	850	1.39
σ	1150	0.43
ϕ	1150	0.57
ρ	1150	2.41
f	1150	0.95
t	1150	1.98

Table 7.5: *Coupling needed such that the expected upper limit significance reaches 95% confidence.*

couplings at the limit are still lower. A second important aspects is the use of the Asimov method. When testing H_0 assuming $H_1(g)$ is realized, i.e. discovery, the uncertainties on the signal do not come into play. Only the parameters of H_0 are fitted. One tries to see if the assumed data can be explained with H_0. Signal plays a role in that $n = s + b$ is assumed. Here, the estimators from Monte Carlo enter, not their uncertainty. In the case of limit setting, it is the other way around, giving the model all the additional freedom of $H_1(g)$ to explain a measurement that would have been expected form the SM H_0.

For all resonances and both mass points, the minimal coupling needed to expect exclusion is below $g = 2.5$ for a vector isovector ρ of mass $m = 1150\,\mathrm{GeV}/c^2$ and reaches down to $g = 0.43$ for a scalar isoscalar σ at both mass points $m = 850\,\mathrm{GeV}/c^2$ and $m = 1150\,\mathrm{GeV}/c^2$. A SM Higgs boson like scalar isoscalar as well as a Padé scalar should be excluded in this channel.



(a) $m = 850\,\mathrm{GeV}/c^2$ **(b)** $m = 1150\,\mathrm{GeV}/c^2$

Figure 7.23: *Expected limit significance plots against coupling* g *for all examined resonances and the two mass points* $m = 850\,\mathrm{GeV}/c^2$ *and* $m = 1150\,\mathrm{GeV}/c^2$. *For a particular significance, one can read off the coupling needed by finding the appropriate point along a horizontal line. Note that the* g *scale is different than for discovery. The* 1.64σ *(95%) CL contour is marked with a solid line.*

8 Conclusions and Outlook

In this thesis, a Monte Carlo simulation analysis of the dileptonic vector boson scattering channel is presented. Longitudinal vector boson scattering is the one channel, where electroweak symmetry breaking must have an effect that the current Standard Model of particle physics cannot account for.

For the first time a new effective model combining the electroweak chiral Lagrangian, vector boson scattering resonances and K-Matrix unitarization was studied experimentally. It greatly reduces the amount of assumptions about new physics necessary compared to current models, like the Padé model studied previously in ATLAS. In combination with the implementation in the Whizard event generator of the same authors, the effective W approximation is no longer necessary. The treatment incorporates off-shell effects and the full 6-fermion final state including irreducible background. Angular correlations, which are an important experimental signature, are conserved.

The Whizard event generator was validated for ATLAS. To that end, the correspondence between validated processes and specific parameter points of the new K-Matrix model was exploited. A generic interface for LHEF files was adapted to interface Whizard to the ATLAS software Athena.

To allow the reuse of precious fully simulated signal samples, a reweighting technique was implemented and tested. It was shown that a reweighting of couplings g to lower values $g' < g$ is possible down to vanishing couplings. The K-Matrix model makes definite predictions in the high-energy region beyond the first resonance of a hypothesized resonance spectrum. It makes sure that the increase of cross section due to unitarity violation is suppressed, removing the most important source of spurious sensitivity in earlier analyses without unitarity constraints. In this thesis, null and signal hypotheses were defined within the K-Matrix model to further reduce any residual effects from overestimation of cross sections at highest energies. The concrete implementation of the signal via subtraction of samples reweighted to vanishing couplings $g = 0$ reduces the statistical uncertainty from Monte Carlo.

A multivariate analyzer was used for the main selection to improve the significance compared to conventional cut analyses. Angular correlations, now correctly simulated by Whizard, were used as discriminating variables. Both measures allowed to approximately recover the sensitivity disadvantage the dileptonic channel has compared to channels which allow m_{VV} reconstruction.

The statistical analysis was performed using a profile likelihood method correctly accounting for statistical uncertainties in the Monte Carlo predictions. A corresponding tool was implemented and appropriate ensemble tests performed validating the method. Systematical uncertainties were included and their effect on the final analysis was estimated in simulations of the whole analysis chain. Pileup was not considered, but a follow-up study is underway. No drastic changes have been seen so far.

It was shown that ATLAS has potential for discovery and limit setting for new resonances in vector boson scattering with $\mathscr{L} = 100\,\mathrm{fb}^{-1}$ of data. The scalar resonances σ and ϕ are the easiest to detect, both having expected discovery limits around $g = 1.0$ for a mass of $m = 850\,\mathrm{GeV}/c^2$. The vector isovector resonance ρ is the least detectable. At $m = 850\,\mathrm{GeV}/c^2$ a coupling of $g = 2.8$ is needed to expect discovery, which is on the border of what would seem natural even for strong electroweak symmetry breaking. For higher masses, even evidence for a ρ seems unlikely with $\mathscr{L} = 100\,\mathrm{fb}^{-1}$ of data. The tensor isotensor t is also very difficult to detect at $m = 1150\,\mathrm{GeV}/c^2$. All other resonances are discoverable for coupling strengths that still seem reasonable. 95% limits are consistently easier to reach than discovery, due to the lower confidence level needed, despite additional uncertainty from signal modeling. For the ρ resonance and the t resonance at the higher mass point expected limits are close to the edge of reasonable couplings. For the other resonances, a larger part of the parameter space may be excluded. A SM Higgs boson should be (re-)discovered at $m = 850\,\mathrm{GeV}/c^2$. At $m = 1150\,\mathrm{GeV}/c^2$ only evidence is expected, as for a Padé scalar of $m = 850\,\mathrm{GeV}/c^2$.

The conclusion about the Padé scalar gives a handle for comparison with an earlier analysis of the semileptonic channel. Taking into account the newly included systematic uncertainties, the overall sensitivity is comparable to the earlier study. With the first data, understanding of the detector is becoming the most important task. For vector boson scattering, both forward jets and highly boosted objects in the central regions are particular challenges. Data will allow to model these more reliably. As the semileptonic channel is advantaged in terms of cross section and reconstruction of the invariant diboson mass m_{VV}, a repetition of the analysis using the upgraded model, generator and statistical methods seems promising. The m_{VV} distribution is interesting candidate for an extraction of the background from data, reducing systematic uncertainties.

Whether the standard model Higgs mechanism is realized or an alternative, any model will have to account for longitudinal vector boson scattering. The LHC will be the first collider to measure vector boson scattering in the region of manifestation of electroweak symmetry breaking effects.

Bibliography

[1] S. L. Glashow. «Partial Symmetries of Weak Interactions.» *Nucl. Phys.*, **22**:pp. 579–588, 1961.

[2] S. Weinberg. «A Model of Leptons.» *Phys. Rev. Lett.*, **19**:pp. 1264–1266, 1967.

[3] A. Salam. *in Elementary Particle Theory, p. 367.* Almqvist and Wiksell, Stockholm, 1968.

[4] P.W.Higgs. «Spontaneous Symmetry Breakdown without Massless Bosons.» *Phys. Rev.*, **145**:p. 1156, 1965.

[5] T. L. E. W. Group. «Blueband Plot.», 2010. URL `http://lepewwg.web.cern.ch/LEPEWWG/`.

[6] T. Appelquist and G. Wu. «Electroweak chiral Lagrangian and new precision measurements.» *Phys. Rev. D*, **48**:pp. 3235–3241, 1993. `arXiv:hep-ph/9304240`.

[7] M. Veltman. «Second Threshold in Weak Interactions.» *Acta Phys. Pol. B*, **8**:p. 475, 1977.

[8] M. Veltman. «Limit on mass differences in the Weinberg model.» *Nuclear Physics B*, **123(1)**:pp. 89 – 99, 1977. URL `http://www.sciencedirect.com/science/article/B6TVC-472T6BB-2H1/2/0888826eee83f354e180fe3c3e2737d9`.

[9] P. Sikivie et al. «Isospin breaking in technicolor models.» *Nuclear Physics B*, **173(2)**:pp. 189 – 207, 1980. URL `http://www.sciencedirect.com/science/article/B6TVC-4719T2G-2BH/2/75988705bc6a1bf257e97ab0e4a38d1c`.

[10] J. Bagger et al. «Strongly interacting WW system: Gold-plated modes.» *Phys. Rev. D*, **49**:pp. 1246–1264, 1994. `arXiv:hep-ph/9306256`.

[11] J. Bagger et al. «CERN LHC analysis of the strongly interacting WW system: Gold-plated modes.» *Phys. Rev. D*, **52**:pp. 3878–3889, 1995. `arXiv:hep-ph/9504426`.

[12] J. Goldstone, A. Salam and S. Weinberg. «Broken Symmetries.» *Phys. Rev.*, **127**:pp. 965–970, 1962.

[13] D. J. Gross and F. Wilczek. ≪Ultraviolet Behavior of Non-Abelian Gauge Theories.≫ *Phys. Rev. Lett.*, **30(26)**:pp. 1343–1346, 1973.

[14] E. Gildener and S. Weinberg. ≪Symmetry breaking and scalar bosons.≫ *Phys. Rev. D*, **13(12)**:pp. 3333–3341, 1976.

[15] S. Weinberg. ≪Implications of dynamical symmetry breaking: An addendum.≫ *Phys. Rev. D*, **19(4)**:pp. 1277–1280, 1979.

[16] L. Susskind. ≪Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory.≫ *Phys. Rev. D*, **20(10)**:pp. 2619–2625, 1979.

[17] E. Farhi and L. Susskind. ≪Grand unified theory with heavy color.≫ *Phys. Rev. D*, **20(12)**:pp. 3404–3411, 1979.

[18] S. Dimopoulos and L. Susskind. ≪Mass without scalars.≫ *Nuclear Physics B*, **155(1)**:pp. 237 – 252, 1979. URL `http://www.sciencedirect.com/science/article/B6TVC-472T23W-162/2/cf435b626c937e28c883de35dc4e45a1`.

[19] E. Eichten and K. Lane. ≪Dynamical breaking of weak interaction symmetries.≫ *Physics Letters B*, **90(1-2)**:pp. 125 – 130, 1980. URL `http://www.sciencedirect.com/science/article/B6TVN-46YKGTF-VK/2/4616b90d40d2e98be1062c3d45dae882`.

[20] W. A. Bardeen, C. T. Hill and M. Lindner. ≪Minimal dynamical symmetry breaking of the standard model.≫ *Phys. Rev. D*, **41(5)**:pp. 1647–1660, 1990.

[21] C. T. Hill. ≪Topcolor: top quark condensation in a gauge extension of the standard model.≫ *Physics Letters B*, **266(3-4)**:pp. 419 – 424, 1991. URL `http://www.sciencedirect.com/science/article/B6TVN-46YT4MC-4Y5/2/caa1aed3e5ab7dfa790df1e504681df9`.

[22] C. T. Hill. ≪Topcolor assisted technicolor.≫ *Physics Letters B*, **345(4)**:pp. 483 – 489, 1995. URL `http://www.sciencedirect.com/science/article/B6TVN-3YMWP3J-CH/2/31bee56491532e5e7e65a00fd5c88fd9`.

[23] K. Lane and E. Eichten. ≪Natural topcolor-assisted technicolor.≫ *Physics Letters B*, **352(3-4)**:pp. 382 – 387, 1995. URL `http://www.sciencedirect.com/science/article/B6TVN-3YMWPXF-90/2/f46548b5a7809914a04dec7e7c6ac6e0`.

[24] C. T. Hill, M. A. Luty and E. A. Paschos. ≪Electroweak symmetry breaking by fourth-generation condensates and the neutrino spectrum.≫ *Phys. Rev. D*, **43(9)**:pp. 3011–3025, 1991.

[25] S. P. Martin. ≪Dynamical electroweak symmetry breaking with top-quark and neutrino condensates.≫ *Phys. Rev. D*, **44(9)**:pp. 2892–2898, 1991.

[26] E. Akhmedov et al. ≪Dynamical left-right symmetry breaking.≫ *Phys. Rev. D*, **53(5)**:pp. 2752–2780, 1996.

[27] E. Akhmedov et al. ≪Left-right symmetry breaking in NJL approach.≫ *Physics Letters B*, **368(4)**:pp. 270 – 280, 1996. URL `http://www.sciencedirect.com/science/article/B6TVN-3VTC5PC-4J/2/951b4998e0b0468df8c47e279cfe106b`.

[28] B. A. Dobrescu and C. T. Hill. ≪Electroweak Symmetry Breaking via a Top Condensation Seesaw Mechanism.≫ *Phys. Rev. Lett.*, **81(13)**:pp. 2634–2637, 1998.

[29] R. S. Chivukula et al. ≪Top quark seesaw theory of electroweak symmetry breaking.≫ *Phys. Rev. D*, **59(7)**:p. 075003, 1999.

[30] C. Csaki. ≪Higgsless electroweak symmetry breaking.≫ 2004. `hep-ph/0412339`.

[31] R. S. Chivukula et al. ≪A Three site Higgsless model.≫ *Phys. Rev. D*, **74(7)**:p. 075011, 2006.

[32] N. Arkani-Hamed et al. ≪The Littlest Higgs.≫ *Journal of High Energy Physics*, **7**:pp. 34–+, 2002. `arXiv:hep-ph/0206021`.

[33] A. Collaboration. *Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics*. CERN, Geneva, 2009.

[34] A. Alboteanu, W. Kilian and J. Reuter. ≪Resonances and unitarity in weak boson scattering at the LHC.≫ *Journal of High Energy Physics*, **11**:pp. 10–+, 2008. `0806.4145`.

[35] J. Goldstone. ≪Field theories with "Superconductor" solutions.≫ *Il Nuovo Cimento (1955-1965)*, **19**:pp. 154–164, 1961. 10.1007/BF02812722, URL `http://dx.doi.org/10.1007/BF02812722`.

[36] Y. Nambu. ≪Axial Vector Current Conservation in Weak Interactions.≫ *Phys. Rev. Lett.*, **4(7)**:pp. 380–382, 1960.

[37] J. Goldstone, A. Salam and S. Weinberg. ≪Broken Symmetries.≫ *Phys. Rev.*, **127(3)**:pp. 965–970, 1962.

[38] W. Kilian. ≪Electroweak Symmetry Breaking.≫ *Springer Tracts Mod. Phys.*, **198/2004**, 2004.

[39] F. Gangemi et al. ≪Electroweak physics in six-fermion final states at future e+e- colliders.≫ *ArXiv High Energy Physics - Phenomenology e-prints*, 2000. `arXiv:hep-ph/0001065`.

[40] P. Krstonosic et al. «Experimental studies of Strong Electroweak Symmetry Breaking in gauge boson scattering and three gauge boson production.» *ArXiv High Energy Physics - Phenomenology e-prints*, 2005. `arXiv:hep-ph/0508179`.

[41] M. Mertens et al. «Monte Carlo study on anomalous quartic couplings in the scattering of weak gauge bosons with the ATLAS detector.» *Technical Report ATL-COM-PHYS-2007-021*, CERN, Geneva, 2007.

[42] S. N. Gupta. *Quantum. Electrodynamics, Second Edition.* Gordon and Breach, New York, 1981.

[43] M. S. Chanowitz. «Quantum corrections from nonresonant WW scattering.» *Physics Reports*, **320(1-6)**:pp. 139 – 146, 1999. URL `http://www.sciencedirect.com/science/article/B6TVP-3YDNWYC-D/2/ff22cc0e656187b43228a25c92ad32c6`.

[44] A. Dobado and J. R. Peláez. «Inverse amplitude method in chiral perturbation theory.» *Phys. Rev. D*, **56**:pp. 3057–3073, 1997. `arXiv:hep-ph/9604416`.

[45] R. Casalbuoni, S. de Curtis and D. Dominici. «Indirect effects of new resonances at future linear colliders.» *Physics Letters B*, **403**:pp. 86–92, 1997. `arXiv:hep-ph/9702357`.

[46] J. Reuter. «Private Communication.»

[47] D. Rainwater and D. Zeppenfeld. «Observing $H \to W^{(*)}W^{(*)} \to We^{+/-}\mu^{-/+}p\!\!/_T$ in weak boson fusion with dual forward jet tagging at the CERN LHC.» *Phys. Rev. D*, **60(11)**:pp. 113004–+, 1999. `arXiv:hep-ph/9906218`.

[48] S. Asai et al. «Prospects for the search for a standard model Higgs boson in ATLAS using vector boson fusion.» *The European Physical Journal C - Particles and Fields*, **32**:pp. s19–s54, 2004. 10.1140/epjcd/s2003-01-010-8, URL `http://dx.doi.org/10.1140/epjcd/s2003-01-010-8`.

[49] T. S. Pettersson and P. Lefèvre. «The Large Hadron Collider: conceptual design. oai:cds.cern.ch:291782.» *Technical Report CERN-AC-95-05 LHC*, CERN, Geneva, 1995.

[50] A. Horvath, 2006. Http://commons.wikimedia.org/wiki/File:LHC.svg.

[51] The ATLAS Collaboration. «The ATLAS Experiment at the CERN Large Hadron Collider.» *JINST 3*, **S08003**, 2008.

[52] W. Kilian, T. Ohl and J. Reuter. «WHIZARD: Simulating Multi-Particle Processes at LHC and ILC.» *ArXiv e-prints*, 2007. `0708.4233`.

[53] M. Moretti, T. Ohl and J. Reuter. «O'Mega: An Optimizing Matrix Element Generator.» *ArXiv High Energy Physics - Phenomenology e-prints*, 2001. `arXiv:hep-ph/0102195`.

[54] T. Sjöstrand, S. Mrenna and P. Skands. ≪PYTHIA 6.4 physics and manual.≫ *Journal of High Energy Physics*, **5**:pp. 26–+, 2006. `arXiv:hep-ph/0603175`.

[55] A. Dobado et al. ≪CERN LHC sensitivity to the resonance spectrum of a minimal strongly interacting electroweak symmetry breaking sector.≫ *Phys. Rev. D*, **62(5)**:pp. 055011–+, 2000. `arXiv:hep-ph/9912224`.

[56] G. Azuelos et al. ≪Electroweak Physics.≫ **(hep-ph/0003275. CERN-TH-2000-102)**:p. 114 p, 2000.

[57] M. S. Chanowitz and M. K. Gaillard. ≪Multiple production of W and Z as a signal of new strong interactions.≫ *Physics Letters B*, **142(1-2)**:pp. 85 – 90, 1984. URL `http://www.sciencedirect.com/science/article/B6TVN-472JRBS-171/2/dcdb0c6788fcc6ed2020e5e2531b7d5b`.

[58] G. L. Kane, W. W. Repko and W. B. Rolnick. ≪The effective W, Z0 approximation for high energy collisions.≫ *Physics Letters B*, **148(4-5)**:pp. 367 – 372, 1984. URL `http://www.sciencedirect.com/science/article/B6TVN-46YD1TS-7W/2/629ba9901c1d6da7e98b83e73e20f75b`.

[59] S. Dawson. ≪The effective W approximation.≫ *Nuclear Physics B*, **249(1)**:pp. 42 – 60, 1985. URL `http://www.sciencedirect.com/science/article/B6TVC-4C9YVN6-1S/2/ef554147d633ec61783243d6ffa52384`.

[60] J. Lindfors. ≪Distribution functions for heavy vector bosons inside colliding particle beams.≫ *Zeitschrift für Physik C Particles and Fields*, **28**:pp. 427–432, 1985. 10.1007/BF01413605, URL `http://dx.doi.org/10.1007/BF01413605`.

[61] J. F. Gunion, J. Kalinowski and A. Tofighi-Niaki. ≪Exact Calculation of $ff \to ffWW$ for the Charged-Current Sector and Comparison with the Effective-W Approximation.≫ *Phys. Rev. Lett.*, **57(19)**:pp. 2351–2354, 1986.

[62] R. Brun and F. Rademakers. ≪ROOT – An object oriented data analysis framework.≫ *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **389(1-2)**:pp. 81 – 86, 1997. New Computing Techniques in Physics Research V, URL `http://www.sciencedirect.com/science/article/B6TJM-3SPKX96-1F/2/3aa2b2cb72c9a4316a842802541bf317`.

[63] F. Pérez and B. E. Granger. ≪IPython: a System for Interactive Scientific Computing.≫ *Comput. Sci. Eng.*, **9(3)**:pp. 21–29, 2007. URL `http://ipython.scipy.org`.

[64] G. Rossum. ≪Python reference manual.≫ *Technical report*, Amsterdam, The Netherlands, The Netherlands, 1995.

[65] E. Boos et al. ≪Generic User Process Interface for Event Generators.≫ *ArXiv High Energy Physics - Phenomenology e-prints*, 2001. `arXiv:hep-ph/0109068`.

[66] J. Alwall et al. ≪A standard format for Les Houches event files.≫ *Comput. Phys. Commun.*, **176**:pp. 300–304, 2007. hep-ph/0609017.

[67] J. Pumplin et al. ≪New generation of parton distributions with uncertainties from global QCD analysis.≫ *JHEP*, **07**:p. 012, 2002. hep-ph/0201195.

[68] W. Giele et al. ≪The QCD / SM working group: Summary report.≫ 2002. hep-ph/0204316.

[69] M. R. Whalley, D. Bourilkov and R. C. Group. ≪The Les Houches Accord PDFs (LHAPDF) and Lhaglue.≫ 2005. hep-ph/0508110.

[70] D. Bourilkov, R. C. Group and M. R. Whalley. ≪LHAPDF: PDF use from the Tevatron to the LHC.≫ 2006. hep-ph/0605240.

[71] S. Jadach, J. H. Kuhn and Z. Was. ≪TAUOLA - a library of Monte Carlo programs to simulate decays of polarized [tau] leptons.≫ *Computer Physics Communications*, **64(2)**:pp. 275 – 299, 1991. URL http://www.sciencedirect.com/science/article/B6TJ5-46DF8RV-14/2/29fa22b8f712b31273284719b0dfbf8f.

[72] E. Barberio and Z. Was. ≪PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0.≫ *Computer Physics Communications*, **79(2)**:pp. 291 – 308, 1994. URL http://www.sciencedirect.com/science/article/B6TJ5-46FXC28-8W/2/1f6e33bc911f4a5e73297cc1a34fe3ad.

[73] S. Agostinelli et al. ≪G4–a simulation toolkit.≫ *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506(3)**:pp. 250 – 303, 2003. URL http://www.sciencedirect.com/science/article/B6TJM-48TJFY8-5/2/23ea98096ce11c1be446850c04cfa498.

[74] J. Allison et al. ≪Geant4 developments and applications.≫ *IEEE Transactions on Nuclear Science*, **53**:pp. 270 – 278, 2006.

[75] R. Bonciani et al. ≪NLL resummation of the heavy-quark hadroproduction cross-section.≫ *Nuclear Physics B*, **529(1-2)**:pp. 424 – 450, 1998. URL http://www.sciencedirect.com/science/article/B6TVC-3V7RK90-J/2/9aae964a16439c5dd8753252eaae78d6.

[76] S. Frixione and B. R. Webber. ≪Matching NLO QCD computations and parton shower simulations.≫ *JHEP*, **06**:p. 029, 2002. hep-ph/0204244.

[77] S. Frixione, P. Nason and B. R. Webber. ≪Matching NLO QCD and parton showers in heavy flavour production.≫ *JHEP*, **08**:p. 007, 2003. hep-ph/0305252.

[78] J. Campbell and F. Tramontano. ≪Next-to-leading order corrections to Wt production and decay.≫ *Nuclear Physics B*, **726(1-2)**:pp. 109 – 130, 2005. URL `http://www.sciencedirect.com/science/article/B6TVC-4H0J7H6-4/2/a911ef8c7dd544ffffd9b5b9109d3ae1`.

[79] B. P. Kersevan and E. Richter-Was. ≪The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5.≫ *ArXiv High Energy Physics - Phenomenology e-prints*, 2004. `arXiv:hep-ph/0405247`.

[80] M. L. Mangano et al. ≪ALPGEN, a generator for hard multiparton processes in hadronic collisions.≫ *Journal of High Energy Physics*, **7**:pp. 1–+, 2003. `arXiv:hep-ph/0206293`.

[81] S. Hoeche et al. ≪Matching Parton Showers and Matrix Elements.≫ *ArXiv High Energy Physics - Phenomenology e-prints*, 2006. `arXiv:hep-ph/0602031`.

[82] G. Corcella et al. ≪HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes).≫ *Journal of High Energy Physics*, **1**:pp. 10–+, 2001. `arXiv:hep-ph/0011363`.

[83] G. Corcella et al. ≪HERWIG 6.5 Release Note.≫ *ArXiv High Energy Physics - Phenomenology e-prints*, 2002. `arXiv:hep-ph/0210213`.

[84] P. Anger. To be published.

[85] C. M. Buttar, R. Harper and K. Jakobs. ≪Weak boson fusion $H \to WW(*) \to l + l - Pt$-miss as a search mode for an intermediate mass SM Higgs boson at ATLAS.≫ *Technical Report ATL-PHYS-2002-033. SHEP-HEP-2002-12*, Southampton Univ. Phys. Dept., Southampton, 2002.

[86] P. Bechtle et al. ≪Benchmark Analysis for Z to tau tau to lepton hadron with the First 100 pb−1.≫ *Technical Report ATL-PHYS-INT-2010-075*, CERN, Geneva, 2010.

[87] R. K. Ellis et al. ≪Higgs decay to [pi]+[pi]-: A possible signature of intermediate mass Higgs bosons at high energy hadron colliders.≫ *Nuclear Physics B*, **297(2)**:pp. 221 – 243, 1988. URL `http://www.sciencedirect.com/science/article/B6TVC-47199H0-8S/2/786af0185bff97884c5dd72f3609496e`.

[88] D. L. Rainwater. *Intermediate-mass Higgs searches in weak boson fusion*. Ph.D. thesis, THE UNIVERSITY OF WISCONSIN - MADISON, 1999.

[89] A. Hoecker et al. ≪TMVA - Toolkit for Multivariate Data Analysis.≫ *ArXiv Physics e-prints*, 2007. `arXiv:physics/0703039`.

[90] L. Breiman et al. *Classification and Regression Trees*. Wadsworth, 1984.

[91] J. R. Andersen and J. M. Smillie. ≪QCD and electroweak interference in Higgs production by gauge boson fusion.≫ *Phys. Rev. D*, **75(3)**:pp. 037301–+, 2007. `arXiv:hep-ph/0611281`.

[92] G. Cowan and E. Gross. ≪Discovery significance with statistical uncertainty in the background estimate.≫ *Technical report*, ATLAS Statistics Forum, 2008. URL `www.pp.rhul.ac.uk/~cowan/stat/notes/SigCalcNote.pdf`.

[93] J. Neyman. ≪Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability.≫ *Phil. Trans. Royal Soc. London*, pp. 333–380, 1937.

[94] G. J. Feldman and R. D. Cousins. ≪Unified approach to the classical statistical analysis of small signals.≫ *Phys. Rev. D*, **57**:pp. 3873–3889, 1998. `arXiv:physics/9711021`.

[95] S. S. Wilks. ≪The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses.≫ *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, **9**:pp. 60–62, 1938.

[96] W. Verkerke and D. Kirkby. ≪The RooFit toolkit for data modeling.≫ *ArXiv Physics e-prints*, 2003. `arXiv:physics/0306116`.

[97] R. Barlow. ≪Systematic Errors: facts and fictions.≫ *ArXiv High Energy Physics - Experiment e-prints*, 2002. `arXiv:hep-ex/0207026`.

[98] T. A. Collaboration. *ATLAS computing: Technical Design Report.* Number ATLAS-TDR-017 ; CERN-LHCC-2005-022 in Technical Design Report ATLAS. CERN, Geneva, 2005.

[99] L. Lyons. ≪Open statistical issues in Particle Physics.≫ *Ann. Appl. Stat*, **2(3)**:pp. 887–915, 2008.

Acknowledgments

Many people have supported me over the years and deserve my thanks, not all of whom I can possibly name in these brief paragraphs. My sincere apologies to anyone whom should I should have mentioned but failed to do so. As is custom in high energy physics I will use titles sparingly.

For giving me the opportunity to undertake this task, changing from one branch of physics to another in the process and for all the help and encouragement along the way, I wish to thank my supervisors Prof. Michael Kobel and Prof. Norbert Wermes.

As an experimentalist one always depends on the help of colleagues from the theoretical branch of particle physics. In Jürgen Reuter, Wolfgang Kilian and Ana Alboteanu I was lucky to find patient collaborators who were always willing to answer my questions about Whizard and phenomenology.

Both working groups I have become affiliated to during this work in Bonn and Dresden, I wish to thank for the professional and friendly atmosphere. Never have I failed to find someone willing to help search for the best answer to any problem. Wolfgang Mader in particular I would like to point out for the invaluable advice and discussions during the preparation of this work. To Gia Khoriauli, Nicolas Möser and Martin Schmitz I am indebted for much help in procuring my backgrounds.

To Duc Bao Ta and Nicolas Möser, the system administrators who worked with and took over from me and all those that followed, a cordial thank you for all the great work. While sometimes an administrator's job is behind the scenes I know very well how important it is.

The Friedrich-Ebert-Stiftung I thank for support, both financially and ideally.

Some people outside of the world of particle physics have been drawn into this unusual occupation of mine. My friends and family have taken it upon themselves to tolerate my quest for knowledge for knowledge's sake. Thank you for your indulgence.