

Study of Missing Transverse Energy in Heavy Flavor Decays



Diplomarbeit

VORGELEGT DER
FAKULTÄT FÜR PHYSIK
DER
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

VON
Jeanette Miriam Lorenz

MÜNCHEN, DEN 14. DEZEMBER 2010

Erstgutachterin: Prof. Dr. Dorothee Schaile
Zweitgutachter: Prof. Dr. Gerhard Buchalla

Abstract

The ATLAS experiment at the Large Hadron Collider (LHC) near Geneva aims at discovering the Higgs boson and looks for physics beyond the Standard Model. A promising theory of the physics beyond the Standard Model is Supersymmetry which introduces a symmetry between fermions and bosons. The background from various physical processes to SUSY signal regions must be understood well. In the estimation of the background data-driven methods are preferred.

In this thesis the QCD background in SUSY 1-lepton searches is studied with focus on the 1-muon channel. Special attention is paid to the background from QCD $b\bar{b}$ events. Only events with one isolated muon, multiple jets and a high Missing Transverse Energy can survive the selection cuts of this channel. Therefore, the QCD background is expected to be low in the 1-muon channel.

The estimation of the QCD $b\bar{b}$ background in the SUSY signal region is done in two steps: First, a QCD $b\bar{b}$ control sample is constructed by using non-isolated muons and two b-tagging cuts. Outgoing from this pure control sample, the second step uses the similarity of the distributions of the transverse momentum of the non-isolated muon and of the Missing Transverse Energy. The shape of the distribution of the Missing Transverse Energy can be predicted and so an estimation of the number of QCD $b\bar{b}$ events in the signal region obtained.

The ATLAS detector has recorded approximately an integrated luminosity of 40 pb^{-1} of proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$. The developed method was applied to a part of this recorded data. A good agreement between data and Monte Carlo is found.

In addition, a matrix method which aims at estimating the QCD background in the signal region is developed and tested with data.

Both methods predict a tiny QCD background in the SUSY signal region.

Zusammenfassung

Das ATLAS Experiment am LHC nahe Genf möchte insbesondere das Higgs-Boson als letztes noch fehlende Teilchen des Standard Modells der Teilchenphysik finden. Aber auch nach neuer Physik jenseits des Standard Modells wird gesucht. Eine vielversprechende Theorie hierbei ist die Supersymmetrie, die eine Symmetrie zwischen Fermionen und Bosonen annimmt. Jedoch muss der Untergrund von anderen, bekannten physikalischen Prozessen bestimmt werden. Hierbei sind Methoden, die komplett auf Daten beruhen und keine Informationen aus Monte Carlo Samples benötigen, vorzuziehen. In dieser Arbeit wird der QCD Untergrund im 1-Myon Kanal in SUSY Suchen näher untersucht, wobei besonderes Augenmerk auf den durch QCD $b\bar{b}$ Ereignissen hervorgerufenen Untergrund gelegt wird. Da sich eine typische SUSY Signatur in diesem Kanal durch fehlende transversale Energie und mehrere Jets und ein isoliertes Myon auszeichnet, sollte der QCD Untergrund in der SUSY Signal Region vernachlässigbar klein sein.

Der QCD $b\bar{b}$ Untergrund wird in zwei Schritten bestimmt. Der erste Schritt besteht aus der Konstruktion eines nahezu reinen QCD $b\bar{b}$ Control Samples. Dies wird durch die Verwendung von einem nicht-isolierten Myon und dem Einsatz zweier b-tagging Schnitte erreicht. Da sich die Verteilung des transversalen Impulses des nicht-isolierten Myons als ähnlich zu der Verteilung der transversalen fehlenden Energie erweist, ist es möglich mit Hilfe des transversalen Impulses des Myons den Schwanz der Verteilung der fehlenden transversalen Energie zu approximieren. Daraus kann eine Abschätzung für die Signal Region gewonnen werden. Die Methode wird auf Daten angewandt, wobei nur ein Teil der 40 pb^{-1} Daten bei einer Schwerpunktsenergie von 7 TeV verwendet wird. Monte Carlo und Daten zeigen in dieser Methode eine gute Übereinstimmung.

Eine Alternative zu dieser Methode, die allerdings den gesamten QCD Untergrund in der Signal Region bestimmt, ist eine Matrix Methode. Diese benötigt zwei Gruppen von unterschiedlichen Objekt Definitionen.

Schlussendlich aber sagen beide Methoden einen verschwindenden QCD Untergrund voraus.

Contents

1	Introduction	5
2	The Standard Model and SUSY: An overview	7
2.1	The Standard Model of Particle Physics (SM)	7
2.2	Deficiencies of the SM - or hints for SUSY	9
2.3	A brief introduction to SUSY	9
2.4	Monte Carlo event generation	12
3	The LHC and ATLAS	15
3.1	The LHC	15
3.2	ATLAS	18
3.3	Event reconstruction and computing	22
4	Object definitions	24
4.1	Electrons	24
4.1.1	Electrons in SUSY analysis	25
4.2	Muons	25
4.2.1	Reconstruction and Identification of muons	25
4.2.2	Object definition in SUSY analysis	26
4.3	Jets	26
4.3.1	Reconstruction of jets	27
4.3.2	Bad jets	27
4.3.3	Jet Energy Scale Calibration (JES)	28
4.3.4	SUSY selection cuts	28
4.4	Missing Transverse Energy (MET)	28
4.4.1	Calculation of MET	29
4.4.2	SUSY	30
5	SUSY search in the 1-lepton channel	31
6	Construction of a QCD $b\bar{b}$ control sample in the 1 lepton channel	34
6.1	b-tagging – principle and efficiency	34
6.1.1	The SV0-algorithm	35
6.1.2	Efficiency and Rejection	35
6.2	First studies with 10 TeV MC samples	38
6.2.1	Applying a SV0-cut	38
6.2.2	Using the isolation	39
6.2.3	Applying both cuts	41
6.2.4	Summary: All cuts used to create a pure QCD $b\bar{b}$ control sample	43
6.3	A center-of-mass energy of 7 TeV	46

7	Estimation of the QCD $b\bar{b}$ background in the signal region	49
7.1	The agreement in their shapes of the distributions of p_T (muon) and MET	50
7.2	Definition of control and signal regions	57
7.2.1	QCD $b\bar{b}$	58
7.2.2	Other backgrounds	59
7.3	The angle between a true muon neutrino and the closest reconstructed jet . .	59
7.4	Suppression of other backgrounds	64
7.4.1	Lowering the p_T and the MET cut	64
7.4.2	Estimation of the QCD $b\bar{b}$ events in the signal region (method 1) . . .	65
7.4.3	An additional b-tagging cut in region A (method 2)	66
7.5	Summary	67
8	QCD $b\bar{b}$ background estimation in data	69
8.1	Analysing data	69
8.2	QCD $b\bar{b}$ control sample in 1 muon + 3 jets channel	70
8.3	QCD $b\bar{b}$ background in data	72
8.3.1	Method 1 (without b-tagging for isolated muons):	72
8.3.2	Method 2 (with b-tagging for isolated muons), period A - F	74
8.4	Update of the jet cuts	75
8.5	Outlook	80
9	An alternative method in order to estimate the QCD background in the 1-muon channel	81
9.1	QCD estimation with the matrix method	82
9.2	Check matrix method with MC samples	84
9.2.1	Determination of N^{tight} and N^{loose}	84
9.2.2	Determination of ϵ_{QCD}	87
9.2.3	Estimation of the number of QCD events in the W control region and in the signal region	88
9.3	Applying the matrix method to data	92
10	Summary	95
A	Monte Carlo samples	97
A.1	QCD $b\bar{b}$	97
A.2	QCD udsc	98
A.3	QCD dijet	99
A.4	$t\bar{t}$	99
A.5	W+jets	99
A.6	W+bb	99
A.7	Z+jets	100
B	Quality criteria for tracks used in SV0-algorithm	101
C	Event displays in the 1-Lepton + 3 jet channel	102

Chapter 1

Introduction

The Standard Model of Particle Physics describes all matter and their weak, electromagnetic and strong interactions surprisingly well. Only the gravitation is not included in this theory. However, at energies near the Planck scale the strength of gravitational force becomes comparable with electromagnetic force. So it is expected, that the Standard Model is not a final theory. Indeed, it bears some deficiencies and blemishes as the hierarchy problem related to the Higgs mass. The Higgs boson itself is the last missing particle of the Standard Model. It is looked for extensively at current experiments. Also, it is known that there is some *Dark Matter* and *Dark Energy* in the universe, whose nature are completely unknown and not describable with the Standard Model.

A promising theory beyond the Standard Model is Supersymmetry, which introduces a symmetry between fermions and bosons. A brief introduction hereof is given in chapter 2. This theory is able to solve some of the problems of the Standard Model. In particular, depending on the kind of supersymmetric theory, various candidate particles are proposed as explanation for the Dark Matter.

In order to answer most of the burning questions in particle physics as for example “What is the nature of Dark Matter?”, “How does matter obtain mass?”, “Do extra dimensions exist?” and “Why is there not more antimatter?”, the Large Hadron Collider was built near Geneva. This proton-proton collider reached successfully in spring of 2010 a center-of-mass energy of 7 TeV. This energy allows the experiments in the collider ring to already probe the frontiers of the Standard Model. ATLAS is one of the two general purpose experiments which aim to answer the questions mentioned. ATLAS, as well as the LHC, are briefly described in chapter 3.

However, before these questions can be attacked, it is essential to understand the expected Standard Model properties. In the case of Supersymmetry, many, often rather advanced, proposals have been studied in the past on how to discover supersymmetric events. Such events usually contain a large Missing Transverse Energy. Searches are carried out in various channels. A typical search strategy in the 1-lepton-channel is outlined in chapter 5. Searches in the 1-lepton channel are optimized for SUSY signatures with one isolated lepton, multiple jets and a high Missing Transverse Energy. Various studies in the past for a center of mass energy of 10 TeV have already shown, that QCD events should be well suppressed after applying the optimized 1-lepton search strategy. However, the LHC is running at 7 TeV in contrast to the center-of-mass energy studied. The QCD background should be estimated again with the modified conditions. Hereby, data-driven methods are preferred, as the predictions should be as independent from Monte Carlo simulations as possible.

This work will take special care of the QCD $b\bar{b}$ background. Neutrinos of the decay of bottom quarks (or B hadrons) can result in a considerable amount of Missing Transverse Energy in QCD $b\bar{b}$ events, as neutrinos escape the detector undetected. So, some QCD $b\bar{b}$ events will survive the SUSY selection cuts. A very pure QCD $b\bar{b}$ control sample is constructed

by exploiting the special characteristics of QCD $b\bar{b}$ events in chapter 6. With this control sample and some modifications it is possible to predict the number of QCD $b\bar{b}$ events in the SUSY signal region as defined in chapter 5 in the 1-muon channel. The treatment of data is in some sense much more complicated than working simply with Monte Carlo simulations, as the real and not ideal detector and various perturbations like cosmic muons must be taken into account. Considering these, the method to estimate the QCD $b\bar{b}$ background is applied to data, which was taken between April and August of 2010, in chapter 8. Although the developed method is found to work quite well, it bears some problems, too. Therefore, an additional method is studied in chapter 9. This method, known as matrix method, has been used successfully for estimating the QCD background to other physical processes than SUSY. Here, it has the additional advantage of being a method which estimates the whole QCD background and not only the QCD $b\bar{b}$ background.

Chapter 2

The Standard Model and SUSY: An overview

The Standard Model of Particle Physics (SM) augmented by neutrino masses describes with astonishing precision all known phenomena. The SM describes all known fundamental fermions (which are the constituents of ordinary matter) and their electromagnetic, weak and strong interactions. These interactions are mediated by gauge bosons. However, the SM bears some deficiencies, like the hierarchy problem, which are outlined below. A very promising and advanced theory beyond the Standard Model is Supersymmetry (SUSY), which is capable to solve many deficiencies of the Standard Model. This theory is outlined briefly after a small introduction about the Standard Model.

2.1 The Standard Model of Particle Physics (SM)

It is impossible to explain all the aspects of the Standard Model of Particle Physics on a few pages. Therefore, only some few points will be highlighted. Detailed descriptions can be found for example in [2] and in [3]. A brief summary can be found in [4], too.

The SM counts in total six leptons (and their antiparticles). These appear in three families or generations. In addition, six quarks (and their antiparticles) are known, which can be ordered in three generations, too (table 2.1). Why there are (as known so far) exactly three generations is an unanswered question in the SM. Leptons and quarks bear spin 1/2 and are hence fermions. The interactions between these fermions are mediated by gauge bosons with spin 1 (these are summarized in table 2.2). The electromagnetic force is mediated by the photon γ , the weak force by W^\pm and Z bosons. The strong interaction is mediated by 8 gluons. In addition, a scalar boson, the Higgs boson, is needed (this will be explained in the following). This Higgs boson has not been discovered yet; its discovery is one of the main challenges of the two general purpose experiments at the proton-proton collider LHC near Geneva.

fermion	generation			electric charge	color	weak isospin		spin
	1	2	3			left handed	right handed	
leptons	ν_e	ν_μ	ν_τ	0	-	1/2	-	1/2
	e	μ	τ	-1	-	1/2	0	1/2
quarks	u	c	t	+2/3	r, b, g	1/2	0	1/2
	d	s	b	-1/3	r, b, g	1/2	0	1/2

Table 2.1: The fermions in the Standard Model. [5]

interaction	couples to	mediator	mass (GeV)
strong	color	8 gluons (g)	0
electromagnetic	electric charge	photon γ	0
weak	weak charge	W^\pm, Z	$m_{W^\pm} = 80.399 \pm 0.023 \text{ GeV}$ $m_Z = 91.1876 \pm 0.0021$ [7]

Table 2.2: Interactions and their mediators in the Standard Model. [5]

Theoretically, the SM is described by two renormalizable Quantum Field Theories (QFT). Hereby, the strong interaction is described by quantum chromodynamics. The electroweak part is described by the unification of two theories, the quantum electrodynamics and the Glashow-Salam-Weinberg gauge theory of weak interactions.

A QFT is formulated by using a Lagrangian, which contains the dynamics of the system in question. The equations of motion of the particular system are obtained by using the Least Action Principle. Quantities are usually calculated with the help of perturbation theory.

Both the electroweak theory as well as QCD rely on symmetry groups (they are Yang-Mills theories). Indeed, the combined symmetry group of the Standard Model is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ ($SU(3)$ is the symmetry group of QCD, $SU(2) \otimes U(1)$ the symmetry group of electroweak theory). As gauge theory, the basic equations of the theory need to be invariant under local gauge transformations. This requirement leads to the introduction of gauge fields. For example, in the case of $U(1)_{em}$ (the electromagnetic part of the electroweak theory) this field transforms as asked by Maxwell's equations and so describes a massless photon field with spin 1. In the electroweak theory, $SU(2) \otimes U(1)$, the emerging gauge fields result in a triplet of W bosons (W^-, W^+, W^0) and a neutral B boson. The W^0 and the B boson mix to the Z boson and to the photon, γ . In QCD, eight gauge fields are needed: the gluons, which are massless and colored bosons with spin 1.

A serious problem for the theory is the fact that fermions as well as the W and Z gauge bosons are massive. But any mass term in the Lagrangian of $SU(2)_L$ will destroy its symmetry. This problem is solved by the Higgs mechanism. Here the gauge invariance is spontaneously broken by introducing a complex scalar $SU(2)_L$ doublet ϕ with a non-zero vacuum expectation value. This scalar field couples to fermions and W^\pm, Z bosons. Due to the non-zero vacuum expectation value, W^\pm, Z and fermions obtain a mass, but not the photon. A massive and scalar boson, the Higgs boson, also appears. Although the vacuum expectation value of the scalar field can be calculated to $v = 246 \text{ GeV}$ by its coupling to the known W -mass, the Higgs mass itself is a free parameter. The Higgs boson has not been discovered so far. Constraints for the Higgs mass were obtained for example by the e^+e^- collider LEP (1989-2000). They are at the moment $114.4 \text{ GeV} < m_H < 186 \text{ GeV}$ (95 % C.L.) (lower bound from LEP, upper bound from indirect constraints, latest Tevatron results not included). Currently, it is searched for at Tevatron and - since this year - also at the LHC. Its discovery would complete the Standard Model, as it is the only missing particle.

The QCD part of the Standard Model has two important features: *Confinement* describes the behavior of the strong force between colored particles, which increases very fast when trying to separate two colored particles. However, at very small distances or at high energies the interaction between two colored particles becomes weak. This behavior is called *asymptotic freedom*. Important consequences are, that perturbation theory can be used at high energies, but not at low energies, and that no quark or gluon can exist free. Instead, they undergo hadronization where the colored objects fragment into color singlets (so color neutral). These color neutral objects are mesons or hadrons and consist of an antiquark and a quark or of three quarks, respectively. In particular, a parton from a hard interaction will hadronize after undergoing soft and collinear showering. This results in collimated bunches

which are referred to as jets. The SM (with massless neutrinos) counts 19 arbitrary parameters [7]: Three gauge couplings, g_3, g, g' ; 9 charged fermion masses; 4 mixing angles in the CKM matrix; v , the Higgs vacuum expectation value; λ , the quartic coupling related to the Higgs; the strong CP violation parameter θ . These arbitrary parameters have to be taken from measurements.

2.2 Deficiencies of the SM - or hints for SUSY

Although the SM, as presented above, works quite well, it also has some serious problems and blemishes.

- At the reduced Planck scale $M_P = (8\pi G_{Newton})^{-1/2} = 2.4 \cdot 10^{18} \text{ GeV}$ the gravitational force gets comparable to the electromagnetic force between two charged particles. Additional theories beyond the SM become necessary.
- The inverse gauge couplings $\alpha_1^{-1}(Q^2)$, $\alpha_2^{-1}(Q^2)$ and $\alpha_3^{-1}(Q^2)$ of the SM run linearly with $\ln Q^2$ at one loop order. In the SM, the gauge couplings approach each other at high energies, but do not actually meet. In contrast to the MSSM (presented below), where they actually meet at $Q^2 \approx (10^{16} \text{ GeV})^2$. This is illustrated in figure 2.1.
- Gravitation is not included in the SM.
- Astronomical observations have shown that the major part of matter in our universe does not consist of known matter described by the SM, but of Dark Matter. Its nature is completely unknown.
- *Hierarchy problem or fine-tuning problem:* The Higgs mass suffers from loop corrections from all particles that couple to the Higgs field. The correction to the Higgs mass is:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 + \dots \quad (2.1)$$

Λ_{UV} is the ultraviolet cutoff and is often assumed to be on the order of M_P . λ_f depends on the SM coupling to the Higgs field. These corrections result in a Higgs mass many orders of magnitude larger than the electroweak scale. However, the Higgs mass should be around the electroweak scale for consistency reasons.

2.3 A brief introduction to SUSY

The above mentioned hierarchy problem is solved beautifully by introducing a symmetry between fermions and bosons. Then the loop contributions to the Higgs mass from bosons and fermions bear a relative minus sign. The contribution of the Λ_{UV}^2 term is canceled. Schematically, the symmetry between fermions and bosons can be written as following:

$$Q|Boson\rangle = |Fermion\rangle \quad (2.2)$$

$$Q|Fermion\rangle = |Boson\rangle \quad (2.3)$$

Q generates the transformations and carries spin angular momentum 1/2, as Q is a fermionic operator. The same is true for Q^\dagger . Supersymmetry must be a spacetime theory [8]. As the Haag-Lopuszanski-Sohnius extension of the Coleman-Mandula theorem restricts

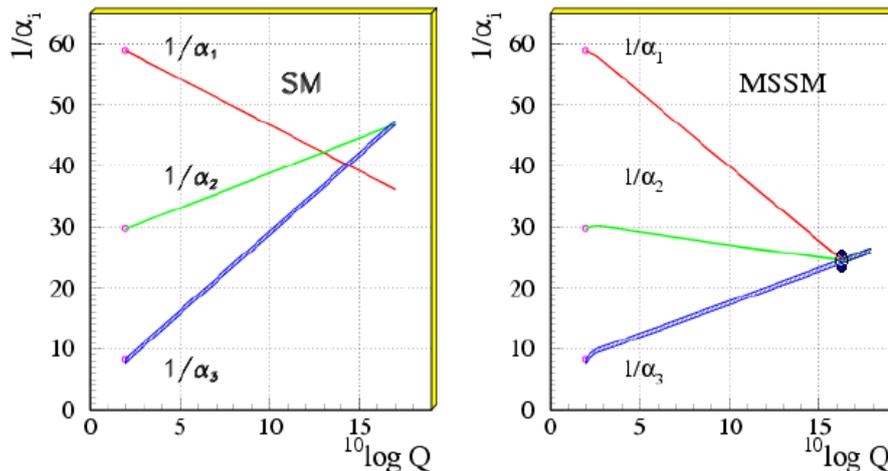


Figure 2.1: The inverse coupling constants do not meet in the Standard Model at high Q values (left figure), whereas they meet in the MSSM (right figure).

the form of such theories [8], the operators Q and Q^\dagger have to satisfy this schematic form of commutation and anticommutation relations [10]:

$$\{Q, Q^\dagger\} = P^\mu \quad (2.4)$$

$$\{Q, Q\} = \{Q^\dagger, Q^\dagger\} = 0 \quad (2.5)$$

$$[P^\mu, Q] = [P^\mu, Q^\dagger] = 0 \quad (2.6)$$

$$[Q, \text{generator of gauge transformations}] = 0 \quad (2.7)$$

Equation 2.4 shows that two SUSY transformations executed after each other will result in a transformation in spacetime. Therefore, a connection between SUSY and General Relativity is expected. A fermion and a boson which are related by equation 2.2 are referred to as superpartners. Equation 2.7 indicates that the superpartners have the same quantum numbers. Equation 2.6 indicates that superpartners have the same mass, because P^2 commutes with the operators Q and Q^\dagger . As no superpartner to any known fermion or boson has been observed so far, SUSY must be broken. However, this breaking can be only soft, because otherwise quadratic corrections to the Higgs mass would be created again. Also, the theory has to stay renormalizable. This results in constraints on the dimensions of the terms breaking Supersymmetry. It can be shown [8] that a supersymmetric theory has the same number of fermionic and bosonic degrees of freedom. The superpartners of leptons or of quarks are called sleptons and squarks, respectively, and the superpartners of the gauge bosons are indicated with *-ino* at the end. The supersymmetric model with the smallest additional number of particles and interactions is called *Minimal Supersymmetric Standard Model (MSSM)*. Some important points are summarized without going into the details (these can be found in [8] or in [7] for example):

- Each known particle has a superpartner with spin differing by 1/2 from the particle.
- All superpartners of particles of the Standard Model are really new particles and no superpartner can be identified with any SM state.
- The MSSM knows about five physical spin-zero particles in the Higgs sector: A charged Higgs boson pair (H^\pm), two CP-even neutral Higgs bosons (h^0, H^0) and a CP-odd neutral Higgs boson (A^0). [7]

Field Content of the MSSM					
Super-Multiplets	Boson Fields	Fermionic Partners	SU(3)	SU(2)	U(1)
gluon/gluino	g	\tilde{g}	8	1	0
gauge/	W^\pm, W^0	$\tilde{W}^\pm, \tilde{W}^0$	1	3	0
gaugino	B	\tilde{B}	1	1	0
slepton/	$(\tilde{\nu}, \tilde{e}^-)_L$	$(\nu, e^-)_L$	1	2	-1
lepton	\tilde{e}^-_R	e^-_R	1	1	-2
squark/	$(\tilde{u}_L, \tilde{d}_L)$	$(u, d)_L$	3	2	1/3
quark	\tilde{u}_R	u_R	3	1	4/3
	\tilde{d}_R	d_R	3	1	-2/3
Higgs/	(H_d^0, H_d^-)	$(\tilde{H}_d^0, \tilde{H}_d^-)$	1	2	-1
higgsino	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^0)$	1	2	1

Figure 2.2: Particle content of the MSSM. [7]

- The charged gauginos (\tilde{W}^\pm) and the charged higgsinos (H_u^+ and H_d^-) mix to charginos ($\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^\pm$).
- The neutral gauginos (\tilde{B} and \tilde{W}^0) and the neutral higgsinos (\tilde{H}_d^0 and \tilde{H}_u^0) mix to neutralinos ($\tilde{\chi}_i^0$ with $i = 1, \dots, 4$).
- The MSSM possesses 124 independent parameters, hereof 105 are new compared to the SM.

All particles of the MSSM are summarized in table 2.2.

Concerning the SUSY breaking mechanism, a *hidden* sector is usually introduced with particles which are completely neutral with respect to the SM gauge group. All particles of the MSSM are in the *visible* sector. Supersymmetry breaking takes place in the hidden sector and its effects are transmitted to the visible sector by some unknown mechanism. A common model, *mSUGRA*, assumes that the symmetry breaking is mediated by effects of gravitational strength. It is possible to reduce the number of independent parameters from 105 to 5 in addition to the 19 free parameters of the Standard Model by simplifying assumptions [7]. A common choice is a universal scalar mass m_0 , a universal trilinear coupling A_0 , a universal gaugino mass $m_{1/2}$ (all defined at the scale of Grand Unification), $\tan \beta$ (ratio of the Higgs vacuum expectation values: $\frac{\langle H_u \rangle}{\langle H_d \rangle}$) and $\text{sgn}(\mu)$ (the sign of the supersymmetric Higgs-mixing-mass term).¹ In principle, in supersymmetric theories terms would be possible which violate the baryon number or lepton number (if this would be the case the proton could decay for example, which has never been observed). The *R-parity* is introduced in order not to allow such terms. It is defined as

$$P_R = (-1)^{3(B-L)+2S} \quad (2.8)$$

B is the baryon number, L the lepton number, S the spin. All SM's particles have R-parity 1, whereas all SUSY particles have R-parity -1. This fact has important consequences:

- SUSY particles can be produced only in pairs and at each vertex an even number of SUSY particles is allowed only.

¹ In this work the SU4 point, a low mass point close to Tevatron bound, is used for Monte Carlo studies. It is defined as $m_0 = 200 \text{ GeV}$, $A_0 = -300 \text{ GeV}$, $m_{1/2} = 160 \text{ GeV}$, $\tan \beta = 10$ and $\mu > 0$

- Therefore, the Lightest Supersymmetric Particle (LSP) needs to be stable. It must be neutral and colorless because otherwise it would have been observed in astrophysics. Therefore, it is a good candidate for the Dark Matter.
- A typical SUSY signature, as assumed in this work, should have a high Missing Transverse Energy, because the LSP carries away energy (because it does not interact with the detector). In addition, a typical signature should contain jets and lepton(s): Gluinos and/or squarks can be produced in a collision of partons. They can decay via $\tilde{g} \rightarrow q\tilde{q}$, $\tilde{g} \rightarrow q\tilde{q}\tilde{\chi}_i^0$, $\tilde{g} \rightarrow q\tilde{q}'\tilde{\chi}_j^\pm$ and $\tilde{q} \rightarrow q\tilde{g}$, $q\tilde{\chi}_i^0$, $\tilde{q}'\tilde{\chi}_j^\pm$, into neutralinos and charginos. The charginos can decay to lepton and corresponding neutrino and the LSP (in this model $\tilde{\chi}_1^0$).

2.4 Monte Carlo event generation

The outcome of parton-parton collisions possess a complex structure. In particular, hadronization and soft radiation can not be calculated exactly. Therefore, events have to be simulated by Monte Carlo event generators, where various parameters have to be set or a physical model has to be assumed. In general, Monte Carlo event generators simulate hypothetical events whose distributions agree with the prediction of the theory. Various Monte-Carlo event generators are available. In this work events simulated by PHYTHIA [11] and AlpGen [12] are mainly used. A short and comprehensive overview about all frequently used generators is given in [17]. Events can be filtered for specific physical processes at generator level. The interaction of particles with the detector (described in the next chapter) is simulated by Geant4 [18], which simulates the hits in the detector. These hits are transformed into Raw Data Objects (RDQ) [14] by *digitization* (this step includes the detector response in form of voltage and current at the readout channels). RDQs are the input to a reprocessing chain, which will be explained briefly in the next chapter. In contrast to real data, events generated by Monte Carlo generator can be accompanied by a *Truth* information. This Truth information of the event contains all particles in the event with their true (kinematical and spatial) properties. This information is useful as in the interaction of particles with a real detector some misidentifications or mismeasurements of particles can happen and these effects should be understood.

Apart from a Monte Carlo dataset containing supersymmetric events (the SU4 point assumed), various background - in particular QCD - Monte Carlo datasets were used. These background Monte Carlo datasets and their physical processes are summarized briefly in the following.

The cross sections and generator efficiencies for the various Monte Carlo event samples used can be found in the appendix together with the number of simulated events. The various Monte Carlo samples often need to be combined. This is done according to their integrated luminosity. The integrated luminosity (L) is connected to the cross section σ and number of events in the Monte Carlo sample (N) by $L = \frac{N}{\sigma}$.²

Background processes:

- **QCD:** The QCD background includes all final states of parton-parton collisions which contain only light quarks (as u , d and s), charm and bottom quarks and their reactions or decays (so processes with the top quark are not included).

Different Monte Carlo samples for different generators are available and used in this work in the case of QCD processes. *AlpGen* simulates tree-level matrix elements with

²The generator efficiency must be considered as well in the calculation of the integrated luminosity for some Monte Carlo samples and is multiplied with the cross section.

a defined number of final hard particles at lowest order of perturbation theory. The showering and hadronization of quarks and gluons must be provided by another generator (usually HERWIG or PYTHIA) [17]. Light quarks as u , d , s and c are usually assumed to be massless, whereas the mass of the bottom or the top quark is included. In the case of QCD two different sets of Monte Carlo samples exist: One that contains only light quarks (u , d , s and c) and is referred to as **QCD udsc** or QCD light flavors in this work. The other set contains exclusively $b\bar{b}$ processes and is referred to as **QCD $b\bar{b}$** .

$t\bar{t}$ processes are treated separately and not as QCD events (see below). As the cross section for multijet production is rather large, a p_T slicing is applied. Hereby the p_T of the leading jet is used and Monte Carlo samples are generated for a specific range of p_T of the leading jet. Thanks to this slicing it is possible to generate a sufficient statistic for events with high p_T jets (the cross section falls with increasing jet p_T) without generating to many events with low p_T jets. As the number of events in the slices for low p_T jets is still too big, an additional filter is applied to these samples. Common approaches are to apply a filter on the jets (like asking for at least three jets with $p_T > 25 \text{ GeV}$ within $|\eta| < 2.8$ and the leading jet must satisfy $p_T > 60 \text{ GeV}$) or to apply a filter on the muons (at least one muon with $p_T > 10 \text{ GeV}$ within $|\eta| < 2.8$). In the Monte Carlo QCD samples at a center-of-mass energy of 7 TeV only the second approach was used. Therefore, in the main part of this work muon filtered QCD samples were used. However, some early studies rely on Monte Carlo samples with a center-of-mass energy of 10 TeV with a jet filter applied. Later in this work a QCD $b\bar{b}$ control sample is constructed. Any cuts in the construction of this control samples must take into account that the different QCD Monte Carlo samples are combined. Therefore, some basic cuts of the control samples are already determined by the jet filter.

However, it is still difficult to provide enough statistics in the tail of some distributions like the distribution of the Missing Transverse Energy. As this is an important distribution in particular for this work, Monte Carlo QCD samples generated by *PYTHIA* were used in addition to the AlpGen Monte Carlo samples and in the second part of this work as default (in particular in all estimations of the QCD background in the signal region).

The **PYTHIA** generator itself is a general purpose generator [17], which is able to simulate also the showering and hadronization in addition to the actual subprocess. Only subprocesses with two incoming and two outgoing particles are generated in the case of QCD events, all at leading order, although some next-to-leading order effects are included in the showering of the partons. PYTHIA is powerful in describing the decay processes of unstable particles (with the Lund string model [17]). However, in the case of QCD $b\bar{b}$ processes better models are available and these are combined with PYTHIA in the PYTHIAB generator. In this work general QCD dijet Monte Carlo samples with a muon filter (at least one muon with $p_T > 8 \text{ GeV}$ within $|\eta| < 2.5$) are used if general QCD events are studied. Two Monte Carlo samples generated by PYTHIAB (one with an electron filter and one with a muon filter) are used in the study of QCD $b\bar{b}$ events. Both PYTHIAB Monte Carlo samples were transformed into n-tuples from AOD format by the SUSYD3PDMaker 00-05-04 [16] for this work. The large size of both datasets can be seen from the fact, that it needed 3000 CPU hours to produce them (an effort which was only possible thanks to the Grid). The properties of QCD $b\bar{b}$ which are important for this work events are discussed in chapter 6.

The main background studied in this work consists of QCD $b\bar{b}$ events (or of QCD events). High energetic neutrinos from the decay of the bottom quark (or the B hadron) result in a large Missing Transverse Energy, as the neutrinos cannot be detected with the ATLAS

detector. Because of this large Missing Transverse energy, these events might pass the SUSY selection cuts as explained in chapter 5. Different backgrounds are referred to as *other backgrounds*. Hereby, an important background consists of $t\bar{t}$ events.

- **$t\bar{t}$:** Two Monte Carlo samples are used for $t\bar{t}$ events. One sample contains $t\bar{t}$ events where a top quark (anti-top quark) decays into a W boson and a bottom quark and the W boson decays into leptons. In the second sample this W boson decays to light quarks instead. Leptons (muons or electrons) from top decays are mainly well isolated. These events were generated by MC@NLO [13] which includes NLO corrections.
- **W+jets:** This background is composed of W bosons which decay either in e and ν_e , μ and ν_μ or in τ and ν_τ plus additional jets composed of light quarks. b -jets are treated separately in $Wb\bar{b}$ Monte Carlo samples. Leptons from W+jets are expected to be isolated. These events are generated by AlpGen.
- **Z+jets:** Here the Z boson decays to ee , $\mu\mu$, $\tau\tau$ or $\nu\nu$ and is accompanied by light jets. As this work concentrates on the 1-lepton channel, the Z+jets background should be negligible. (However it is possible, that one lepton is not reconstructed.) These events are again generated by AlpGen.

Chapter 3

The LHC and ATLAS

As it was outlined briefly in the last chapter, the last ingredient of the Standard Model of Particle Physics, the Higgs boson, has not been discovered yet, although it was looked for at LEP and at Tevatron. Further, there are hints that the Standard Model cannot be a final theory. A new proton-proton accelerator and collider has been built near Geneva at CERN, the LHC (**L**arge **H**adron **C**ollider) in order to discover the Higgs boson and to find physics beyond the Standard Model. The LHC is designed for a center-of-mass energy of $\sqrt{s} = 14$ TeV. It is expected that a Higgs boson - if it exists - should be discovered with this center of mass energy. Depending on the theoretical model, supersymmetric particles are also in reach. Other theories beyond the Standard Model, including extra dimension, will be tested as well. The experiments in the collider ring have been designed and optimized for these challenges.

After a problematic start in 2008, the LHC restarted successfully in late 2009 with a center-of-mass energy of 900 MeV. After further technical improvements, the LHC succeeded in reaching a center-of-mass energy of $\sqrt{s} = 7$ TeV in late March of 2010. Up to this very day, the LHC has recorded more than 40 pb^{-1} of integrated luminosity of proton-proton collisions, which are available for physics analysis and were partly used in this work. In the meantime, heavy-ion collisions have taken place as well and interesting results were obtained. Last summer the experiments showed at various conferences that they understand their detectors well and are in the best position to search for new phenomena.

In the following section a short overview of the LHC will be given. It will be followed by a more detailed overview of the ATLAS detector, which is a general purpose detector in the LHC ring.

3.1 The LHC

In 1994/96 it was decided to build the new proton-proton collider LHC in the existing collider tunnel of the LEP (electron-positron collider), which had been in operation since 1989 and was closed in 2000 to allow the construction of the LHC. The information in this section was mainly extracted from [20] and [21]. This existing LEP tunnel as well as the existing injection chain were strong arguments for building the LHC at CERN. This tunnel is 26.7 km long and 45 m up to 170 m under the surface with a inclination of 1.4 % towards the the Léman lake. It consists of 8 arcs and 8 straight sections. The straight sections were necessary for the LEP to compensate synchrotron radiation losses, although they are not wished for a proton-proton collider. The LEP tunnel was reused nevertheless due to cost savings. As a particle-particle collider two rings with opposite rotating beams were necessary for the LHC. As the LEP tunnel is rather small with a diameter of 3.7 m, it was not possible to install two proton rings in the LEP tunnel. Therefore the so-called twin-bore magnet design was used, which was proposed by John Blewett (Brookhaven laboratory). Eight interaction points (IP)

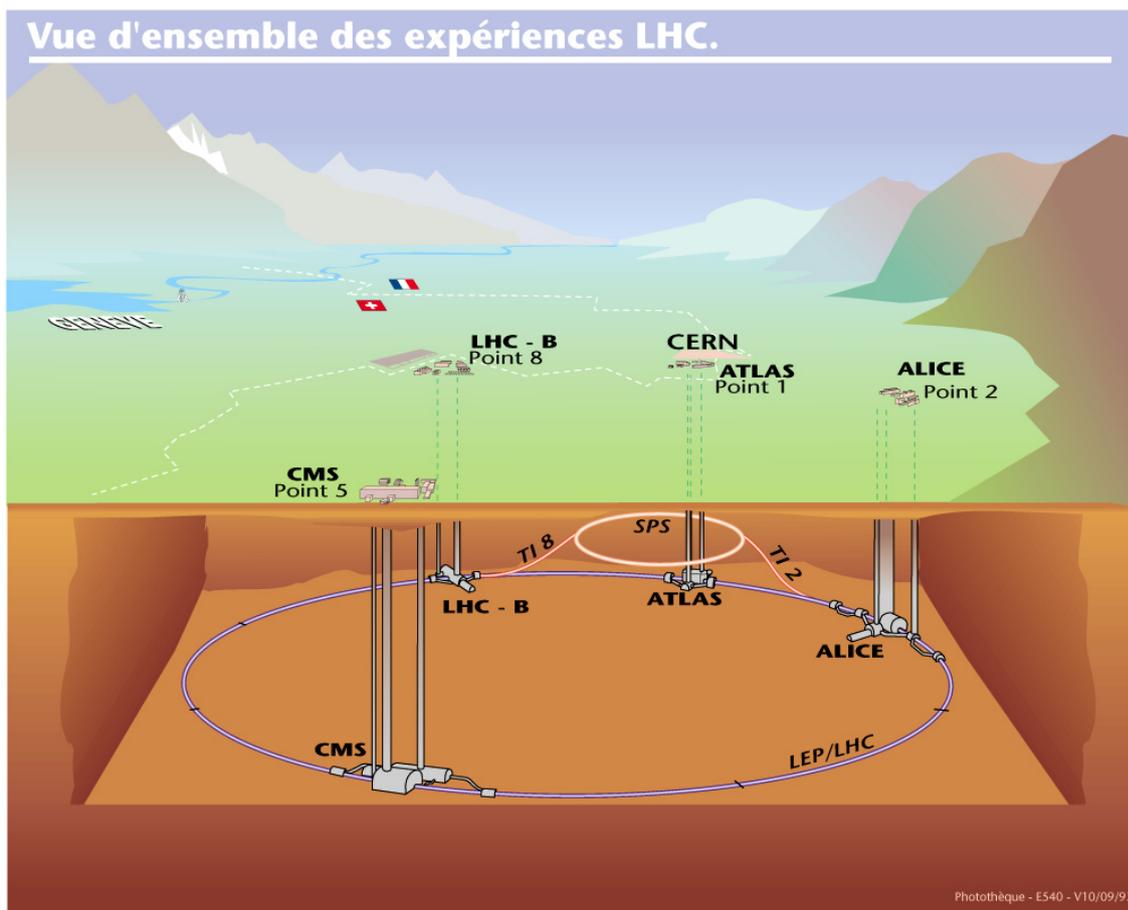


Figure 3.1: The LHC ring with its experiments is embedded between Léman Lake and the French Jura [48].

exist where the two proton beams could collide. But only four of them are used for collisions, where in the sum six different experiments are installed for different purposes:

The two general purpose detectors:

- ATLAS (**A** **T**orodial **L**H**C** **A**pparatu**S**)
- CMS (**C**ompact **M**uon **S**olenoid)

and the specialised detectors:

- LHCb (specialised to B-physics)
- ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment, a heavy-ion detector which shall study quark-gluon plasma)
- TOTEM (**T**OTAL **E**lastic and diffractive cross section **M**easurement, aimed at studying the proton-proton interaction cross section)
- LHCf (simulates cosmic rays by using very forward particles created inside the LHC)

The design parameters of the LHC are summarized in table 3.1

The underground and surface structures at Point 1 and 5 (ATLAS and CMS) are new, whereas existing structures could be used for ALICE and CMS at Point 2 and 8. The LHC

Beam injection energy (TeV)	0.45
Beam energy (TeV)	7.0
Number of particles per bunch	$1,15 \times 10^{11}$
Number of bunches per beam	2808
Max. stored beam energy (MJ)	362
Norm. transverse emittance ($\mu m rad$)	3.75
Colliding beam size (μm)	16
Bunch length at 7 TeV (cm)	7.55

Table 3.1: Nominal LHC parameters [21]

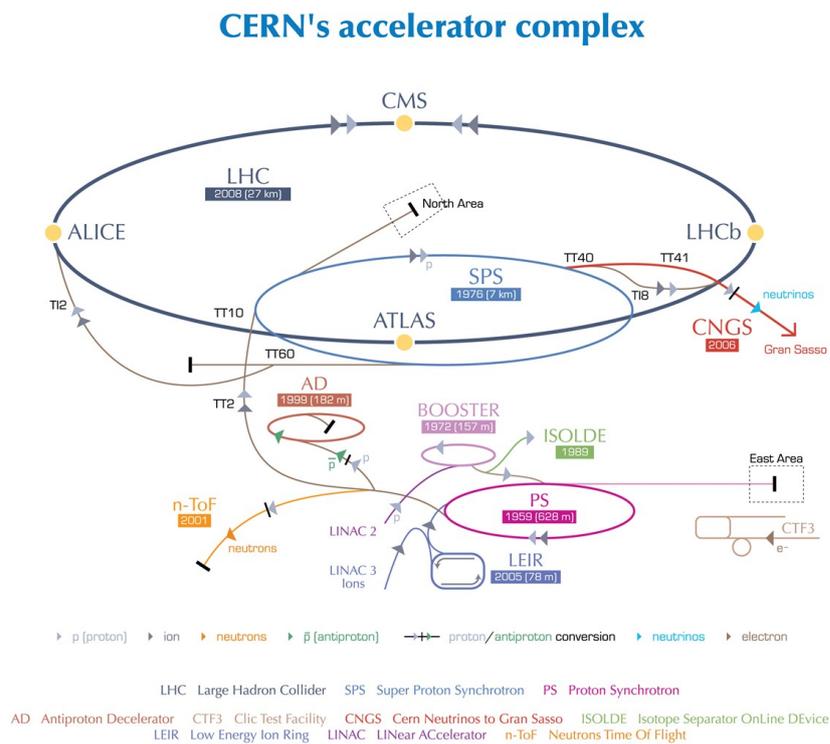


Figure 3.2: The LHC with its injection chain [48]

ring is connected with the CERN accelerator complex by two transfer tunnels. A schematic view of the whole accelerator complex including all pre-accelerators can be found in figure 3.2. It starts with a bottle of Hydrogen. In Linac 2 (a Alvarez's drift-tube) the protons are accelerated to 50 MeV and thereafter injected in the PS Booster. It increases the intensity ($> 10^{13}$ protons obtained) and accelerates to 1.4 GeV. The next step is the Proton Synchrotron (PS) which is followed by the Super-Proton Synchrotron (SPS) with a circumference of 6.9 km. Hereby, the protons are accelerated to 450 GeV and thereafter injected into the LHC. The procedure for heavy-ion beams (lead ions; with a maximal designed center-of-mass energy of 5.5 TeV and a designed luminosity of $10^{27} \frac{cm^2}{s}$) is similar, although the linear accelerator Lineac 3 is used as first step and is followed by LEIR (Low-energy ion ring). The PS Booster is not needed.

3.2 ATLAS

Both the physical processes to study as well as the high energy collisions are challenging tasks for a general purpose detector like ATLAS at Point 1¹. Apart from high precision measurements of the already discovered parts of the Standard Model (SM) including the study of the coupling and spin of the top quark, the ATLAS detector is optimized for the discovery and study of the Higgs Boson, the only missing part of the well established SM. As the Higgs Boson is produced and decays by various mechanisms depending on its mass, it is a challenging task for a detector to cover the whole mass range, so subdetectors have been optimised for this. Physics beyond the SM, like massive Gauge Bosons W' and Z' , mono-jets with high p_T , supersymmetric decays or exotic processes, is also looked for. The LHC will deliver bunches containing up to 10^{11} protons, which will collide 40 million times per second. A center-of-mass energy of 14 TeV in the proton collisions and a luminosity of $10^{34} \frac{cm^2}{s}$ are aimed at. This means for a detector, that it has to handle high interaction rates, radiation doses, particle multiplicities and particle energies. Therefore, the detector has to fulfill the following conditions:

- Fast and radiation-hard electronics and sensors
- High granularity
- Large pseudorapidity (explained below) range
- Very good momentum resolution of charged particles and high reconstruction efficiencies in the Inner Detector (ID); a vertex detector is needed around the interaction region in order to detect secondary vertices
- An electromagnetic calorimeter for the electron and photon identification
- A hadronic calorimeter for precise Jet and Missing Transverse Energy (MET, explained later) measurements
- Muon identification and a very good resolution of high p_T muons

Before this work summarizes how the ATLAS detector satisfies these conditions, some basic naming convention are explained which are used in the following.

Naming conventions and definitions: The origin of the ATLAS coordinate system is the nominal interaction point. The z-axis is defined in beam direction, whereas the x-y plane is transverse to the beam. All transverse quantities refer to this plane. The positive x-axis is in the direction of the middle of the LHC ring and the positive y-axis points upwards. The

¹The information in this section is taken mainly from [22]

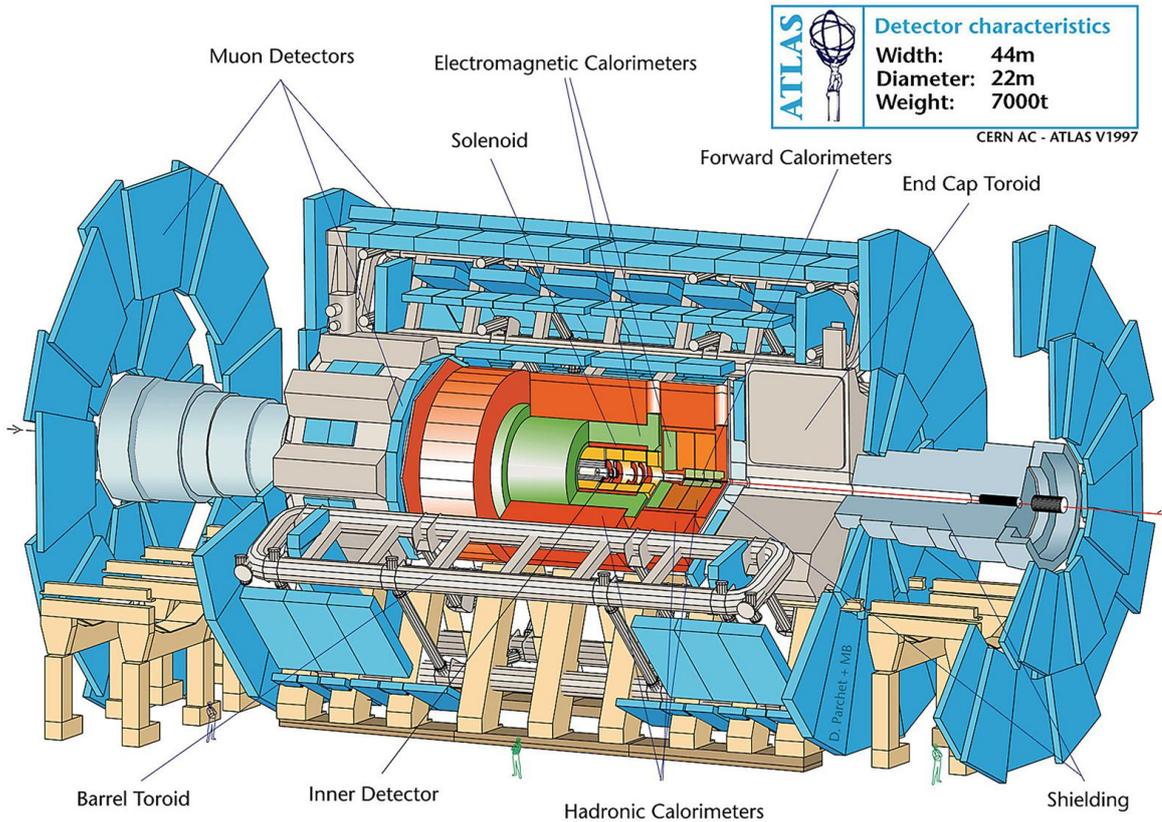


Figure 3.3: The layout and composition of the ATLAS detectors. [48]

ATLAS detector has an A-side in positive z direction and a C-side in negative z direction. The angle ϕ is measured around the beam axis ($\phi = 0$ corresponds to the positive x -axis; the angle increases clockwise if looking in direction of the beam axis) and the polar angle θ is measured around the x -axis with ($\theta = 0$ is in the positive z direction and the angle increases clockwise if looking into the negative x -direction). The pseudorapidity is defined as $\eta = -\ln \tan(\frac{\theta}{2})$. The distance of objects to each other is usually expressed in $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The layout of the ATLAS subdetectors: The ATLAS detector consists of in total 4 groups of subdetector systems and a system of magnets. An overview of the ATLAS detectors is shown in figure 3.3. The ATLAS subdetectors are ordered in an onion-like structure around the beam line. The Inner Detector (ID) is placed directly around the beam and is itself surrounded by a thin superconducting solenoid magnet (with a magnetic field of 2 T, 5.3 m long, diameter of 2.5 m). First an electromagnetic and then a hadronic calorimeter surround the ID. Three magnet systems composed of superconducting toroids (in the barrel and in the two endcaps), enclose the calorimeters. The outer part consists of a muon spectrometer. The same sequence is used in the two endcaps.

Inner Detector: A high track density of approximately 1000 particles emerging from the interaction point every 25 ns within $|\eta| < 2.5$ has to be handled by the Inner Detector. For the detection of interesting physical processes a high momentum resolution as well as a high vertex resolution (of primary and secondary vertices) is necessary. To fulfill these requirements, the Inner Detector consists of three parts (a more detailed cut-away view is given in figure 3.4): Its innermost part consists of a pixel detector ($|\eta| < 2.5$), where the pixel sensors have a size of $R \cdot \phi \times z = 50 \times 400 \mu\text{m}^2$. The accuracy is $10 \mu\text{m}$ in $R \cdot \phi$ and

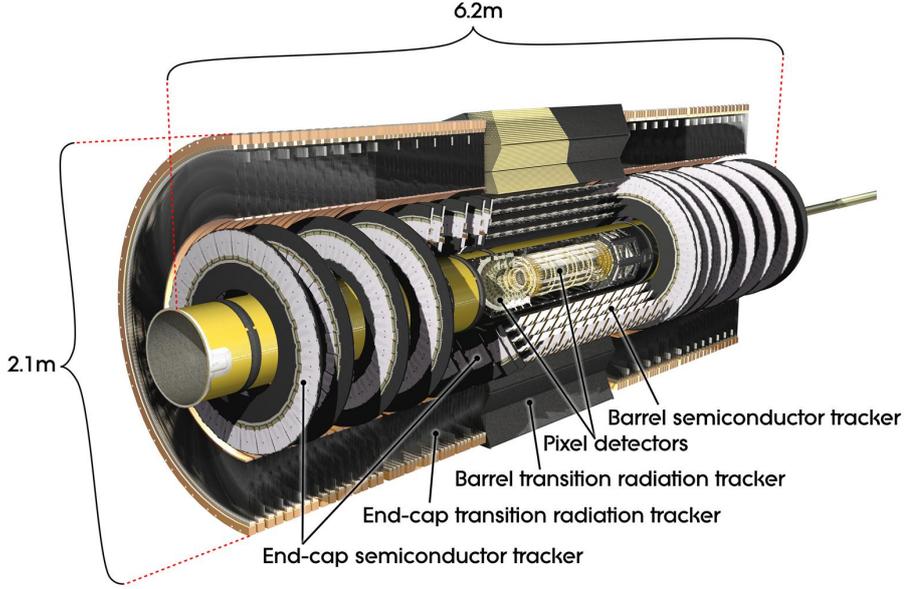


Figure 3.4: Cut-away view of the Inner Detector. [48]

$115 \mu\text{m}$ in z (Barrel) or in R (endcap region). The innermost layer of the pixel detector, the B-layer is important in the measurement of secondary vertices. In total the pixel detector has 80.4 million readout channels. It is surrounded by a Silicon microstrip detector, the semiconductor tracker (SCT) ($|\eta| < 2.5$) using stereo strips and 6.3 million readout channels. Its accuracy is $17 \mu\text{m}$ in $R \cdot \phi$ and $580 \mu\text{m}$ in R (barrel) and in z (endcap region). The third part around the SCT, the Transition Radiation Tracker, consists of 4 mm straw tubes, filled with a xenon-based gas mixture ($|\eta| < 2.5$). It delivers only $R \cdot \phi$ information and provides an accuracy of $130 \mu\text{m}$. It has 351000 readout channels. The combination of the three systems allows a precise tracking measurement.

Calorimeters: The Inner Detector is surrounded by an electromagnetic calorimeter (precision measurements of electrons and photons) which is itself surrounded by a hadronic calorimeter (precision measurement of jets and MET). Both parts have to limit the punch-through to the muon system. Its coverage is $|\eta| < 4.9$. The electromagnetic calorimeter has a thickness of more than $X_0 = 22$ radiation lengths in the barrel and more than $X_0 = 24$ radiation lengths in the endcap. The electromagnetic calorimeter is a lead-LAr detector with accordion-shaped electrodes and lead absorber plates [22]. The barrel part covers $|\eta| < 1.475$ and the endcap part $1.375 < |\eta| < 3.2$. The hadronic calorimeter comprises of two different parts: A tile calorimeter with a barrel $|\eta| < 1.0$ and two extended barrels $0.8 < |\eta| < 1.7$. It uses steel as absorber and scintillating tiles as active material. It is complemented by a LAr hadronic endcap calorimeter (HEC) with two independent wheels directly behind the EM Calorimeters in the endcaps and covers $1.5 < |\eta| < 3.2$. It combines copper plates as absorber with LAr as active medium. These systems are completed by the LAr Forward Calorimeter (FCal) which covers $3.1 < |\eta| < 4.9$ and measures electromagnetic showers as well as hadronic showers with copper and tungsten as absorber. The calorimeter structure is shown in figure 3.5

Muon Spectrometer: The high precision measurement of muons in the muon system needs a magnetic field, which is provided by a barrel toroid system ($|\eta| < 1.4$) and two smaller endcap toroid coil systems ($1.6 < |\eta| < 2.7$). In the transition region, $1.4 < |\eta| < 1.6$, barrel

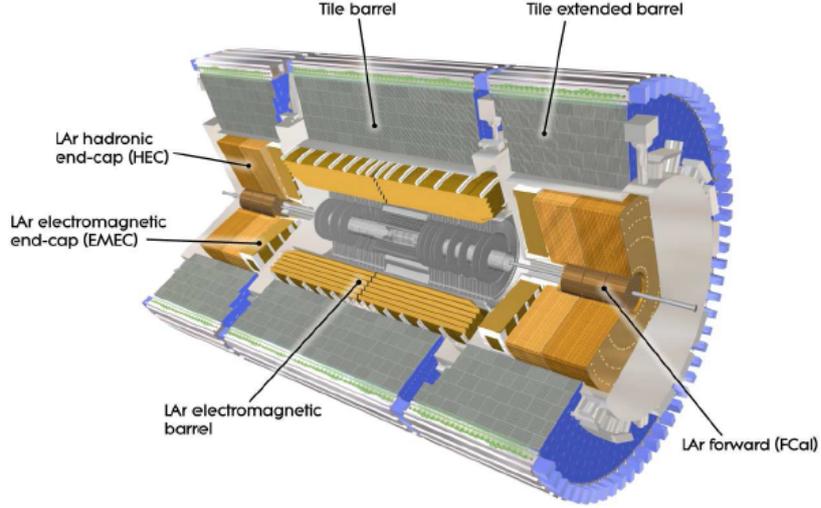


Figure 3.5: The structure of the calorimeters. [22]

and endcap magnetic fields have to be combined. The barrel toroid consists of eight radially and symmetrically grouped coils around the beam axis outside of the hadronic calorimeters. The measurement of the muon track is performed by muon chambers which are arranged in three cylindrical layers around the calorimeters and in three layers orthogonal to the beam axis in the endcaps. Four different types of muon chambers exist: Monitored Drift Tubes (MDT's), Cathode Strip Chambers (CSC's, multiwire proportional chambers with cathodes segmented into strips) which are used at higher $|\eta|$ ($2 < |\eta| < 2.7$) and have a higher granularity and two chamber types which are assigned to the trigger system (more about the trigger system below), the Resistive Plate Chambers (RPC's) in the barrel and Thin Gap Chambers (TGC's) in the endcaps. The TGC's measure the muon coordinate part orthogonal to the part measured by the RPC's. Because of the required precision in the muon momentum measurement the alignment of all the components of the muon system must be known and is supervised by sensors.

Forward detectors: In addition to the described detector components three smaller detector systems exist. Of particular interest are the two detectors measuring the luminosity: LUCID (LUminosity measurement using Cerenkov Integrating Detector), $\pm 17 m$ away from the interaction point and close to the beam, and ALFA (Absolute Luminosity For ATLAS, uses scintillating fibre trackers located inside Roman pots [22]), $\pm 240 m$ away from the interaction point and close to the beam.

Trigger: Because of limited data storage place, it is not possible to store all collision data. Therefore, an advanced decision system in three steps has been developed in order to reduce the data amount and to filter the events of interest. In total the data rate has to be reduced to 200 Hz. This trigger system is shown in detail in figure 3.6. The first step, Level 1 (L1), is purely hardware based and uses only a part of the whole detector information and looks for electrons, photons, jets, hadronically decaying taus or muons with high p_T , as well as high MET or high transverse energy. Hereby the two types of trigger chambers in the muon spectrometer and the calorimeter information is used. The L1 trigger also define regions of interest (RoI's) in η and in ϕ . The L1 trigger rate cannot exceed 75 kHz and a decision has to be taken within $2.5 \mu s$. In interesting events, the second level, L2, uses the complete detector

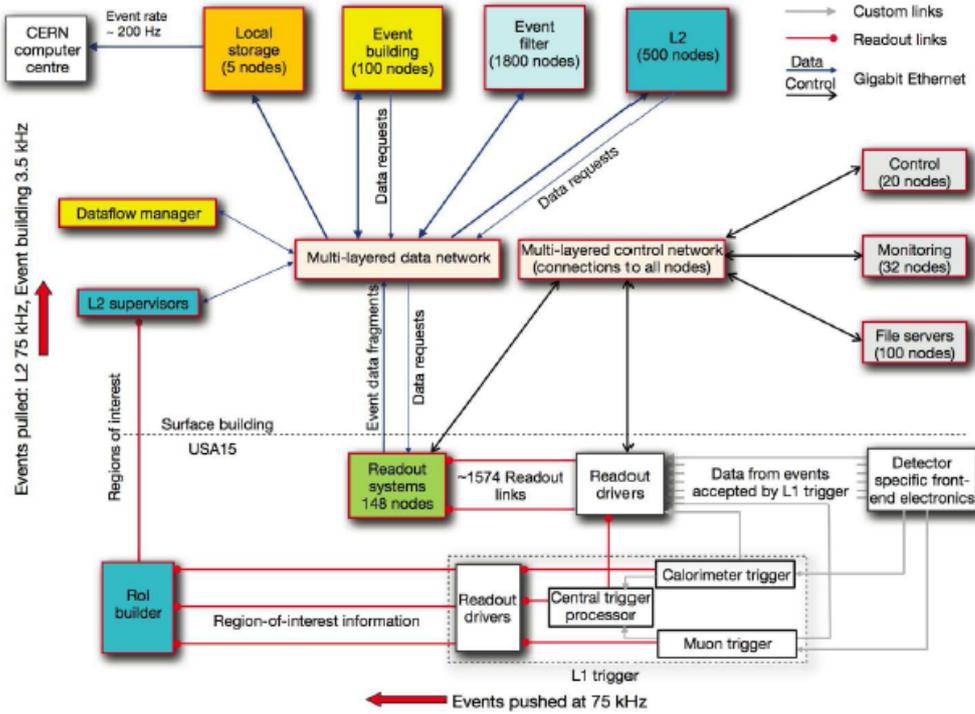


Figure 3.6: The trigger system of the ATLAS system. The incoming data rate has to be reduced by this system to 200 Hz. [22]

information in the found Regions of Interest. Only two percent of the available data is used. The trigger rate is again further reduced to 3.5 kHz. The event processing time is 40 ns. The final step, with the use of offline analysis procedures, is the Event Filter (EF). It reduces the trigger rate to 200 Hz and the event processing time is 4s. Different triggers are available for physics analysis. Although this work cannot treat the triggers in detail, it will be necessary to explain the triggers used in more detail later in the data chapter.

3.3 Event reconstruction and computing

Simulated Monte Carlo events as well as data are stored in various formats. Outgoing from the Byte-stream Data (output from the triggers) and the Raw Data Objects (explained earlier in section 2.4) two data formats are obtained: *Event Data Summary (ESD)* and *Analysis Object Data (AOD)*. ESDs contain the detailed detector reconstruction output information, whereas AODs contain summaries of the reconstructed events. In the process of reconstruction between Raw Data and ESDs the whole detector information is used to obtain reconstructed objects. The AODs (sometimes also the ESDs) are often condensed into n-tuples (D3PDs). These can be read easily by macros based on the ROOT framework [15]. N-tuples, created with the SUSYD3PDMaker² [16], were used for this work. The analysis of data at a single computer center is complicated or not possible due to two reasons: The analysis jobs often need more resources than a single computer center can provide. Second and highly problematic, the input datasets, whether ESD, AOD or n-tuples, are so big that it is often impossible

²mainly the versions 00-05-05, 00-06-14 and 00-06-20

to store them locally. A solution to this is provided by the Grid: The datasets are stored at a huge amount of storage elements all over the world and it is the analysis job which is sent to the data and processed at the place of the data with the help of huge computing resources. The Grid was used heavily for this work.

Chapter 4

Object definitions

The whole ATLAS detector, described in the previous chapter, is necessary to reconstruct and identify electrons, muons, jets and MET (Missing Transverse Energy). For a better understanding of the following parts the mechanism of reconstruction and identification of electrons, muons, jets and MET will be explained.

4.1 Electrons

The information about the electron reconstruction and identification was extracted from [23], [24], [25], [26].

Reconstruction of electrons: The Inner Detector as well as the electromagnetic calorimeter are both used and necessary in the reconstruction of electrons. Two electron reconstruction approaches exist for electrons within $|\eta| < 2.5$: The first starts with the calorimeter information and combines it with the ID information. This is the standard reconstruction algorithm and used mainly for isolated high- p_T electrons. The second algorithm works vice-versa and is used in the reconstruction of low energetic electrons and for the reconstruction of electrons in a jet.

The **standard reconstruction algorithm** looks for a cluster with a deposited energy $p_T > 2.5 \text{ GeV}$ in the second layer of the electromagnetic calorimeter. This cluster has a size of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$. This cluster is matched to an appropriate track found in the ID by a matching procedure. First, it compares η and ϕ of cluster and of all tracks found in the ID with $p_T > 0.5 \text{ GeV}$. The tracks with a reasonable agreement in η and in ϕ are extrapolated to the electromagnetic calorimeter. The closest track to the cluster is usually used as the matched track.

The second algorithm (**track-based algorithm**) starts from a track in the ID. Only a limited amount of tracks in the Inner Detector is chosen by applying quality cuts to the tracks. These quality cuts avoid fakes. These quality cuts contain requirements on the number of hits in the different parts of the Inner Detector. These tracks are extrapolated to the second layer of the electromagnetic calorimeter and thereof an electron cluster is obtained.

Identification of electrons: Three different categories of electrons are commonly used; they differ in the jet rejection rate and the electron identification efficiency. **Loose** electrons show a good electron identification efficiency, but a low jet rejection rate. In the selection the leakage into the hadronic calorimeter¹ and the shower shape in the middle layer of the electromagnetic calorimeter are used. **Medium** electrons require basic track quality criteria and use more calorimeter information in addition to the criteria for loose electrons. They

¹defined by the fraction of transverse energy in the first sampling of the hadronic calorimeters behind the cluster and the energy of the cluster divided by $\cosh\eta$ (of cluster in second sampling)[27]

have a better jet rejection rate by a factor of 3-4, but also reduced electron identification efficiency by 10 %. **Tight electrons** show the highest jet rejection rate, but also the lowest electron identification efficiency. At least one hit in the B-layer, more track requirements and in formation of the TRT is needed in addition to the requirement for medium electrons. Electron candidates which are matched to reconstructed photon conversions are rejected.

4.1.1 Electrons in SUSY analysis

In this work the common SUSY object definition cuts for electrons are used.

- **Medium** electrons are required
- Electron candidates which were reconstructed by an algorithm which starts from seeds in the calorimeter are selected. (*author* = 1 or *author* = 3)
- Only electrons within $|\eta| < 2.5$ are used
- $p_T > 20 \text{ GeV}$
- **Crack electrons are rejected:** These are electrons within $1.37 < |\eta| < 1.52$. This is the intermediate region between barrel and endcaps where a low performance is expected.
- **Isolated electrons** are defined by $etcone20 < 10\text{GeV}$ and the overlap removal (explained below). *etcone20* is the energy deposited in a cone of $\Delta R = 0.2$ around the electron without taking into account the expected energy of the electron itself.

4.2 Muons

4.2.1 Reconstruction and Identification of muons

In the coverage of $|\eta| < 2.7$, a good muon reconstruction and identification is possible for muons with $3 \text{ GeV} < p_T < 1 \text{ TeV}$ [29]. The resolution of the muon momentum measurement in the muon spectrometer varies with p_T : For low p_T muons the fluctuations of the energy loss in front of the muon spectrometer are crucial. For intermediate p_T muons multiple scattering in the muon spectrometer has to be taken into account. For muons with $p_T > 300 \text{ GeV}$, the resolution is determined by the single hit resolution. The muon momentum resolution is limited by the detector characteristic, alignment and calibration [29]. As a further momentum measurement of the muon from the Inner Detector can be obtained, both the measurements from the Inner Detector and the muon spectrometer are combined for $30 \text{ GeV} < p_T < 200 \text{ GeV}$. Below 30 GeV, the measurement from the Inner Detector is more accurate, above the measurement of the muon spectrometer. Three different reconstruction approaches exist and for each of them two different algorithms are available: Staco, which is the default in the SUSY working group at the moment, and Muid. These reconstruction approaches are [30],[31]:

- **Standalone Muon:** Hits in the muon spectrometer are combined into segments and tracks. The tracks are extrapolated to the beam. A correction for energy loss in the calorimeter and the multiple scattering in the muon spectrometer is applied. Although this approach has a slightly bigger coverage in $|\eta|$ up to 2.7 ² as it uses only the muon spectrometer, it is not possible to reconstruct low energy muons with it, which do not reach the muon spectrometer. Muons produced by π or Kaon decays in the hadronic

²But there are some holes in the coverage in $|\eta| = 0.0$ and 1.2

calorimeter can also be misidentified as muons from collisions. The Staco algorithm is called Muonboy, the Muid algorithm Moore.

- **Combined Muon:** Tracks in the Muon Spectrometer are combined with Inner Detector tracks by a statistical methods. Therefore the Inner Detector as well as the muon spectrometer measurement are used. An important measure of the quality of this match is given by the match chi-square value.³
- **Tagged muon:** Inner Detector tracks are extrapolated to the muon spectrometer and the algorithm looks for close segments in the muon spectrometer. These must be close enough, so that the tagged muon candidate is found. This approach might be useful for low p_T muons. The momentum measurement in the Inner detector is used. Tracks in the Inner Detector used for combined muons or segments in the muon spectrometer used by Muonboy are not used for tagged muons in the Staco algorithm. Therefore, tagged muons show no overlap with combined muons in the Staco algorithm.

4.2.2 Object definition in SUSY analysis

Combined muons of the **staco** family are usually used in SUSY studies. Further, the following requirements must be fulfilled:

- The fit of the tracks in Muon Spectrometer and in Inner detector must be successful ($0 \leq matchchi \leq 100$).
- The muon transverse momentum must satisfy $p_T > 20GeV$.
- Only muons within $|\eta| < 2.5$ are selected.
- **Isolation:** Isolated muons are defined by $etcone < 10GeV$ and by $\Delta R > 0.4$ between muon and closest reconstructed jet. The last criteria is usually applied in a overlap removal (explained below). $etcone20$ is again the energy deposit in a cone of $\Delta R = 0.2$ around the muon (without the muon energy).

4.3 Jets

In this section, the hadronization of quarks in the calorimeters, the so-called jets, is described in more details. The definition of a jet itself depends on the jet reconstruction algorithm and there is a variety of algorithms available. However, ideally each algorithm should satisfy the following theoretical requirements. These are taken from [32], [33].

- **Infrared safety:** The presence or absence of any soft particles between two particles which belong to the same jet should not affect the recombination of these two particles to the jet. The number of reconstructed jets should not be affected by any soft particle which come from the fragmentation of a hard scattered parton.
- **Collinear safety:** The jet reconstruction should be independent from a particle splitting into two collinear particles.
- **Order dependence:** The same hard scattering process should be reconstructed independently at parton-, particle- or detector level [32].

³This is the difference between outer and inner track vectors weighted by their combined covariance matrix. $\chi_{match}^2 = (\mathbf{T}_{MS} - \mathbf{T}_{id})^T (\mathbf{C}_{ID} + \mathbf{C}_{MS})^{-1} (\mathbf{T}_{MS} - \mathbf{T}_{ID})$. \mathbf{T} is vector of five track parameters at the point of closest approach to beam line, \mathbf{C} is the covariance matrix.

Some experimental requirements should be satisfied as well: In particular the influence of the detector on the jet reconstruction should be minimal. So the detector resolution (spatial and energetic resolution) must have only a minimal influence; electronic noise, inactive material or cracks between the detectors should also not affect the reconstruction of jets. The jet reconstruction has to be as independent from processes in its environment as possible, e. g. from the underlying event, pile-up, multiple interactions. Further, jets from energetic partons should be reconstructed with high efficiency.

4.3.1 Reconstruction of jets

Although a variety of jet algorithms is available for physics studies, only the AntiKt4H1Topo jets are described here in more detail, as these are the current default for SUSY studies. The anti- k_{\perp} algorithm can be understood as generalization of sequential recombination algorithms [35]. The distance between two objects, d_{ij} , is defined in these algorithms as

$$d_{ij} = \min(p_{T_i}^{2k}, p_{T_j}^{2k}) \frac{(\Delta R)_{ij}^2}{R^2} \quad (4.1)$$

and between an object and the beam as

$$d_{iB} = p_{T_i}^{2k} \quad (4.2)$$

with

$$(\Delta R)_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \quad (4.3)$$

R is the distance parameter (and in principle a free parameter) and is used to control the size of the jets. A common value for R is 0.4 (for narrow jets, for wide jets $R = 0.6$ is used). Of a list of all d_{ij} and d_{iB} the smallest quantity is chosen. If this is a d_{ij} both objects contributing to d_{ij} are combined. If it is a d_{iB} which is the smallest quantity, the object contributing to d_{iB} is considered as jet and it is removed from the list. Jet algorithms differ by the value of k : $k = 1$: k_{\perp} , $k = 0$: Cambridge/Aachen algorithm and $k = -1$: anti- k_{\perp} . The last algorithm with a cone size of 0.4 is used commonly in SUSY-analysis (AntiKt4H1Topo jets are used). Its advantage is, that the jet boundaries are unaffected by soft radiation, because within $\Delta R < R$ soft objects are merged with the hard object. If two hard objects are found within $\Delta R < R$ of each other, both objects are merged; if they are found within $R < \Delta R < 2R$ to each other, the energy will be shared according the relative transverse momentum and their distance.

4.3.2 Bad jets

In data events it is possible to find events with "bad" jets, e. g. fake jets originating from hardware problems, cosmic ray showers or LHC beam conditions. The definitions and sources are described in detail elsewhere and detailed recommendations by the JetEtmiss group exist on how to identify bad jets [36]. If such a bad jet is found in the event, this particular event is rejected. In particular bad jets can be [39]

- Sporadic noise bursts in the HEC can cause fake jets.
- Fake jets in the electromagnetic calorimeter due to coherent noise bursts in neighboring cells.
- Jets which are reconstructed from energy deposits in the calorimeter not related to a collision event, so originating from cosmics or beam background.

4.3.3 Jet Energy Scale Calibration (JES)

The measurement of jets takes place at electromagnetic energy scale (EM) in the calorimeters. The electromagnetic energy scale is correct for electrons and photons and has been verified in test beam measurements. However, this electromagnetic energy scale is not correct for the measurement of jets due to instrumental/detector effects. The response of the calorimeter to hadrons is lower than to electrons, as binding energy is lost in the calorimeter and not all particles, like soft neutrons or neutrinos for example, can be detected. In addition, energy loss due to dead material has to be taken into account. Not all particles of the jet might be contained in the calorimeter or in the jet. Finally, inefficiencies in the calorimeter clustering or in the jet reconstruction can occur. Therefore, the energy and the momentum of the jets have to be corrected. This is done with reference to Monte Carlo truth jets. The correction is applied for each jet using its p_T and η . Detailed information about the calibration can be found in [37].

4.3.4 SUSY selection cuts

In the SUSY analysis, as well as in this work, **AntiKt4H1TopoJets** are used. Events with one or more bad jets are rejected. Only jets within $|\eta| < 2.5$ are used.

Between identified electrons, muons and jets an **overlap removal** is applied. In this overlap removal jets within a radius of 0.2 of a selected electron are removed (so the object is considered as electron). The second part of the overlap removal can be understood as additional isolation criteria for electrons and muons. Hereby, isolated electrons and isolated muons are removed if they are closer than $\Delta R = 0.4$ to a jet. This guaranties that no electron or muon within a jet is selected. This work will use extensively **non-isolated** muons (sometimes also non-isolated electrons). These non-isolated leptons (electrons and muons are referred to as leptons, taus are considered separately) are defined as being in a jet. Hence, there is expected activity around the electron or the muon. The definition of non-isolated muons or electrons consists of two parts: First the second part of the overlap removal is inverted: The muon or the electron, respectively, need to have a distance of less than 0.4 to a jet ($\Delta R < 0.4$), so has to be in the jet. The second part inverts the *etcone20* requirement defined above: Non-isolated muons or electrons need to have *etcone20* > 10 GeV.

4.4 Missing Transverse Energy (MET)

As explained in the introduction the typical SUSY signature in the 1 lepton channel is usually 1 lepton + jets + MET. If all particles could be detected, their energy deposits in the subdetectors in the transverse plane with respect to the two colliding beams should sum to zero because of transverse momentum conservation. However, it is not possible to detect neutrinos or the LSP, which escape the detectors. Therefore, the energy deposits in the transverse plane will not cancel if such particles escape. The negative of the vectorial sum of all found energy deposits is the Missing Transverse Energy (MET or E_T^{miss}). Using all reconstructed objects (electrons, photons, jets, taus and muons) the transverse momentum balance reads:

$$\sum_{reconstructed\ particles} \vec{p}_T + \vec{p}_T^{miss} = 0 \quad (4.4)$$

or

$$\vec{p}_T^{miss} = - \sum_{reconstructed\ particles} \vec{p}_T \quad (4.5)$$

In general masses are neglected in the definition of E_T^{miss} leading to

$$E_T^{miss} := |\vec{p}_T^{miss}| = \left| - \sum_{\text{reconstructed particles}} \vec{p}_T \right| \quad (4.6)$$

Since all the different physical objects have to be considered in the calculation of MET, the information of the whole detector is necessary and the calculation of the MET is very sensible to the miscalibration of any input physical objects. Imperfections in the detector, which do not correspond to physical objects can also cause fake E_T^{miss} . Fake E_T^{miss} can originate from electronic noise, dead regions in the detector, finite detector resolutions, the limited coverage of some detectors, inactive transition regions between calorimeters and dead or noisy readout channels. A short physical introduction of MET can be found in reference [38], more technical descriptions about the calculation of the MET and the performance can be found in [39] (recently published) and in [40] (in some points out of date). The calculation of the MET is rather complex and various MET definitions exist. In the following, only the commonly used definition in the SUSY group is explained in more detailed. However, much more details are available on the twiki-page of the JetEtmisss group [41].

4.4.1 Calculation of MET

In general the x- and y-components of the MET are calculated as follows:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss, \text{calo}} + E_{x(y)}^{miss, \text{cryo}} + E_{x(y)}^{miss, \text{muon}} \quad (4.7)$$

This calculation takes into account all the different contributions to the MET. The first term, $E_{x(y)}^{miss, \text{calo}}$, is calorimeter based. In the calorimeter, it would mean adding all cells in which energy was deposited. However, it is important to apply a noise suppression to limit the number of contributing cells. Therefore only the deposited energy in 3D-topological calorimeter clusters (TopoClusters) is taken into account. Such a TopoCluster is built from a seed with $|E_{Cell}| > 4 \sigma_{noise}$. σ_{noise} is defined as the Gaussian width of the cell energy distribution measured in randomly triggered events. To this seed are added all neighbor cells with $|E_{Cell}| > 2 \sigma_{noise}$. Cells at the boundary of the TopoCluster need to have $|E_{Cell}| > 0$. These thresholds suppress electronic noise and pile-up from minimum bias events. In total, a common TopoCluster contains about 2500 cells. However, it is found that this reconstructed MET does not result in a good measurement of real MET. Therefore, a refined computation is done on top. Different calibration schemes are available and were used, especially for the treatment of first data. For Monte Carlo, as well as for data, a refined calibration (RefFinal) is used after initial calibrations. Here, the calorimeter cells are associated with parent physical objects in the following order: Electrons, photons, hadronically decaying τ -leptons, jets and muons:

$$E_{x(y)}^{miss, \text{calo, calib}} = E_{x(y)}^{miss, e} + E_{x(y)}^{miss, \gamma} + E_{x(y)}^{miss, \tau} + E_{x(y)}^{miss, jets} + E_{x(y)}^{miss, \text{calo, } \mu} + E_{x(y)}^{miss, \text{CellOut}} \quad (4.8)$$

For each term the negative sum of the calibrated cell energies inside the corresponding objects is used [39]. $E_{x(y)}^{miss, \text{calo, calib}}$ replaces the uncalibrated term $E_{x(y)}^{miss, \text{calo}}$ of equation 4.7. $E_{x(y)}^{miss, \text{calo, } \mu}$ is the correction for the energy loss of muons in the calorimeter. $E_{x(y)}^{miss, \text{CellOut}}$ includes all cells in TopoClusters which have not been included in reconstructed objects. The second term of equation 4.7, $E_{x(y)}^{miss, \text{cryo}}$, is a correction for the energy loss in the cryostat between LAr barrel electromagnetic calorimeter and the TileCal barrel hadronic calorimeter. This is relevant for hadronic showers, as the cryostat has a length of half an interaction length.

The last term, $E_{x(y)}^{miss, \mu}$ is calculated from the momentum of the muons in range $|\eta| < 2.7$:

$$E_{x(y)}^{miss, \mu} = - \sum_{selected\ muons} E_{x(y)}^{\mu} \quad (4.9)$$

In the range $|\eta| < 2.5$ only combined muons are used (although no p_T -cut is applied). Isolated and non-isolated muons are treated differently:

- Isolated muons ($\Delta R_{(jet, muon)} > 0.3$): p_T of the combined measurement of Inner Detector and Muon Spectrometer is used and the energy lost term $E_{x(y)}^{miss, calo, \mu}$ in the calculation of the refined MET in the calorimeter is not added.
- Non-isolated muons ($\Delta R_{(jet, muon)} < 0.3$): Here the energy lost in the calorimeters cannot be separated from the nearby jet. Therefore, the Muon Spectrometer measurement of p_T is used. If this differs significantly from the Inner Detector measurement, the combined measurement and the parametrized energy lost in the calorimeter is used.

In $2.5 < |\eta| < 2.7$ the measurement of the Muon Spectrometer is used alone, as the Inner detector has no coverage here. In general, muons are lost in regions with no Muon Spectrometer coverage. It might be possible to recover these by using the calorimeter information, though. Further, unmeasured or badly measured and fake muons would result in a large fake E_T^{miss} .

4.4.2 SUSY

With the first data and the ongoing calibration of the various MET calculation methods, different definitions of MET were used in the SUSY group as well. However, this concerns mainly the treatment of data. For the biggest part of this work, where Monte Carlo samples have been used, the **MET_RefFinal** definition has been used, which is described in detail in the last subsection, with exception of the last chapter, where a simplified MET_RefFinal definition has been used, which can be applied to Monte Carlo as well as to data. For the studies with data in chapter 9 an earlier definition of MET had to be used as the current definition had not been calibrated. The definition used was based on a local cluster weighting calibration of the TopoCluster cells in the calorimeter with added muon corrections.⁴

⁴MET_LocHadTopo + MET_MuonBoy - MET_RefMuonTrack

Chapter 5

SUSY search in the 1-lepton channel

In order to understand complex SUSY events and the expected backgrounds better, the SUSY working group developed a SUSY inclusive search strategy. This strategy was described in detail in [46]. SUSY signatures are searched for in different channels. For example only events with some reconstructed jets and no lepton (0 lepton + jets channel) or events with some jets and one lepton (1-lepton + jets channel) define such channels. This work treats only the 1-lepton channel. The QCD background is already suppressed in this channel by asking for 1 lepton. In this channel, studies of the SUSY working group have showed, that applying the following cuts are very effective in order to suppress backgrounds and select SUSY events. As shown in only the $t\bar{t}$ and the W+jets background are present after all cuts whereas all other backgrounds contribute nearly nothing. Typical cuts are¹:

1. Exactly **one isolated lepton** with $p_T > 20 \text{ GeV}$ and no further isolated lepton with $p_T > 10 \text{ GeV}$

2. At least **four jets** with $p_T > 40. \text{ GeV}$, one jet must have $p_T > 100 \text{ GeV}$.

These two first cuts define the channel.

3. $MET \geq 80. \text{ GeV}$

4. $MET \geq 0.2M_{eff}$

$M_{eff} = \sum_{i=1}^{N_{jets}} p_T^{jet,i} + \sum_{i=1}^{N_{lepton}} p_T^{lepton,i} + MET$ is the effective mass.

Both cuts reduce the SM-background. Cut 4 eliminates the Gaussian part of statistical fluctuations of the MET measurement which grow with increasing M_{eff} [46].

5. $S_T > 0.2$

The transverse sphericity S_T is calculated via $S_T \equiv \frac{2\lambda_2}{\lambda_1 + \lambda_2}$. Here $\lambda_{1,2}$ are the eigenvalues of the 2×2 sphericity-tensors $S_{i,j} = \sum_k p_{ki} p_{kj}$, where all jets with $|\eta| < 2.5$ and $p_T > 20. \text{ GeV}$ and all selected leptons are used. This cut is often not applied. This cut is applied for historical reasons and its efficiency is questionable. In principle, it reduces the MET background from mismeasured dijet events, because QCD events tend to have $S_T \sim 0$. In contrast, SUSY events tend to have $S_T \sim 1$, because the initial particles

¹The cuts presented here are very strict. Actually, they were too strict in the treatment of first data. Therefore, they were varied regularly by the SUSY group during this work. The selection cuts which are used in this work are defined at appropriate place (and change during this work). In particular the 1-muon + 3 jets channel is always used with different requirements on the jets' p_T . However, the principle - as defined in this chapter - stays the same.

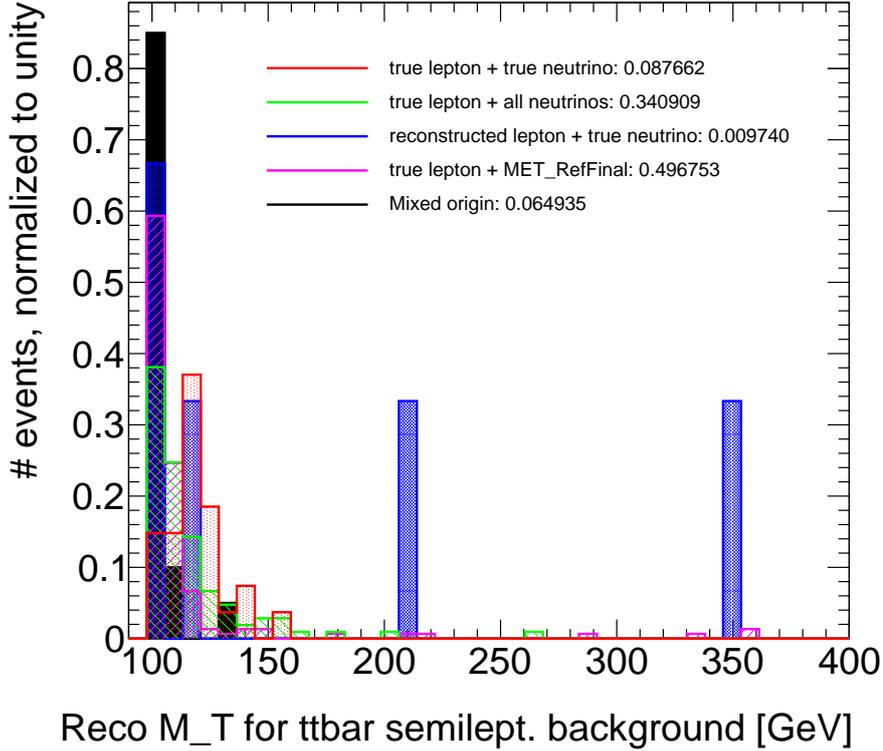


Figure 5.1: Reconstructed M_T (in GeV) for semileptonic $t\bar{t}$ decay topology. The red curve shows the contribution and distribution of events with a true lepton and a true neutrino to M_T (8.7662 %). The green curve shows the distribution of M_T from a true lepton and all neutrinos in the event (34.0909 %). Events with a reconstructed lepton and a true neutrino (blue curve) provide 0.9740 %, events with a true lepton and reconstructed MET (pink curve) 49.6753 %. 6.4935 % events are of mixed origin. All curves are normalized to unity.

are usually produced at rest and emit many particles in various directions during their decay chain [46], [47].

6. $M_T > 100$. GeV

M_T is defined by:

$$M_T^2(\mathbf{p}_T^\alpha, \mathbf{p}_T^{miss}, M_\alpha, m_\chi) \equiv m_\alpha^2 + m_\chi^2 + 2(E_T^\alpha E_T^{miss} - \mathbf{p}_T^\alpha \mathbf{p}_T^{miss}) \quad (5.1)$$

with $E_T^\alpha \equiv \sqrt{(\mathbf{p}_T^\alpha)^2 + m_\alpha^2}$ and $E_T^{miss} \equiv \sqrt{(\mathbf{p}_T^{miss})^2 + m_\chi^2}$

m_α and \mathbf{p}_T^α is the mass and the momentum, respectively, of any visible particle. \mathbf{p}_T^{miss} is the missing transverse-energy two vector. This cut reduce the background from events where a W decays into a lepton and a neutrino.

7. $M_{eff} > 800$ GeV

This cut selects high mass final states.

These cuts were applied to a special MC-sample containing $t\bar{t}$ events whose decay topology is semileptonical. This sample also contains the whole truth information of each event. If the applied cuts would select to 100% efficient SUSY events and reject all other background events, then no $t\bar{t}$ event should survive these cuts.

However, there are some events which survive the applied cuts. Due to the truth information, it is possible to study the causes hereof in more detail: In 29.8 % of all events the event topology of the $t\bar{t}$ decay is semileptonic, in 17.3 % dileptonic and in 52.9 % the decay products include a tau. Of the semileptonically decayed events (see figure 5.1), 8.8 % do indeed have a true $M_T > 100$ GeV (in red): In the Truth information was found a true lepton and a true corresponding neutrino. In 34.9 % of all these events, it was found true MET from neutrinos from bottom or charm decays, but there the neutrino(s) does not correspond to the lepton (in green). In 1.0 % the lepton (in blue) and in 46.7 % the MET of the event (in pink) was misreconstructed. Other causes, which were not studied in more detailed, are responsible of this high M_T in 6.5 % of all cases (in black). In figure 5.1 each distribution was normalized to unity. So, 34.9 % of the events decaying semileptonically survive the selection cuts due to high MET from the decay of heavy flavors. Hence, QCD $b\bar{b}$ events might also survive the selection cuts due to high MET from the decay of heavy flavors. The order of magnitude of the number of QCD $b\bar{b}$ events surviving the selection cuts will be estimated in the next chapters. This will be done via a QCD $b\bar{b}$ control sample which is constructed in the next chapter.

These selection cuts were introduced for SUSY studies at a center-of-mass energy of $\sqrt{s} = 10$ TeV. They were revisited for the first data at a center-of-mass energy of $\sqrt{s} = 7$ TeV. Although the principle and the cuts remained the same, some cut values changed (and not all cuts were applied at the beginning of data taking). The cuts in this work are inspired by the latest version of the SUSY selection cuts (which changed with the time), in particular if an estimation of the QCD background in the signal region is given. However, in this work the full SUSY selection cuts were not applied, as this would reduce the statistics too much for QCD background studies.

Chapter 6

Construction of a QCD $b\bar{b}$ control sample in the 1 lepton channel

This chapter explains how to select a (nearly) pure QCD $b\bar{b}$ control sample with respect to other backgrounds. In order to do this, b-tagging is used and the isolation of the selected lepton is inverted. First, b-tagging is discussed, then the isolation. Finally, both will be combined and some important distributions will be shown. The main part of this chapter uses Monte Carlo datasets produced at a center of mass energy $\sqrt{s} = 10$ TeV. An update to 7 TeV is presented at the end of this chapter.

6.1 b-tagging – principle and efficiency

B-tagging (the application of an algorithm in order to identify b-jets) is a powerful tool to select events with b-jets (so jets coming from the hadronization of b-quarks). It is used in many different physics studies, which aim either at selecting events with b-jets or rejecting such events. Examples can be found in searches for the Higgs boson, in the selection of pure $t\bar{t}$ samples and in very specific physics searches like SUSY with b-jets in the final states [42], [43]. Very recently, it has been used to separate top control regions and W control regions in SUSY background studies. In the study presented here it will be most useful in the selection of a pure QCD $b\bar{b}$ control sample.

B-tagging uses the special properties from B-hadrons to differentiate b-jets from light jets (jets originating from u, d and s quarks) or c-jets. B-hadrons have a high mass with > 5 GeV and obtain a big fraction of the original b-quark momentum, which both result in a large transverse momentum with respect to the jet axis [42]. Many b-tagging algorithms rely on a third important property, which is the relatively long lifetime of B-hadrons (about 1.5 ps). This results in a considerable flight path length. (A B-hadron with $p_T = 50$ GeV will fly about 3 mm in transverse plane before decaying.) The b-tagging algorithms under investigation by the b-tagging group can be divided in three categories: The first two categories use the transverse impact parameter (d_0 , defined as the distance of the closest approach of the track to the primary vertex in the $r - \phi$ -projection) and the longitudinal impact parameters (z_0 , the z-coordinate at this point of closest approach) by using the large impact parameters of decaying B-hadrons or by reconstructing explicitly the secondary vertex. These algorithms are referred to as spatial b-tagging algorithms. The third category uses a completely different approach by tagging the lepton in the jet originating from a semileptonic decaying B-hadron. Despite this variety of b-tagging algorithms, only three are suggested for the use in the early data by the b-tagging group. This is due to the fact that these algorithms are in some sense simple, and therefore expected to be calibrated first. These early taggers are: JetProb (which calculates a jet probability out of the $\frac{d_0}{\sigma_{d_0}}$ distribution for each track associated to the jet

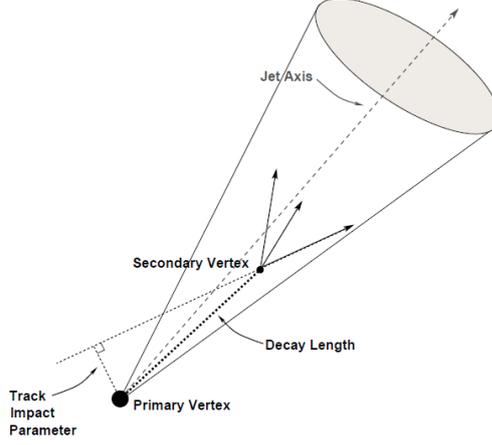


Figure 6.1: The SV0-weight is the signed decay length significance between the primary vertex and the fitted secondary vertex. [44]

with σ_{d_0} being the error on d_0), TrackCounting (a b-jet is tagged if at least a few associated tracks have a large impact parameter or impact parameter significance) and SV0 (cut on the signed decay length significance) [45]. For this study the SV0 algorithm was chosen because of its simplicity.

6.1.1 The SV0-algorithm

The SV0-algorithm builds from tracks fulfilling some quality criteria (theses can be found in the appendix) vertices with two tracks. The tracks need to be associated to the jet (so the tracks are within $\Delta R = 0.4$ of the jet and associated to it via a ΔR -matching). Then, all two-track vertices whose masses are consistent with K_S^0 , Λ^0 or photon conversions or which are consistent with material interaction in the pixel detector layers are removed and a secondary vertex is fitted with the remaining ones. The SV0 weight is obtained by dividing the distance between the primary vertex and the fitted secondary vertex, the decay length, by the error of the decay length (so it is the signed decay length significance). The sign is positive if the projection of the decay length vector on the jet axis is in the direction of the jet. An illustration can be found in figure 6.1.

6.1.2 Efficiency and Rejection

The performance of the b-tagging algorithms with respect to the different physical processes can be described by the following variables:

- The **efficiency** of the b-tagging is defined as the fraction of the tagged jets which are true b-jets with respect to all true b-jets:

$$\epsilon = \frac{N_{true\ b-jets\ and\ tagged}}{N_{true\ b-jets}} \quad (6.1)$$

- The **rejection** is defined as the fraction of all true light jets with respect to the tagged true light jets:

$$r = \frac{N_{true\ uds-jet}}{N_{true\ uds-jet\ and\ tagged}} \quad (6.2)$$

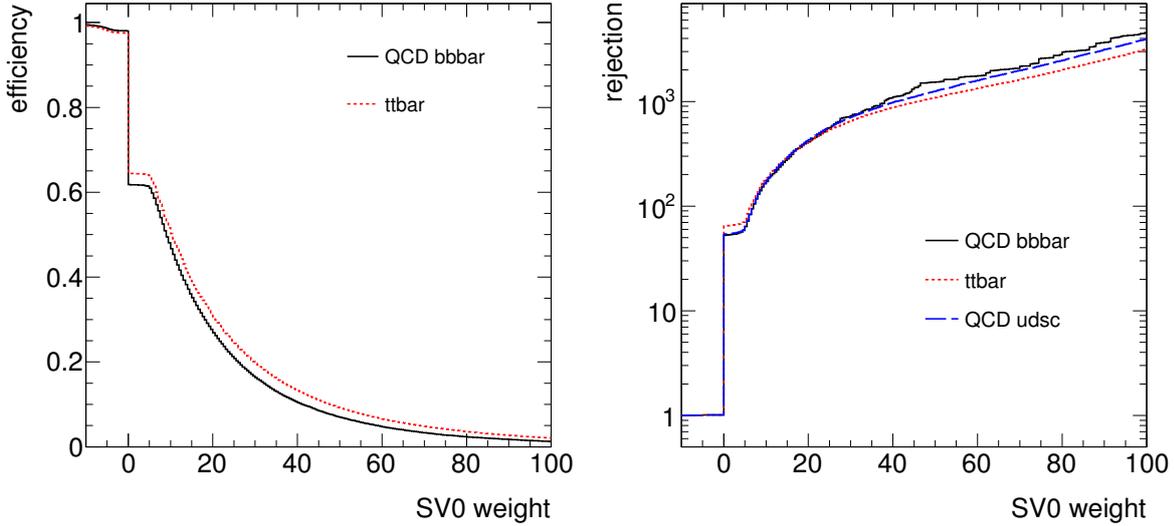


Figure 6.2: On the left: Efficiency (defined in the text) for the QCD $b\bar{b}$ and the $t\bar{t}$ samples. On the right: Rejection (defined in the text) for the QCD $b\bar{b}$, $t\bar{t}$ and QCD udsc samples.

	QCD bbbar	ttbar	QCD udsc
SV0-weight	9	10	7.5
rejection	153	184	130

Table 6.1: SV0-weight and rejection for an efficiency of 50% in the Monte Carlo samples used.

- The **purity** is defined as the fraction of the number of tagged b-jets with respect to the number of all tagged jets:

$$p = \frac{N_{\text{true } b\text{-jets and tagged}}}{N_{\text{all tagged jets}}} \quad (6.3)$$

The efficiency and the rejection were obtained for QCD $b\bar{b}$, QCD udsc and $t\bar{t}$ events in the Monte Carlo datasets used for this work. Only events with at least 3 jets with a transverse momentum of at least 60, 25 and 25 GeV, respectively, were used for the following plots (this cut was necessary to correct for the different filters applied to the Monte Carlo samples in their production). In figure 6.2, on the left side, the efficiency against the SV0 weight is shown for $t\bar{t}$ and QCD $b\bar{b}$. For negative SV0-weight both curves are near 1, since all b-jets are tagged. Close to 0 a very sharp decrease can be observed which is due to the sharp peak in a SV0 distribution at 0, which means a decay length of 0 of the corresponding B-hadron (e. g. if no secondary vertex was found). For positive values of the SV0-weight a monotone decrease can be seen. The efficiency of $t\bar{t}$ events is always above the efficiency of QCD $b\bar{b}$ events, which is probably due to higher energetic jets in the $t\bar{t}$ processes, since the efficiency is p_T -dependent [42]. In the plot on the right the rejection against the SV0-weight is shown, which increases with higher SV0-weights. In figure 6.3 the efficiency is plotted against the rejection. It can be observed that a high efficiency results in a low rejection and vice-versa. Therefore, the optimal SV0-weight for each study depends on whether the efficiency or the rejection should be high. Usual efficiencies taken are 0.5 or 0.6. The corresponding SV0-weights and rejections can be found in table 6.1 and 6.2.

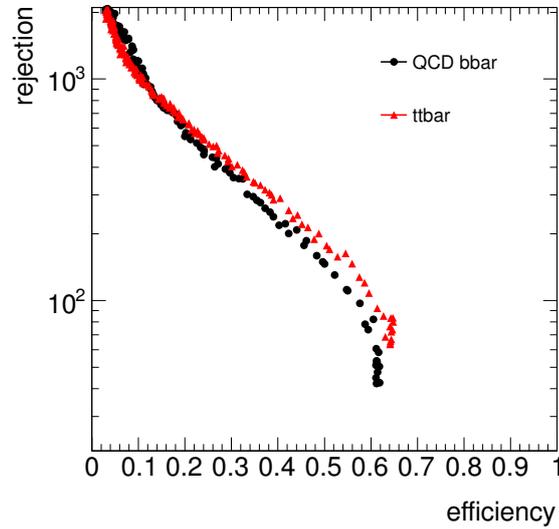


Figure 6.3: Efficiency versus rejection for the QCD $b\bar{b}$ and the $t\bar{t}$ samples.

	QCD bbbar	ttbar	QCD udsc
efficiency	0.61	0.63	-
rejection	63.6	74.0	64.2

Table 6.2: Efficiency and Rejection for a SV0-weight of 5 for the three different types of samples.

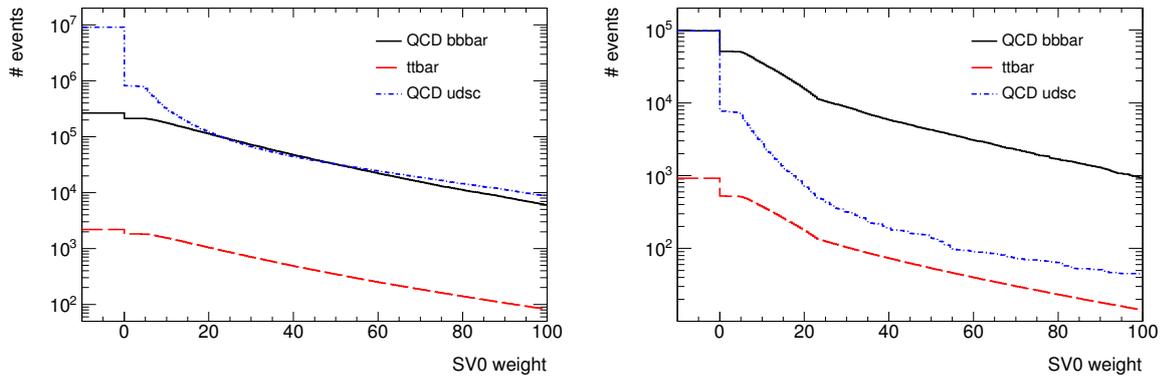


Figure 6.4: Left: Number of events after selecting events with at least one jet above the SV0 weight on the x-Axis. Here no 1-lepton cut is applied. In the region of SV0 weight above 23 the QCD $b\bar{b}$ dominates over QCD udsc and $t\bar{t}$. Right: Number of events after selecting events with at least one jet with a SV0-weight greater than 23 and at least a second jet with a SV0 weight greater than the SV0 weight on the x-Axis. Both plots are for 10 pb^{-1} .

6.2 First studies with 10 TeV MC samples

6.2.1 Applying a SV0-cut

In order to select a pure QCD $b\bar{b}$ sample, $t\bar{t}$ events and QCD udsc events need to be suppressed. Since there are only a few b-jets in the QCD udsc samples (coming from gluon splitting), QCD udsc events should be well suppressed by applying a b-tagging cut. This was first tested without any conditions for the leptons in the event. In figure 6.4 on the left the number of all events which contain at least a jet which has a SV0-weight greater than a given value is plotted. For very low x (which corresponds to no b-tagging) QCD udsc events dominate due to the large cross-sections of the samples. In the region around 20 the number of QCD $b\bar{b}$ and QCD udsc events become comparable, but for no cut value a strong suppression with respect to QCD $b\bar{b}$ events can be obtained (as the QCD udsc samples contain also events with c-hadrons, these events might survive the b-tagging, because of the similar properties of c-hadrons compared to b-hadrons). The number of $t\bar{t}$ events is always a factor 100 lower than the number of QCD events due to the cross section of these samples. In order to suppress QCD udsc events further, a second b-tagging cut can be applied: After selecting events which contain at least one jet with a SV0 weight above 23, a second jet in the event which has a SV0 weight above some x is required. The distributions can be seen in figure 6.4 (right), where the number of events fulfilling these conditions for a given x (on the x-axis) is shown. It can be seen that each cut on positive x is very effective in suppressing QCD udsc events with respect to QCD $b\bar{b}$ events. Therefore, a low cut value at 5 is preferred in the following in order to get a QCD $b\bar{b}$ control sample with a sufficient statistics.

Since this study aims at creating a QCD $b\bar{b}$ control sample in the 1-lepton-channel, a 1-lepton cut (with no isolation requirement and according to the object definitions explained earlier) is added. In figure 6.5 a plot similar to the plot in figure 6.4 on the left is shown, but with a 1-lepton cut. It can be seen that QCD udsc events are already well suppressed after applying this 1-lepton cut and after applying a b-tagging with a SV0-weight greater than 0. This would not be the case for events with one isolated lepton: There $t\bar{t}$ events would be dominant with respect to both QCD $b\bar{b}$ and QCD udsc events. As the 1-lepton (without any criteria on the isolation) cut suppresses already $t\bar{t}$ and QCD udsc events, it is not necessary to chose the b-tagging weight as high as 23 (which is necessary if applying no lepton cut).

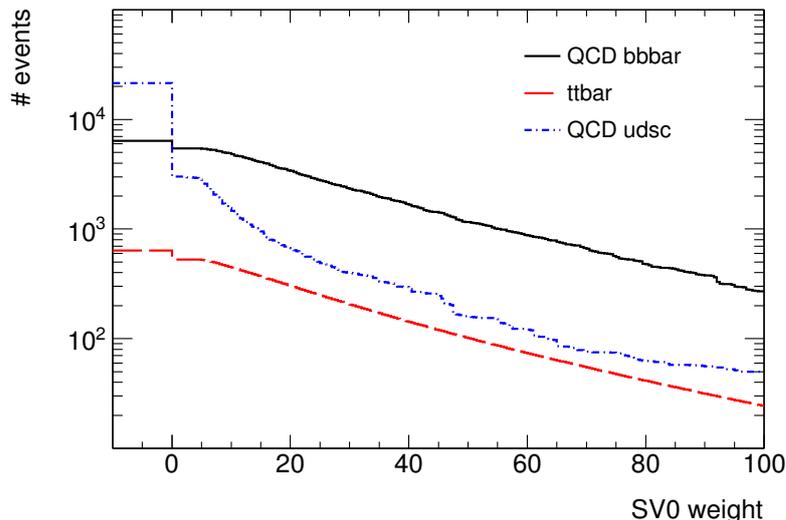


Figure 6.5: The number of events with at least one jet with a SV0 weight greater than the SV0 weight on the x-axis after the applying a 1-lepton cut (10 pb^{-1}).

Hence, a SV0 weight between 5 and 10 is used in the first b-tagging in this work, as this preserves a sufficient statistics in the Monte Carlo samples.

6.2.2 Using the isolation

The b-tagging cuts presented are not able to reduce the number of $t\bar{t}$ events with respect to QCD $b\bar{b}$ events, as a top-quark decays to a bottom quark and a W boson. The latter decays either to a lepton and a neutrino or two light quarks. The lepton of such a decay (of a semileptonic $t\bar{t}$ decay topology) tends to be isolated. So, this can be used to distinguish $t\bar{t}$ events from QCD $b\bar{b}$ events, where the b-quark (the B-hadron) decays preferably to a non-isolated lepton. The isolation of leptons can be well seen in the distribution of the *etcone20* variable, which is the energy deposited in the calorimeter within a cone of $\Delta R = 0.2$ around the lepton.

In figure 6.6 the transverse momentum p_T of the selected lepton (electron or muon) is plotted against the *etcone20* value of this lepton. Clear differences between QCD $u\bar{d}sc$ events (top left), $t\bar{t}$ events (top right) and QCD $b\bar{b}$ events (bottom) can be seen. $t\bar{t}$ events tend to have $etcone20 < 5 \text{ GeV}$, whereas QCD $b\bar{b}$ events tend to have $etcone20 > 10 \text{ GeV}$. The *etcone20* values of QCD $u\bar{d}sc$ events lies in between these two extremes, because of the non-isolated leptons from the decay of c-hadrons and the isolated leptons from conversions or fake leptons. While QCD $b\bar{b}$ events tend to have large *etcone20* values but a low transverse momentum, $t\bar{t}$ events show an opposite behavior. QCD $u\bar{d}sc$ events show both properties. Both variables can be combined into a "relative isolation", defined as $etcone20(\text{selected lepton})/p_T(\text{selected lepton})$. This distribution is shown in figure 6.7, where $etcone20(\text{selected lepton})/p_T(\text{selected lepton})$ of $t\bar{t}$ events peaks between 0 and 0.1, whereas $etcone20(\text{selected lepton})/p_T(\text{selected lepton})$ of QCD $b\bar{b}$ events reaches its maximum between 0.2 and 0.3 with a flatter distribution. Here again, QCD $u\bar{d}sc$ events show an intermediate behavior. It is therefore possible to select QCD $b\bar{b}$ events with a cut on $etcone20(\text{selected lepton})/p_T(\text{selected lepton}) > 0.2$ or to use non-isolated leptons (so leptons with $etcone20 \geq 10 \text{ GeV}$ and $\Delta R \leq 0.4$). Both methods will be studied in the following, but the last method will prove to be more applicable, because it will be easier to get an estimation of the QCD $b\bar{b}$ background in the SUSY signal region with it.

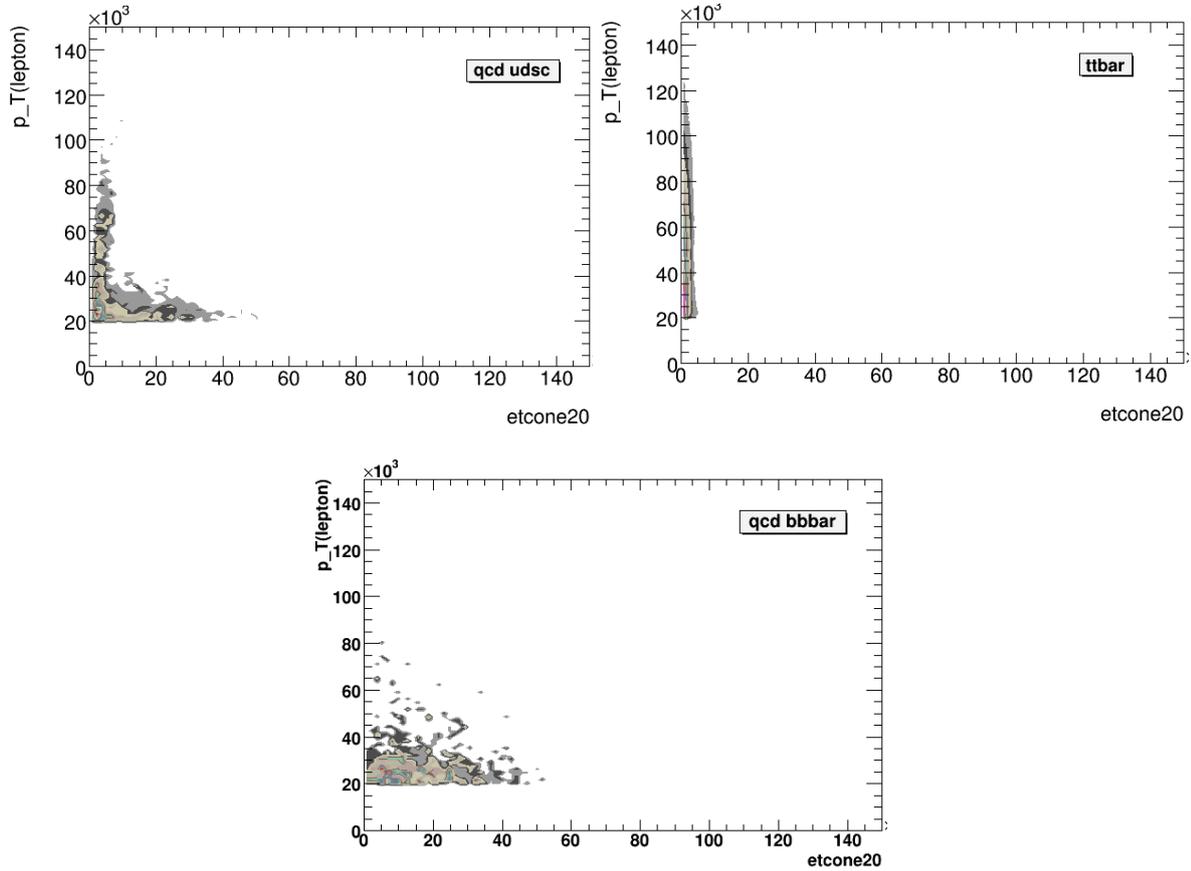


Figure 6.6: Etcone20 [GeV] versus p_T (selected lepton) for QCD $b\bar{b}$ on the bottom, QCD $ud\bar{c}$ on the top left and for $t\bar{t}$ on the top right. QCD $b\bar{b}$ events tend to have lower p_T (selected lepton) and bigger etcone20 than $t\bar{t}$. QCD $ud\bar{c}$ events show partly a distribution like QCD $b\bar{b}$ due to the c -quarks in these samples and partly a distribution with lower etcone20 and larger p_T (selected lepton) values due to isolated leptons from conversions or fake leptons.

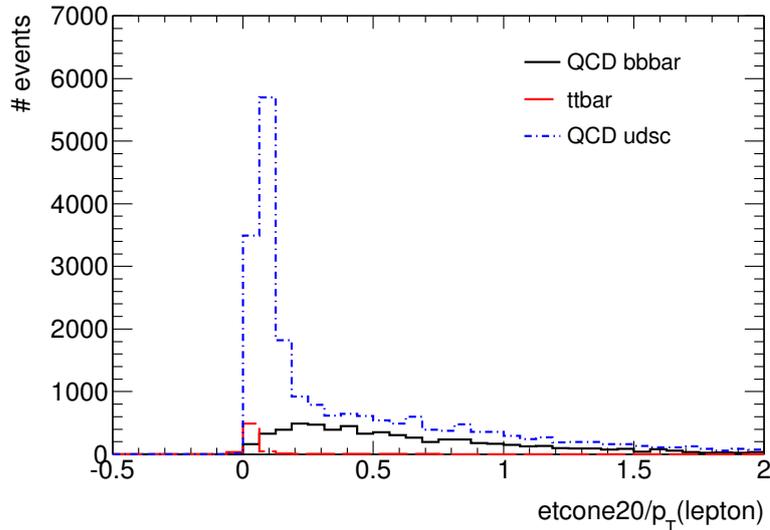


Figure 6.7: The ratio $etcone20/p_T(\text{selected lepton})$ is shown for $t\bar{t}$, QCD udsc and QCD $b\bar{b}$ for events with one selected non-isolated or isolated lepton. Most of the $t\bar{t}$ events have $etcone20/p_T(\text{selected lepton}) < 0.2$. Therefore, a cut on $etcone20/p_T(\text{selected lepton})$ suppresses $t\bar{t}$ events. The plot is for $10 pb^{-1}$.

6.2.3 Applying both cuts

In the previous sections a cut on the isolation has proven to be necessary in order to suppress the number of $t\bar{t}$ event with respect to the number of QCD $b\bar{b}$ events. In order to suppress QCD udsc events, b-tagging is most useful. In figure 6.8 both cuts are combined: The number of events fulfilling the following conditions against the reconstructed Missing Transverse Energy (MET_Reffinal) are shown:

- 1-lepton with $p_T > 20. GeV$
- 3 jets with $p_T > 60, 25, 25 GeV$
- $\frac{etcone20(\text{selected lepton})}{p_T(\text{selected lepton})} > 0.2$
- At least one jet in the event with a SV0-weight greater than 5

In each bin the number of $t\bar{t}$ events is more than a factor 10 less than the number of QCD $b\bar{b}$ events. In contrast, the number of QCD udsc events is only by a factor of approximately 3 lower than the number of QCD $b\bar{b}$ events. Therefore it is necessary to suppress these events further. Increasing both the b-tagging cut value to 8 or 10 and the $\frac{etcone20(\text{selected lepton})}{p_T(\text{selected lepton})}$ -cut to 0.3 was found to be inefficient in suppressing the number of QCD udsc events. But, as was shown earlier, applying a second b-tagging might be useful to suppress the number of QCD udsc events further with respect to the number of QCD $b\bar{b}$ events. The result thereof is shown in figure 6.9. We hence selected events with at least one jet with a SV0-weight greater than 8 and a second jet in the event with a SV0 weight greater than 5. These two b-tagging cuts are motivated from the two expected b-jets in QCD $b\bar{b}$ events. The cut values were chosen low, to preserve an acceptable statistics, but high enough to lead to a good suppression of QCD udsc events. After all cuts, QCD udsc events are found to be suppressed by more than a factor of 10 with respect to QCD $b\bar{b}$ events and $t\bar{t}$ events are negligible.

The Missing Transverse Energy in the SM-events should come mainly from neutrinos and from an imperfect detector (dead layers, incomplete coverage etc.). Hence, as will be

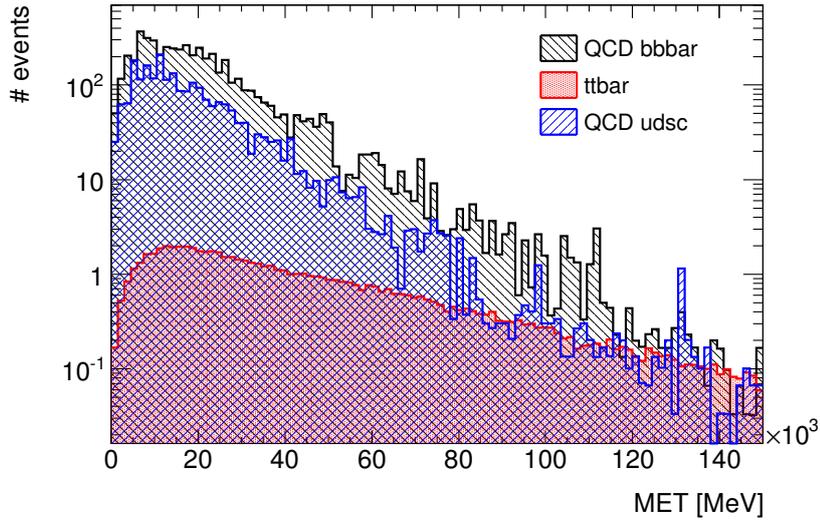


Figure 6.8: Reconstructed MET after applying an $etcone20/p_T(\text{selected lepton})$ cut and selecting events with at least one jet with a SV0-weight greater than 5. QCD $u\bar{d}sc$ events are not sufficiently suppressed by these cuts, while the number of $t\bar{t}$ events is low with respect to the number of QCD $b\bar{b}$ events.

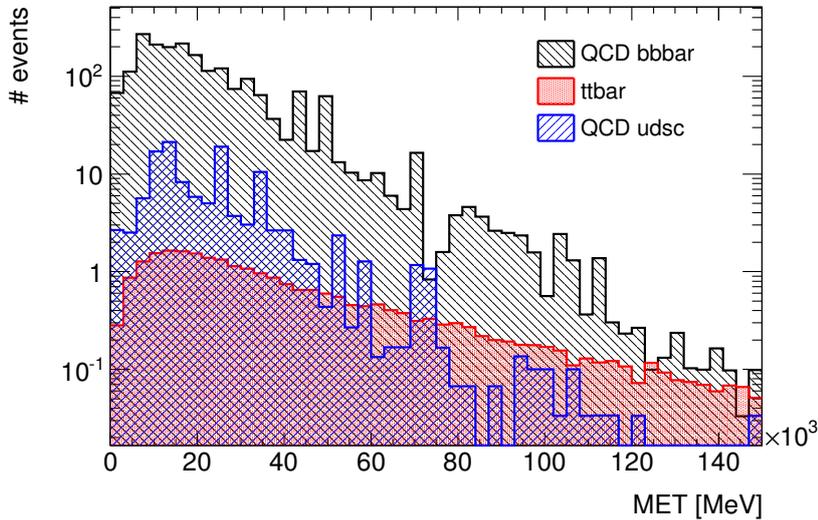


Figure 6.9: Reconstructed MET of $t\bar{t}$, QCD $u\bar{d}sc$ and QCD $b\bar{b}$ events after all selection cuts including the $etcone20/p_T(\text{selected lepton})$ cut and a second b-tagging cut. This second b-tagging cut suppresses the QCD $u\bar{d}sc$ events further compared to figure 6.8.

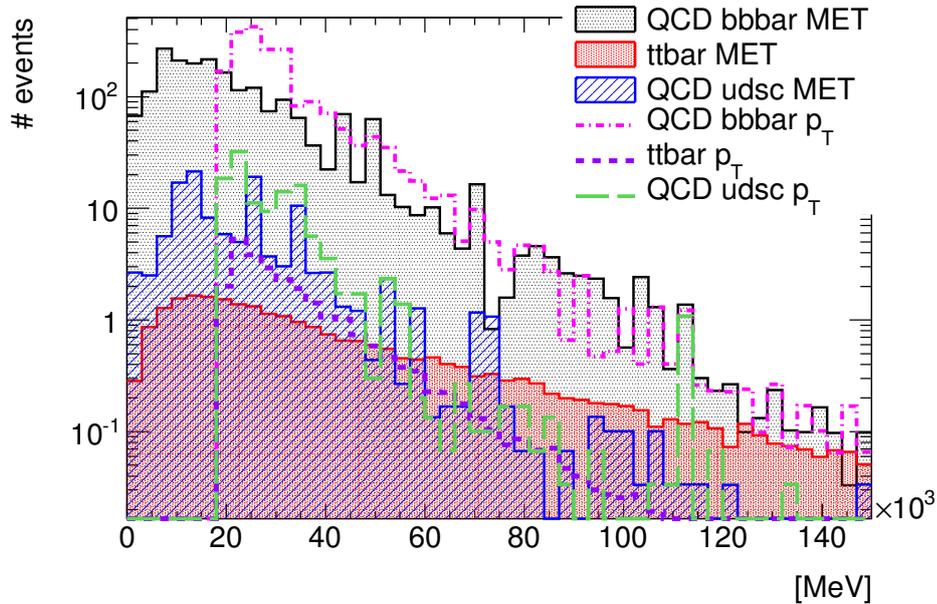


Figure 6.10: Reconstructed MET compared to the distribution of $p_T(\text{selected lepton})$ after all selection cuts including the $etcone20/p_T(\text{selected lepton})$ cut.

explained in the next chapter, the shape of the distribution of $p_T(\text{selected lepton})$ should be similar to the shape of the distribution of MET in QCD $b\bar{b}$ events. Therefore, it is interesting to compare the MET in the event with the $p_T(\text{selected lepton})$, which is shown in figure 6.10. No cut on the MET was applied in the event so far, whereas there is a cut on the p_T of the selected lepton. So, the distribution of p_T begins at the cut value, whereas the MET distribution starts at 0.

Since the control sample will be used to get an estimation of the number of QCD $b\bar{b}$ events in the SUSY signal region (which uses events with a selected isolated lepton), it will be necessary to invert the isolation cut. (Isolation is currently defined as $etcone20 < 10 \text{ GeV}$ and $\Delta R > 0.4$ between lepton and jet. By inverting the isolation, control regions which are clearly separated from the signal region can be defined.) In order to do this in the following non-isolated leptons will be used ($etcone20 \geq 10. \text{ GeV}$ and closer than $\Delta R = 0.4$ to a reconstructed jet). This leads to a relatively big statistics in the control sample, as the leptons in QCD $b\bar{b}$ events are mainly non-isolated, as shown earlier. However, applying an $etcone20/p_T(\text{selected lepton})$ -cut might still be useful as an alternative, because here the statistics after all applied cuts is bigger than in case of using non-isolated leptons. (In addition, the isolation is still under discussion and a $etcone20/p_T(\text{selected lepton})$ -cut is often used as isolation cut as well.) By using non-isolated leptons similar plots as in 6.9 and 6.10 are obtained in 6.11 and 6.12, with the same conclusions: Using non-isolated leptons and applying two b-tagging cuts result in a good suppression of QCD udsc and $t\bar{t}$ events with respect to QCD $b\bar{b}$ events.

6.2.4 Summary: All cuts used to create a pure QCD $b\bar{b}$ control sample

The cuts motivated in the previous section aim at selecting a pure QCD $b\bar{b}$ control sample. They are summarized in the following:

1. 1 non-isolated lepton (muon or electron)
2. At least 3 jets

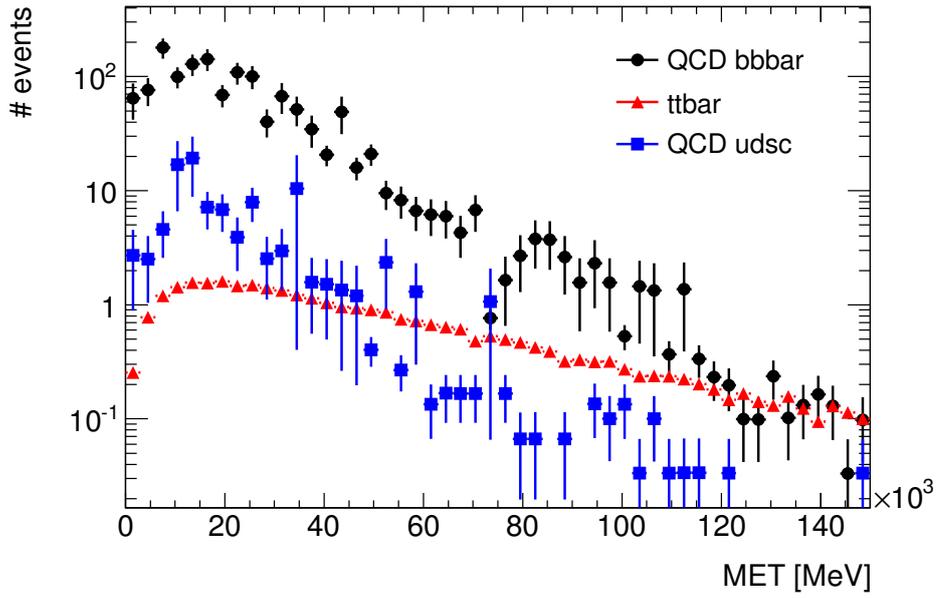


Figure 6.11: Reconstructed MET of $t\bar{t}$, QCD $udsc$ and QCD $b\bar{b}$ events after all selection cuts.

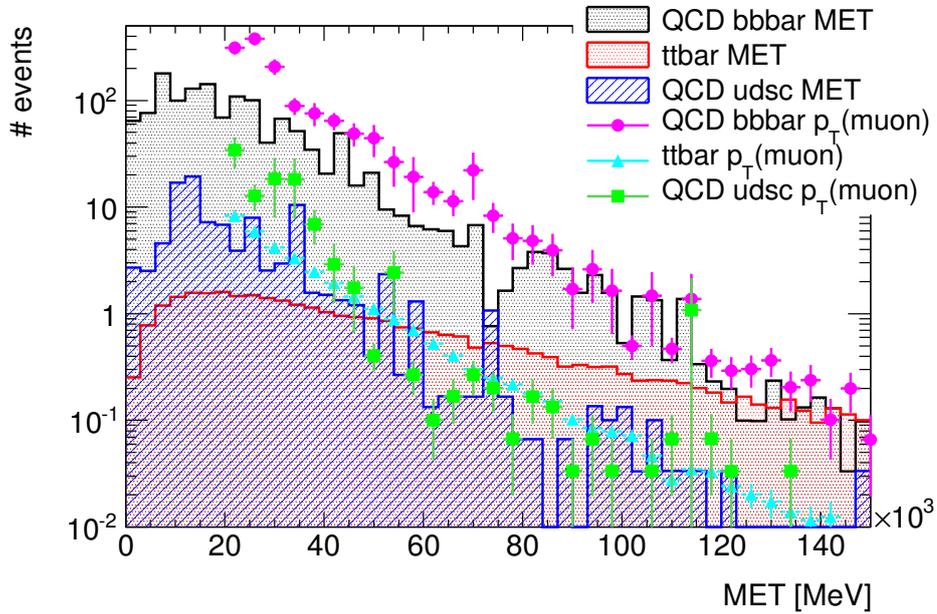


Figure 6.12: Reconstructed MET compared to the distribution of $p_T(\text{selected lepton})$ after all selection cuts and using non-isolated leptons.

	QCD bbbar	ttbar	QCD udsc
3 jets (60,25,25 GeV) + 1 non-isolated lepton	4508.2	84.1	8734.1
1. b-tagging (SV0-weight 8)	3730.2	68.5	1260.4
2. b-tagging (SV0-weight 5)	1683.9	35.8	115.9

Table 6.3: Cutflow for 1 non-isolated lepton + 3 jets with $p_T^{first\ jet} > 60\ GeV$ and $p_T^{third\ jet} > 25\ GeV$.

	QCD bbbar	ttbar	QCD udsc
3 jets (60,25,25 GeV) + 1 non-isolated muon	4505.3	83.6	8379.0
1. b-tagging (SV0-weight 8)	3727.7	68.1	1220.6
2. b-tagging (SV0-weight 5)	1682.7	35.7	105.3

Table 6.4: Cutflow for 1 non-isolated muon + 3 jets with $p_T^{first\ jet} > 60\ GeV$ and $p_T^{third\ jet} > 25\ GeV$.

3. At least one jet in the event with a SV0-weight above 8 (first b-tagging)
4. At least a second jet in the event with a SV0-weight above 5 (second b-tagging)

In the following, different cutflows will be presented for different jet p_T and SV0 weights. In table 6.3 the cuts as described above are applied. Moderate jet cuts with $p_T^{first\ jet} > 60\ GeV$ and $p_T^{third\ jet} > 25\ GeV$ are applied. The number of QCD udsc events is still in the same order of magnitude as the QCD $b\bar{b}$ events after applying the first b-tagging cut at a weight of 8. The second b-tagging cut at a weight of 5 reduces the number of QCD udsc events, so that their fraction is less than a tenth with respect to the number of QCD $b\bar{b}$ events. Only 37.4 % of the QCD $b\bar{b}$ events survive the two b-tagging cuts. The number of $t\bar{t}$ events is low after all cuts, because they are already reduced considerably by demanding a non-isolated lepton.

In the cutflow in table 6.4 only events with a selected non-isolated muon (in contrast to non-isolated lepton of before) are used. Comparing to table 6.3 the numbers do not change much. This shows that the cuts select mainly events with a non-isolated muon. This is due to the electron object definitions, as in this case medium isolated electrons were selected (and this contains already an isolation criteria). The difference between table 6.3 and table 6.4 is bigger in the number of QCD udsc events. In the following, only events with one non-isolated muon will be selected.

The number of QCD $b\bar{b}$ events is sensitive to the applied jet cuts. Increasing the cuts on the transverse momentum of the first three jets from $p_T^{first\ jet} > 60\ GeV$ and $p_T^{third\ jet} > 25\ GeV$ to $p_T^{first\ jet} > 70\ GeV$ and $p_T^{third\ jet} > 30\ GeV$ reduces the number of QCD $b\bar{b}$ events by 20 % (table 6.5 and table 6.6). $t\bar{t}$ or QCD udsc are not reduced by this so strongly. In contrast, slight variations of the SV0 weight in the first b-tagging cut show less effects.

Summarized:

- There are mostly non-isolated muons and only a few non-isolated electrons.
- The difference between different SV0-weights for the first b-tagging is small.
- It is important to apply two b-tagging cuts.
- Tightening the jet cuts leads to a much lower statistic in the final control sample.

	QCD bbbar	ttbar	QCD udsc
3 jets (70,30,30 GeV) + 1 non-isolated muon	3503.7	75.6	6223.6
1. b-tagging (SV0-weight 10)	2761.9	58.8	780.8
2. b-tagging (SV0-weight 5)	1318.1	32.3	88.1

Table 6.5: Cutflow for 1 non-isolated muon + 3 jets with $p_T^{first\ jet} > 70\ GeV$ and $p_T^{third\ jet} > 30\ GeV$.

	QCD bbbar	ttbar	QCD udsc
3 jets (70,30,30 GeV) + 1 non-isolated muon	3503.7	75.6	6223.6
1. b-tagging (SV0-weight 8)	2902.1	62.0	941.6
2. b-tagging (SV0-weight 5)	1347.5	33.0	100.6

Table 6.6: Cutflow for 1 non-isolated muon + 3 jets with $p_T^{first\ jet} > 70\ GeV$ and $p_T^{third\ jet} > 30\ GeV$.

In order to get a pure QCD $b\bar{b}$ control sample, it is also important to suppress other backgrounds like W+jets, Z+jets and W+bb. These are suppressed very well, which the following cutflow 6.7 shows:

	W+jets	Z+jets
3 jets (60,25,25 GeV) + 1 non-isolated muon	32.9	4.8
1. b-tagging (SV0-weight 10)	4.9	0.4
2. b-tagging (SV0-weight 5)	0.7	0

Table 6.7: Cutflow for 1 non-isolated muon + 3 jets with $p_T^{first\ jet} > 70\ GeV$ and $p_T^{third\ jet} > 30\ GeV$ for W+jets and Z+jets.

Wbb is found to be completely negligible.

6.3 A center-of-mass energy of 7 TeV

Since it was decided that the LHC should start with a center of mass energy $\sqrt{s} = 7\ TeV$, it got necessary to produce centrally MC samples at the same center-of-mass energy of 7 TeV. So far, only Monte Carlo samples at a center-of-mass energy $\sqrt{s} = 10\ TeV$ had been available. Since the available statistics in the AlpGen samples used so far is rather low in particular in the tail of the distribution of the Missing Transverse Energy in events with one isolated lepton, a muon-filtered (at least one muon with $p_T > 15\ GeV$) QCD $b\bar{b}$ MC sample which was generated with the PYTHIA generator (PYTHIAB) was included in the analysis. Because of the large number of generated events, the study of QCD $b\bar{b}$ events with one selected isolated lepton was possible. However, the statistics in the tail of the MET distribution remained very limited. The corresponding electron-filtered sample (at least one electron with $p_T > 15\ GeV$) was included in the analysis as well. The application of the study to real data and the comparison to Monte Carlo of the hereof obtained results made it necessary to introduce the application of a trigger in the Monte Carlo analysis and to apply some cleaning cuts in order to remove so-called bad jets (defined in the object definition chapter) [52]. This results in updated cuts in order to select a pure QCD $b\bar{b}$ control sample¹:

¹These cuts are inspired by the usual SUSY selection cuts in summer 2010 (so far the construction of the control sample has only taken into account the different filters of the Monte Carlo samples used). As the

cut	QCD jetjet PYTHIA - no b-jets	QCD $b\bar{b}$ PYTHIA	$t\bar{t}$
1	29655200	62373.1	39.14
2	29654600 (100 %)	62368.1 (100 %)	39.11 (100 %)
3	29625500 (100 %)	62364.9 (100 %)	39.11 (100 %)
4	5881.83 ($2.0 \cdot 10^{-2}$ %)	4665.69 (7.5 %)	3.19 (8.2 %)
5	768.173 (13.1 %)	773.49 (16.6 %)	2.52 (79.0 %)
6	244.934 (31.9 %)	480.09 (62.1 %)	1.83 (72.7 %)
7	18.3323 (7.5 %)	138.50 (28.9 %)	0.80 (43.7 %)

Table 6.8: QCD jetjet PYTHIA Monte Carlo (any events with b-jets were rejected with the help of the truth information), QCD $b\bar{b}$ PYTHIA and $t\bar{t}$.

1. Cut 1: Trigger: L1_MU6 for muons
2. Cut 2: Rejection of events with bad jets
3. Cut 3: At least one vertex with more than 4 tracks (this cut selects good collision events)
4. Cut 4: Exactly one muon with $p_T > 20 \text{ GeV}$
5. Cut 5: 3 jets. The highest energetic jet must have $p_T > 40 \text{ GeV}$, the second and the third most energetic jets, $p_T > 30 \text{ GeV}$
6. Cut 6: At least one jet with a SV0 weight greater than 8.
7. Cut 7: At least one additional jet with a SV0 weight greater than 5.

The cutflows of these selection cuts are presented in table 6.8 and in table. These cutflows are for 1 pb^{-1} .

The most dramatic reduction of the number of events in the QCD $b\bar{b}$ control sample is due to the change from a center-of-mass energy of $\sqrt{s} = 10 \text{ TeV}$ to $\sqrt{s} = 7 \text{ TeV}$ (this can be seen if comparing the QCD $b\bar{b}$ AlpGen samples at a center-of-mass energy of $\sqrt{s} = 10 \text{ TeV}$ with the samples at 7 TeV). As the statistics in the PYTHIA Monte Carlo samples is considerably bigger, these samples will be used in the following and the AlpGen samples neglected. It is known from comparison to data, that simple scaling of the PYTHIA Monte Carlo samples according to their cross sections does not result in an agreement with data [54] (this it also not expected a priori, as the PYTHIA samples use only leading order of the coupling constant in QCD events). Therefore, an additional scaling factor is usually applied in comparison to data for these datasets. This scaling factor is usually obtained in a low MET and/or low M_T region, where QCD events dominate over other physical processes. The number of $t\bar{t}$ events is completely negligible after all selection cuts. The general Pythia dijet samples (referred to as general QCD PYTHIA Monte Carlo samples in the following) contain processes with light quarks, c-quarks as well as with b-quarks. Therefore, these should be understood as a combination of the previously used QCD $b\bar{b}$ and QCD udsc categories. These Monte Carlo samples can be divided by using the Truth information in a part with b-jets (which correspond to the QCD $b\bar{b}$ Monte Carlo datasets used earlier) and a part without any b-jets (which correspond to the QCD udsc Monte Carlo datasets used earlier). Only

QCD $b\bar{b}$ background in the SUSY signal region will be estimated, it is necessary, that the cuts of the control sample are close to the SUSY selection cuts. The cuts of the control sample differ from the SUSY selection cuts by asking for a non-isolated muon instead of an isolated muon and by not applying any SUSY selection cuts after the jet cuts. Of course, the SUSY selection cuts do not include the two b-tagging cuts. The cuts of the control sample as presented here are the basis for the next chapter.

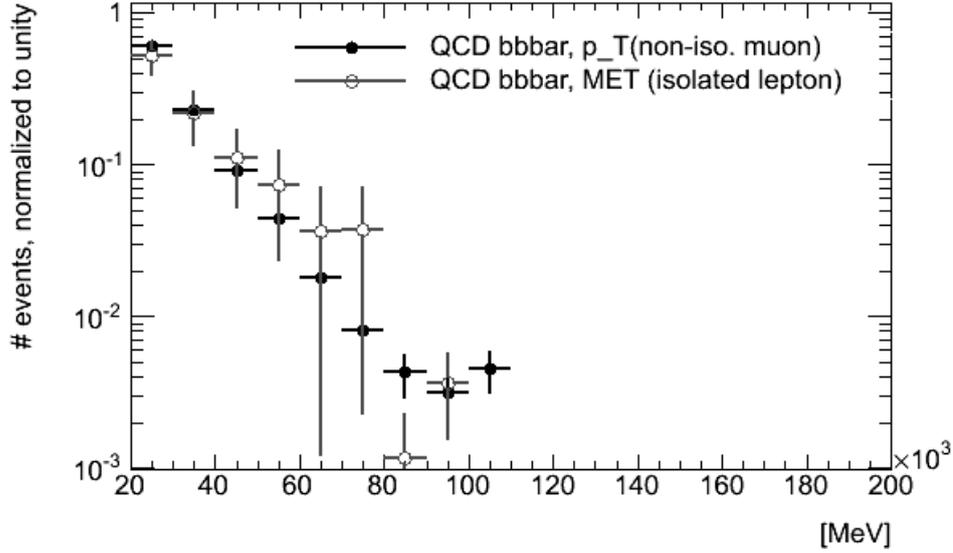


Figure 6.13: p_T (selected non-isolated muon) in comparison to MET (selected isolated lepton). Both distributions are scaled to unity.

the part without any b-jets in the event is shown on the left in table 6.8. The numbers are in agreement with the expectation from the cutflow of the QCD $b\bar{b}$ PythiaB Monte Carlo datasets.

It is interesting to compare the shape of p_T (non-isolated muon), as obtained in the control sample after all selection cuts, with the shape of the distribution of the Missing Transverse Energy. Of particular interest is the MET distribution taken in events where the usual signal region selection cuts have been applied (but without applying MET, M_T and M_{eff} cuts)². If the shape of both distributions is in agreement, then it will be possible to conclude from the p_T distribution the MET distribution. This is interesting in the region of high MET. As the statistics of QCD $b\bar{b}$ events with one selected non-isolated muon is much better than the statistics in QCD $b\bar{b}$ with one selected isolated lepton, it is possible to get a prediction for the tail of the MET distribution of events with one isolated lepton by the distribution of p_T . This prediction can be used to estimate the number of events in the tail of the MET distribution. But this tail corresponds to the SUSY signal region. Therefore, it is possible to estimate the number of QCD $b\bar{b}$ events in the SUSY signal region and so to get a number for the QCD $b\bar{b}$ background in the signal region. However, as can be seen from figure 6.13, the shapes of both distributions are only in rough agreement. This problem is discussed and solved in the next chapter.

²Basically, the following cuts are applied: Trigger, Jet cleaning, vertex cut, exactly one isolated lepton with $p_T > 20 \text{ GeV}$, at least 3 jets with $p_T^{first \text{ jet}} > 40 \text{ GeV}$, $p_T^{third \text{ jet}} > 30 \text{ GeV}$

Chapter 7

Estimation of the QCD $b\bar{b}$ background in the signal region

In the previous chapter a QCD $b\bar{b}$ control was constructed successfully. This control sample is used to get an estimation of the number of QCD $b\bar{b}$ events in the signal region in the following. This is achieved by comparing the shape of the distribution of p_T (non-isolated muon) to the shape of the Missing Transverse Energy distribution. Therefore, the first step is to motivate why these distributions should agree in their shape. The second step will be to prove this with the QCD $b\bar{b}$ Monte Carlo samples. As the statistics in the tail of the MET distribution is essential in the proof of the functionality of this method, PYTHIA Monte Carlo samples are always used in this chapter. All distributions and numbers are for 1 pb^{-1} . It will be seen, that a perfect agreement of the shapes of both distributions cannot be obtained with the unmodified cuts of the control sample. However, it will be shown that with modifications a good agreement is obtained. This agreement in their shapes is used to apply a kind of ABCD-method in order to estimate the QCD $b\bar{b}$ background in the SUSY signal region.

The ABCD-method used in this work uses the similarity of the MET and p_T (non-isolated muon) distributions in their shapes. Due to the similarity of their shapes, it is possible to estimate the tail of the MET distribution by its lower region and the shape of the p_T distribution. Three control regions (A, C and D) are defined in order to do this: A corresponds to the lower region of the MET distribution. The regions C and D are taken in the p_T distribution. All control regions must be dominated by QCD $b\bar{b}$ events. The number of events in the tail of the MET distribution (the tail corresponds to the signal region B) can now be estimated if $\frac{(\text{Number of events in B})}{(\text{Number of events in A})} = \frac{(\text{Number of events in D})}{(\text{Number of events in C})}$ is satisfied (for QCD $b\bar{b}$ events). This application of an ABCD-method is in contrast to a classical ABCD-method which usually uses two uncorrelated variables and the plane spanned by them in order to define control and signal regions.

In the naive application of this method it will be seen that the contamination of other backgrounds than QCD $b\bar{b}$ in one control region is too big. Namely, it is necessary to suppress $t\bar{t}$ events and W +jets events. In order to do so, the cuts for the control sample are again modified. With these modifications, the method is found to work very nicely. The application as well as a critical discussion will mostly be given in the next chapter.

As this chapter contains much information, the reader might want to study the summary at the end of this chapter first in order to get an overview.

7.1 The agreement in their shapes of the distributions of $p_T(\text{muon})$ and MET

In the case of a perfect detector with a complete coverage, no dead or inactive regions, the Missing Transverse Energy in SM processes would come from neutrinos, since neutrinos cannot be detected with the ATLAS detector in contrast to other particles or objects (electrons, muons, jets) which can be detected in principle. However, apart from *true* MET also *fake* MET can occur due to for example dead channels, noisy channels or misreconstruction of particles etc.

If neglecting fake MET and if taking into account only true MET, one could relate the shape of the MET distribution to the shape of the distribution of the transverse momentum of the muon in decays of b-quarks: QCD $b\bar{b}$ events are produced in hadron collisions mainly in s-channel production or in gluon splitting. The bottom quark will combine with other quarks to hadrons. If the bottom quark is accompanied by a light quark like u , d and s , then the decay of this hadron is well described by a weak decay of the bottom quark, the other quarks being only spectator quarks. In most cases, this bottom quark will decay into a charm quark and a virtual W boson, which may decay into a lepton with corresponding neutrino (or to quarks). If neglecting the mass of the lepton, one can expect that the kinematical properties of the lepton and the corresponding neutrino show the same distributions.

Therefore, the shape of $p_T(\text{lepton})$ should be the same as the shape $p_T(\text{neutrino})$. As motivated above, the distribution of $p_T(\text{neutrino})$ corresponds ideally to the distribution of MET. Therefore, one could assume that the shape of $p_T(\text{muon})$ is the same as the shape of MET. As will be shown in detail, this is not directly the case.

Finally, the distributions of $p_T(\text{selected non-isolated muon})$ and of MET (in events with one selected isolated muon or lepton) will be compared in order to get an approximation of the number of QCD $b\bar{b}$ events in the signal region. Some conditions must be satisfied in order to get a reliable estimation of the QCD $b\bar{b}$ background in the signal region. The first condition which must be satisfied, so that this comparison can be done, is that the shape of the MET distributions is the same in events with one selected non-isolated muon, with one selected isolated muon and with one selected isolated lepton. Hereby, the events with one non-isolated muon are selected according to the cuts of the control sample defined in the last chapter. Events with one isolated muon (lepton) are selected in the same way, except for the isolation of the muon (lepton). In figure 7.1 the Missing Transverse Energy in events with a selected isolated muon, isolated lepton or non-isolated muon are compared to each other. This is done after applying the two b-tagging cuts. An agreement of the three distributions is visible within error bars.

The next important condition which must be satisfied is that the two b-tagging cuts do not change the shape of the distributions of MET, since no b-tagging cut is applied in the signal region. Hence, the MET in the event is compared before applying and after applying the two b-tagging cuts in figure 7.2. On the top left this done for events with one non-isolated muon, on the top right for events with one isolated muon and on the bottom for events with one isolated lepton. In the case of events with one isolated muon a poor statistics is available, so that the error bars are large and a conclusion is difficult despite the fact that an agreement is visible within the error bars. In events with one isolated lepton, differences in the tail are visible, but here again the statistics is low. In total both curves in each case are in agreement. The biggest statistics is available in events with one non-isolated muon. Here, a nice agreement can be seen within error bars. In total the assumption is satisfied: The two b-tagging cuts do not vary the shape of the MET distribution dramatically.

The next step is to compare the distributions of MET and of $p_T(\text{non-isolated muon})$ to each other. This is done separately for MET in events with one selected non-isolated muon in figure 7.3, for MET in events with one isolated muon in figure 7.4 and for MET in events

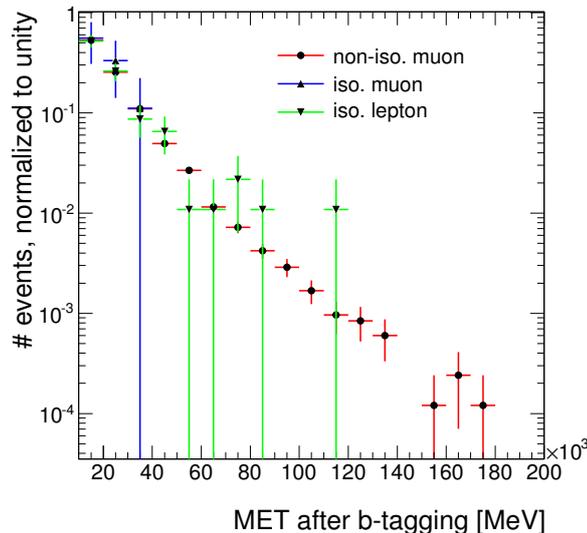


Figure 7.1: MET in the event with one selected non-isolated muon (red), one selected isolated muon (blue), one selected isolated lepton (green) after applying two b-tagging cuts as described earlier. Although a nice agreement of the three curves can be seen, this plots suffers from a very low statistics for isolated muons.

with one selected isolated lepton in figure 7.5. In figure 7.3, in 7.4 and in 7.5 the curves of p_T and of MET are always normalized to unity in the range shown, i.e. between 30. and 200. GeV in order to permit a comparison of both distributions.

In particular, in figure 7.3, MET(in events with one non-isolated muon before applying the b-tagging cuts) is compared to p_T (non-isolated muon after applying the b-tagging cuts). Here, it can be observed that with this way of normalization the shape of both distributions differ. In particular, the MET distribution lies above the p_T distribution in the tail of high energy (a disagreement can be seen around 100. GeV). Actually, this disagreement comes only from the way of normalizing. In order to force a correct normalization an additional MET -cut at the same value as for the p_T -cut (so $MET > 20. GeV$) is introduced in the selection cuts of the control sample. The result of introducing such a cut will be shown below.

Finally, MET should be taken in events with one isolated muon (lepton). Figure 7.4 shows the comparisons between the distribution of MET(isolated muon) (selected with the same cuts as for the control sample but asking for one isolated muon instead of asking for one non-isolated muon) and the distribution of p_T (non-isolated muon). As the statistics in events with one selected isolated muon is low, comparisons of both distributions are not possible. In figure 7.5 the same comparison plot are done for MET in events with one selected isolated lepton. Although the statistics is poor in this case as well, a clear disagreement is visible in nearly all distributions. However, it will be taken care of these problems (low statistics and the disagreement) later.

First, the comparison of p_T (non-isolated muon) and MET(in events with one selected non-isolated muon) is studied further. Introducing an additional MET cut will result a perfect agreement of the shapes of p_T (non-isolated muon) and of the MET(in events with one selected non-isolated muon). Nevertheless, it must be excluded, that the shape of MET is influenced by two additional effects, namely the number of muon in the event and the contribution of jets to the MET.

As explained in the section about the Missing Transverse Energy in the chapter about object definitions, MET_RefFinal is used (the different components of MET_RefFinal are explained in section 4.4.1). MET_RefFinal includes the contribution of various physical objects

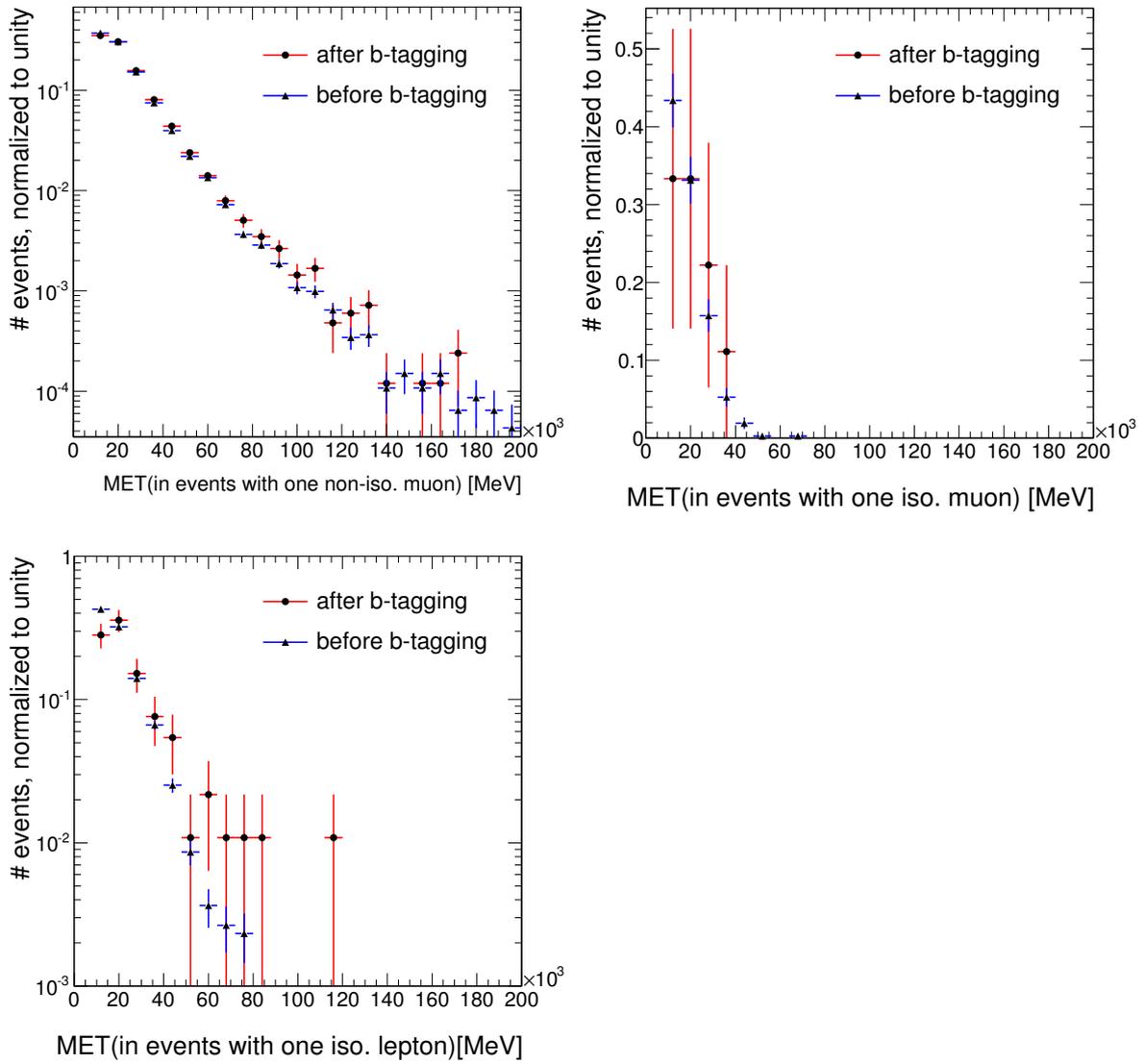


Figure 7.2: Comparison of MET before (blue) and after (red) the b-tagging cuts in events with one selected non-isolated muon (top left), with one isolated lepton (top right) and with one isolated muon (bottom left).

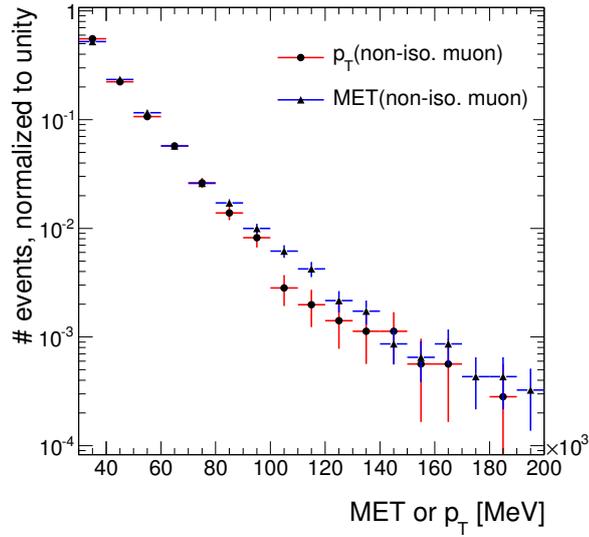


Figure 7.3: MET (blue) is compared to p_T (selected muon) (in red) in events with one selected non-isolated muon. MET is taken before applying the b-tagging cuts as in the signal region, while p_T is taken after applying the b-tagging cuts.

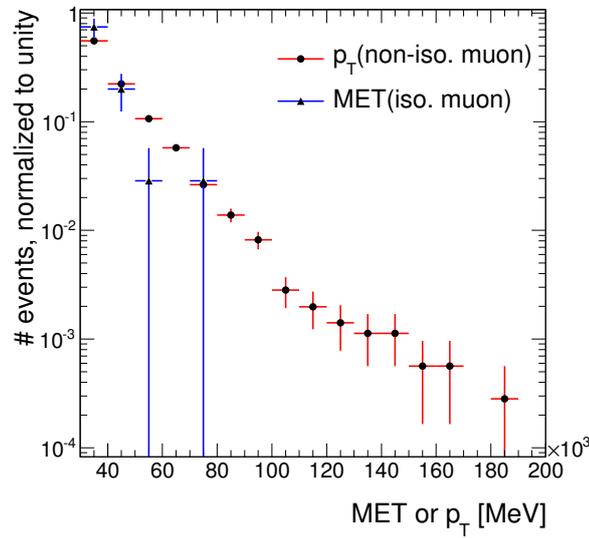


Figure 7.4: Comparison of the distributions of MET(in events with one selected isolated muon) and p_T (non-isolated muon) after applying the b-tagging cuts only for events with one non-isolated muon.

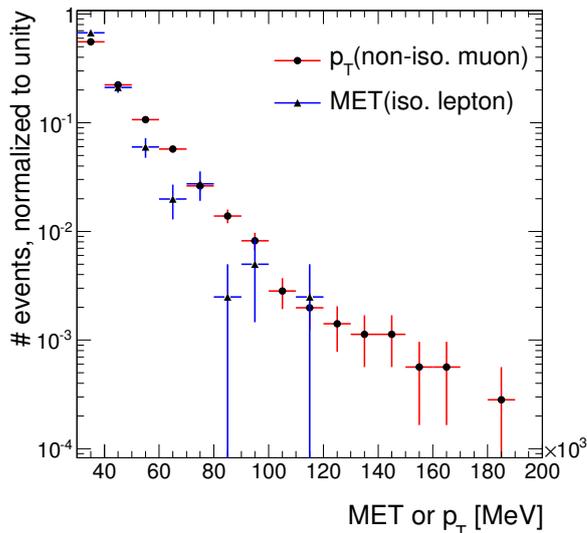


Figure 7.5: Comparison of MET (in events with one selected isolated lepton) with p_T (selected non-isolated muon). Only for events with one selected non-isolated muon both b-tagging cuts were applied.

like electrons, muons and jets.

Figure 7.6 shows the number of reconstructed muons on the left and the components of $MET_{RefFinal}$ on the right. The components of $MET_{RefFinal}$ are compared to the p_T of the selected non-isolated muon, too. The curve of p_T (non-isolated muon) is always above the curve of $MET_{RefFinal}$. This is partly due to the calculation of $MET_{RefFinal}$ which might include more muons than only the selected non-isolated muon, as it can happen that there is more than one muon in the event, for example some low energetic muons or isolated muons in addition to the selected non-isolated muon.

Therefore, it seems so, as if the contribution of the muon term to $MET_{RefFinal}$ do not need to agree with p_T of the selected muon. Apart from this, it can be observed from this plot that the contribution by jets to MET is not negligible in the region around 100 GeV. This indicates, that in this region the MET might not be well approximated by p_T (muon) because of contributions of the jets to the calculation of MET . However, the contributions by jets are usually counterbalanced by the muon term in the calculation of MET (the plot shows only the absolute values of all contribution, but MET is calculated as vectorial sum of its components). Hence, no attention will be paid to the contribution of jets to the MET in the following. In contrast, more muons in the event than the selected muon might influence the MET distribution, because their contributions to the Missing Transverse Energy might partly cancel themselves if these muons are back to back (as the transverse momentum of muons enters the calculation of MET as a vector). And the plot on the left in figure 7.6 shows that indeed - although selecting events with exactly one non-isolated muon - in some selected events more than one muon is present.

Therefore, it will be tested in the following, if also other muons in the event like low- p_T muons or isolated muons should be taken into account in the comparison of a distribution of the transverse momentum of muons to the distribution of MET . For this the sum of $p_{x,y}$ of all muons which fulfill some very basic cuts is used. Hereof p_T (in the following called $|\sum p_{x,y}|$) is calculated. The basic cuts are chosen in agreement with the usual cuts on the muons in the calculation of $MET_{RefFinal}$, namely:

- Only muons within $|\eta| \leq 2.5$ are used.

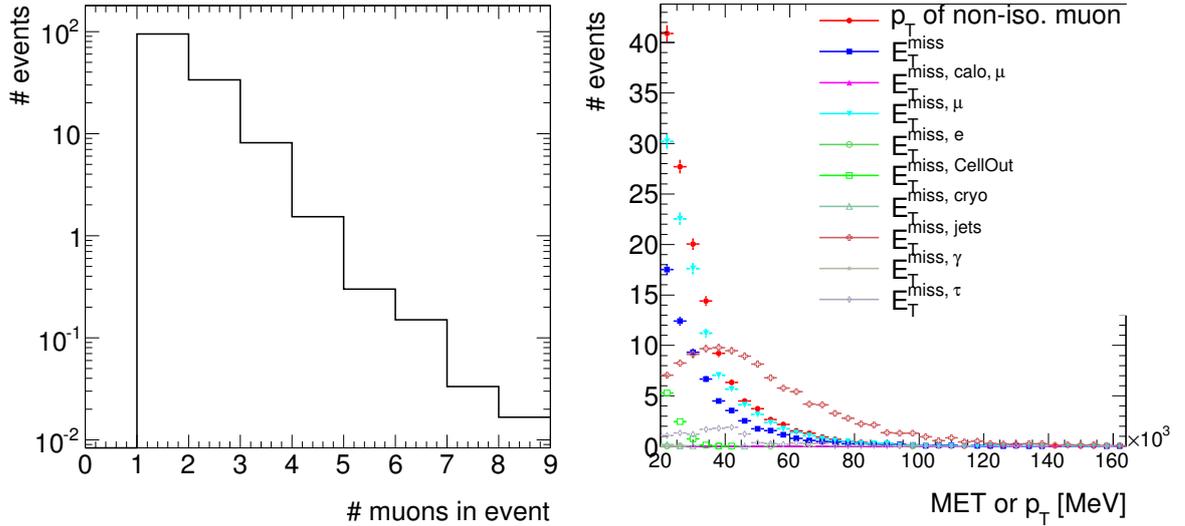


Figure 7.6: On the left: The number of reconstructed muons in the event. There is often more than one reconstructed muon in the event. On the right: The different components of $MET_RefFinal$. MET coming from jets dominates between 40. and 100. GeV over other contributions to MET.

- Only combined muons are used
- The match of tracks in the Muon Spectrometer and of tracks in the Inner Detector must be successful ($0 \leq matchchi2 \leq 100$)
- The muons used must have $p_T > 3$. GeV

The cuts of the control samples were amended by including a cut on MET with $MET > 20$ GeV (which corresponds exactly to the lepton p_T cut) due to normalizing issues. In addition it is tested if $|\sum p_{x,y}|$ approximates better the distribution of MET. With these two improvements a perfect approximation of MET (in events with one selected **non-isolated** muon) by the p_T (selected non-isolated muon) (or by $|\sum p_{x,y}|$) is achieved. This can be seen in figure 7.7. The first plot on the top left compares $MET_RefFinal$, MET_Truth_NonInt (the true MET from non-interacting particles) and $|\sum p_{x,y}|$ directly before the 1-Lepton cut. Both curves, $|\sum p_{x,y}|$ and $MET_RefFinal$, are clearly shifted to each other. The second plot on the top right compares $MET_RefFinal$, MET_Truth_NonInt , $|\sum p_{x,y}|$ and p_T of the selected non-isolated muon after the both jet cuts. A very good agreement is found. This agreement can also be seen in the similar plot after the b-tagging-cuts. It can also be seen, that MET_Truth_NonInt is even better approximated by $|\sum p_{x,y}|$ or p_T as $MET_RefFinal$. This is expected, as the p_T of the selected muon can only approximate the muon contribution to MET. The distributions of p_T and of $|\sum p_{x,y}|$ are in the most cases identical, so that more muons than the selected non-isolated muon have not be taken into account in the following. p_T of the selected non-isolated muon is sufficient.

But the aim of this method is to approximate the MET in the signal region by p_T (selected non-isolated muon). That means, one has to compare MET (in events with a selected **isolated** muon after the jet cuts) with the p_T of the selected non-isolated muon. This is presented in figure 7.8, where PYTHIAB $b\bar{b}$ mu-filtered samples were used. Although the statistics is poor for the events with one selected isolated muon, there seems to be a disagreement of the shapes of MET and the p_T of the muon. Possible reasons for this are discussed in the following, but

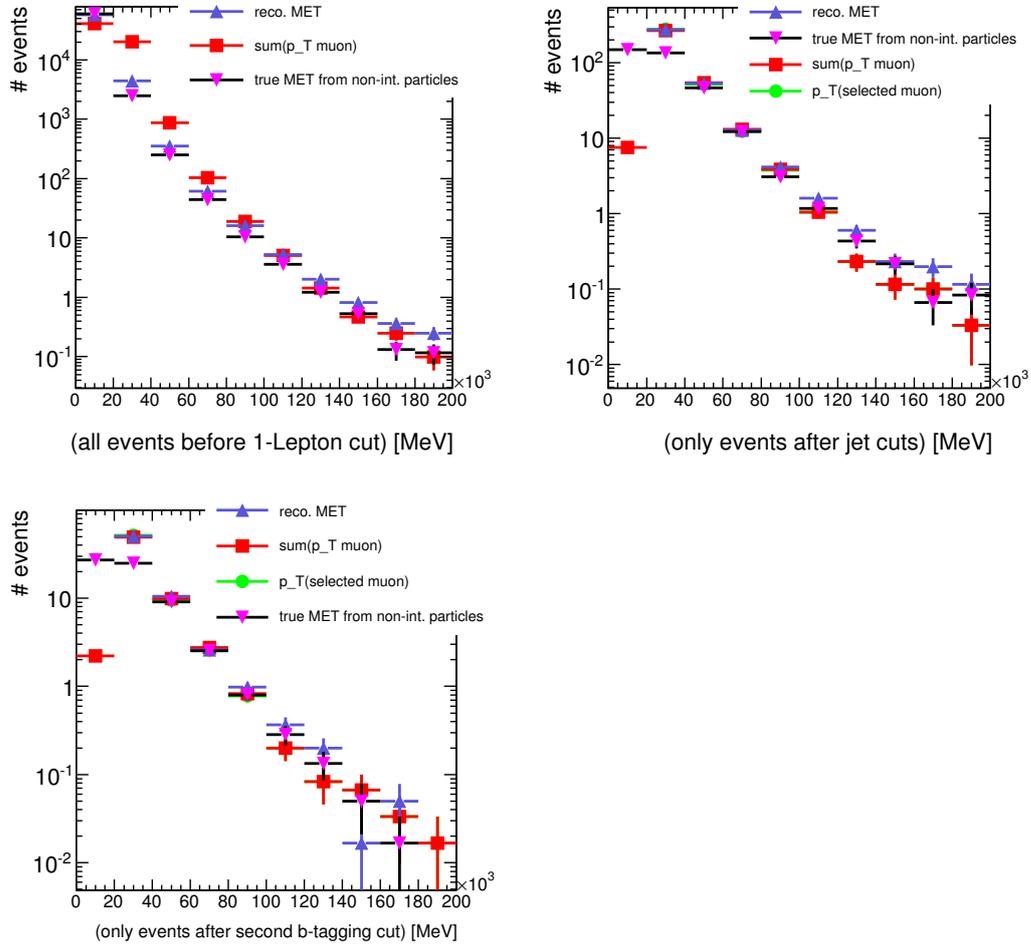


Figure 7.7: The distributions of MET_RefFinal, MET_Truth_NonInt, $p_T(\mu\text{on})$ and $|\sum p_{x,y}|$ are compared in the control sample. Note that no normalization is applied. The top left plot is before the one muon cut, the top right plot after the jets cuts and the plot on the bottom after all cuts (including the b-tagging cuts). A very good agreement is found between the distributions of MET and $p_T(\mu\text{on})$. $|\sum p_{x,y}|$ is identical to $p_T(\mu\text{on})$ in most cases.

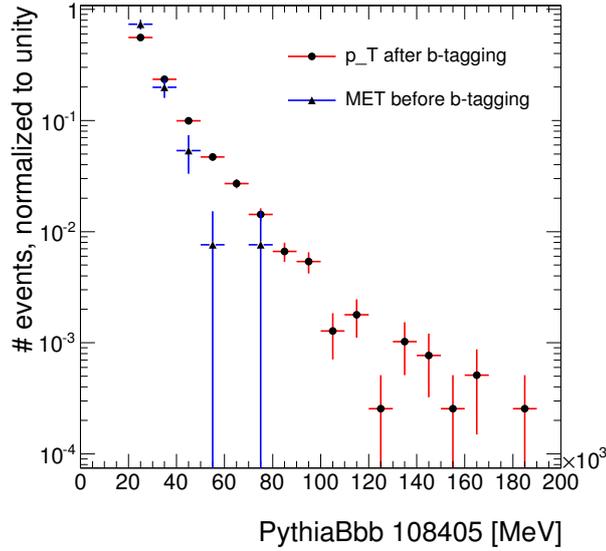


Figure 7.8: Comparison of p_T (non-isolated muon) in the control sample with MET(in events with one isolated muon without applying any b-tagging cuts)

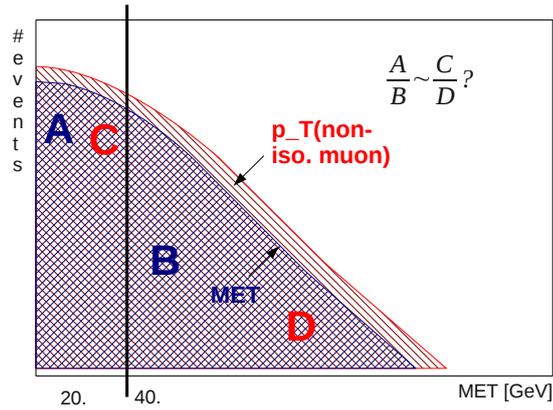


Figure 7.9: definition of control regions and signal region

first the method used to obtain an estimation for the signal region will be described and first results will be presented.

7.2 Definition of control and signal regions

By using the similarity of the distribution of p_T (non-isolated muon) and MET it is possible to estimate from the shape of the distribution of p_T (non-isolated muon) the shape of the distribution of MET. In particular, it is interesting to get a prediction for the tail of the MET distribution. High MET values corresponds (or points) to the signal region in SUSY selection cuts as described in chapter 5. Hence, if the number of QCD $b\bar{b}$ events in the tail of the MET distribution can be estimated, then the size of the QCD $b\bar{b}$ background in the signal region can be predicted. In order to motivate this the following control regions are defined (see figure 7.9 for an illustration of these regions). The cuts of the (modified) control sample are used.

A	B	C	D
274.1 ± 2.1	73.5 ± 1.1	51.75 ± 0.93	13.42 ± 0.48
A/B	C/D		
3.73 ± 0.06	3.86 ± 0.15		

Table 7.1: Population of regions A,B, C and D, which are defined as in section 7.2, if comparing p_T (non-isolated muon) with MET(in events with one non-isolated muon and not applying any b-tagging cuts). Only PythiaB $b\bar{b}$ mu-filtered MC sample are used.

A	B	A/B
2.05 ± 0.19	0.15 ± 0.05	13.67 ± 0.08

Table 7.2: The population of region A and B if taking MET in events with one selected isolated muon and not applying any b-tagging cuts.

- **C:** p_T of selected non-isolated muon after b-tagging, $20 \text{ GeV} < p_T < 40 \text{ GeV}$
- **D:** p_T of selected non-isolated muon after b-tagging, $p_T > 40 \text{ GeV}$
- **A:** MET (in selected events before b-tagging) $20 \text{ GeV} < MET < 40 \text{ GeV}$

The shape of MET distribution and the number of events in the tail shall be estimated in the *signal region*¹:

- **B:** MET (in selected events before b-tagging) $MET > 40 \text{ GeV}$

7.2.1 QCD $b\bar{b}$

A nice test of the similarity of the shapes of p_T and MET can be done in the control sample by trying to estimate the number of events with $MET > 40 \text{ GeV}$ and one selected non-isolated muon (So p_T (non-isolated muon) is compared to MET(in events with one non-isolated muon)). The regions are populated as shown in table 7.1.

As can be seen from these numbers, the number of events in the region B (73.5 ± 1.1 events) is very well approximated by $\frac{N_A \cdot N_D}{N_C} = 71.1 \pm 2.9$ (N_x indicates the number of events in region x). In particular, the ratios C/D and A/B are in perfect agreement as well.

For an approximation of the number of QCD $b\bar{b}$ events in the SUSY signal region in the 1-muon channel², MET has to be taken in events with one selected isolated muon and not in events with one selected non-isolated muon. The number of events in region A and B with MET in events with one selected isolated muon before b-tagging is given in table 7.2.

This results in an estimation for the number of events in the signal region of

$$\frac{N_A \cdot N_D}{N_C} = 0.53 \pm 0.05 \quad (7.1)$$

¹This signal region is not the same as defined in chapter 5. First, the cut on MET is much looser. Second, no M_T or M_{eff} is applied. It is not possible to apply these cuts or to harshen the MET cut at this stage, as the statistics is even too poor in the PYTHIAB samples. Therefore, the quoted numbers in the following have to be understood as upper limits.

²The 1-electron-channel and the 1-muon channel were separated by the SUSY group during first analyses of data. The method presented here will only give an estimation of the QCD $b\bar{b}$ background in the 1-muon channel and is optimized for this channel.

	A	B	A/B	C	D	C/D
ttbar	2.679 ± 0.017	6.544 ± 0.026	0.4093 ± 0.0030	0.634 ± 0.013	0.2171 ± 0.0076	2.92 ± 0.12
W+jets	5.93 ± 0.11	9.68 ± 0.14	0.613 ± 0.014	0.0040 ± 0.0028	0.0020 ± 0.0020	2.0 ± 2.5
Wbb	0.0758 ± 0.0060	0.1397 ± 0.0082	0.543 ± 0.054	0.0074 ± 0.0019	0.0034 ± 0.0013	2.2 ± 1.0
Z+jets	0.645 ± 0.035	0.434 ± 0.029	1.49 ± 0.13	0	0.0020 ± 0.0020	-
QCD udsc	0.33 ± 0.19	0.096 ± 0.089	3.4 ± 3.7	0.13 ± 0.12	0	-

Table 7.3: The population of the various regions by other backgrounds.

events, but only $B = 0.150 \pm 0.050$ events can be found in region B. This is an overestimation by a factor of more than 3. So far only QCD $b\bar{b}$ events have been considered, although other backgrounds might be important, too.

7.2.2 Other backgrounds

The control regions A, C and D and the signal region B are defined as before and the same cuts as before for selecting the events are applied. The number of events in region A, B, C and D for the backgrounds of $t\bar{t}$ (semileptonic and full hadronic event topology), W+bb, W+jets, Z+jets and QCD udsc (with rejection of all events in the QCD udsc with 2, 4,... true b-jets within) is obtained. MET is taken in events with one selected isolated muon and without applying any b-tagging cuts (the MET distribution will be usually used like this in the following). The results are detailed in table 7.3.

In comparison with the number of events of QCD $b\bar{b}$ in all the four regions (see table 7.2 and 7.1), one observes, that events with one selected non-isolated muon in region C and D of the backgrounds $t\bar{t}$, W+bb, W+jets, Z+jets, QCD udsc are well suppressed (this can also be seen at the MET distribution of the various backgrounds in figure 7.10 where all other backgrounds are negligible compared to QCD $b\bar{b}$). In contrast, many events of $t\bar{t}$ and W+jets are found in region A. Since the events of backgrounds like $t\bar{t}$, W+jets, Z+jets, W+bb and QCD udsc cannot be separated from the events of QCD $b\bar{b}$ in real data (without any further cut), the number of QCD $b\bar{b}$ events in the signal region would be overestimated as the number of events in region A enters in the nominator of equation 7.1. As can be seen from figure 7.11, the MET distribution of $t\bar{t}$ events peaks near 40 GeV. Therefore, a possibility to suppress $t\bar{t}$ events is to move region A to lower values.

This section has shown, that two problems have to be solved in order to get an exact estimation of the number of QCD $b\bar{b}$ events in the signal region:

- The similarity between p_T (non-isolated muon) and MET(in events with one selected isolated muon) is not very good \Rightarrow Hereof the next section will take care.
- The contamination of other backgrounds is too big in control region A. Two methods to solve this will be presented in the next sections.

7.3 The angle between a true muon neutrino and the closest reconstructed jet

As can be seen from the number of events in the regions A,B, C and D for QCD $b\bar{b}$ in the previous section, it is not possible with the method presented so far to estimate the number of events in the signal region B through regions A, C and D. Especially the shape of MET in events with one selected isolated muon and in events with one non-isolated muon seems to be different. Since the MET in these events should mainly come from neutrinos, it is

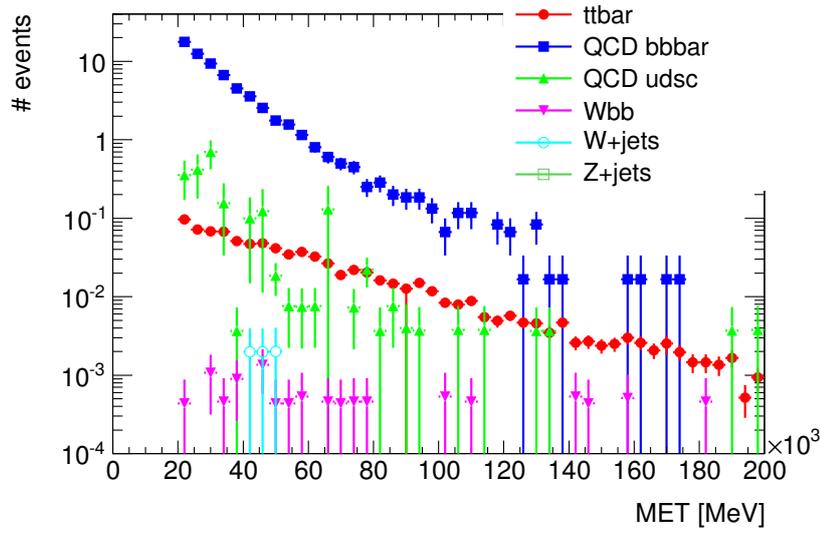


Figure 7.10: The distribution of MET for the various backgrounds in the control sample. All other backgrounds are well suppressed with respect to QCD $b\bar{b}$.

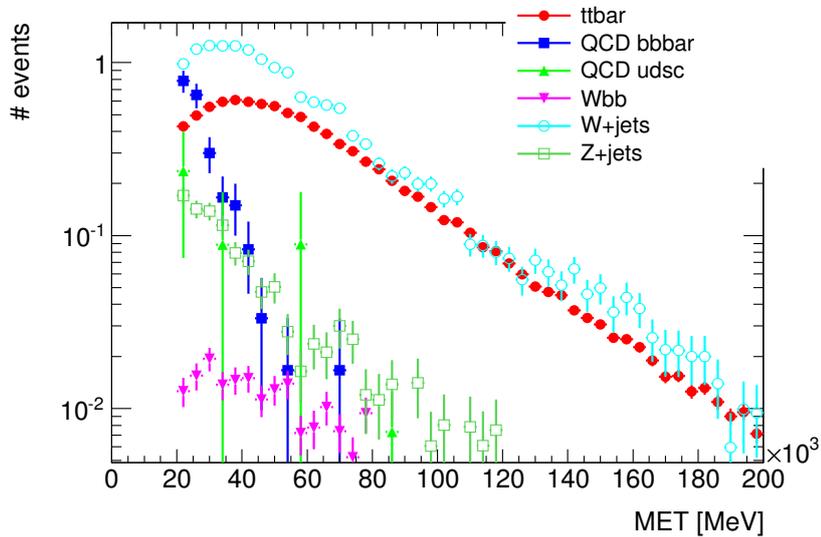


Figure 7.11: The distribution of MET in events with one selected isolated muon. $t\bar{t}$ events peak in region A near 40 GeV.

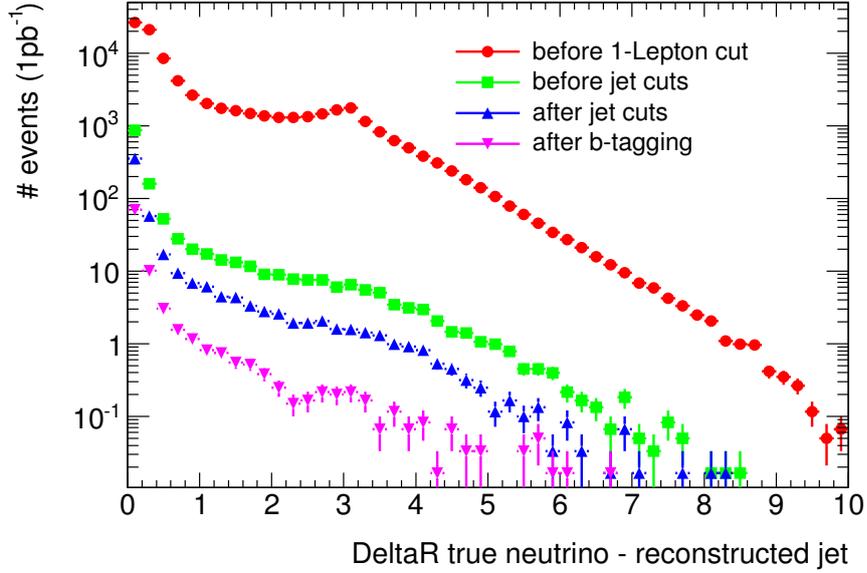


Figure 7.12: The distance between a true muon neutrino and the closest reconstructed jet in events with one non-isolated muon. The red curve shows the distance before the 1 muon cut, the green curve after both the 1 muon and the MET cut. The blue curve is obtained after applying the jet cuts and the pink curve after all cuts.

interesting to check the direction of the neutrinos with respect to the closest reconstructed jet. In figure 7.12 the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between a true muon neutrino and the closest reconstructed jet is shown before the 1-muon-cut (in red), after the 1-lepton and the MET-cut (in green), after the jet cuts (in blue) and after the both b-tagging cuts (in pink).

In the case of events with one selected non-isolated muon a peak near very small values of ΔR can be seen, whereas this behavior is not observed in the case of events with one selected isolated muon (figure 7.13), where a peak can be seen in the region of $\Delta R = 3$. Events with muon neutrinos which are too close to a reconstructed jet might be responsible for the disagreement of the shapes of MET in events with one selected isolated muon and on selected non-isolated muon. In order to get a better agreement of these shapes, non-isolated muons which are too close to a reconstructed jets (in events with one selected non-isolated muon) were rejected. That means that the non-isolated muon is asked to be within $x < \Delta R < 0.4$ to a jet. This was checked for different x of 0.05, 0.1 and 0.2. For $x = 0.2$ only very few events with a non-isolated muon between $\Delta R = 0.2$ and 0.4 are found. These few events lead to bigger error bars. However, there is already an agreement between p_T (of selected non-isolated muon after b-tagging) and MET (in events with one selected isolated muon before b-tagging). In the case of $x = 0.05$ the agreement of the shapes of p_T and MET is not better. In contrast, a very good agreement of both shapes for $x = 0.1$ (see figure 7.14) can be observed. However, the numbers of surviving events with one selected non-isolated muon after all cuts is reduced by a factor of 5 compared to the case of $\Delta R < 0.4$ for non-isolated muons.

Hence, the control sample constructed in chapter 6 was modified with the improvements which were obtained in the last sections:

- The definition of non-isolated muons was changed from $etcone20 > 10 \text{ GeV}$ and $\Delta R < 0.4$ to $etcone20 > 10 \text{ GeV}$ and $0.1 < \Delta R < 0.4$
- An additional MET at the same value as the p_T cut of the muon was introduced

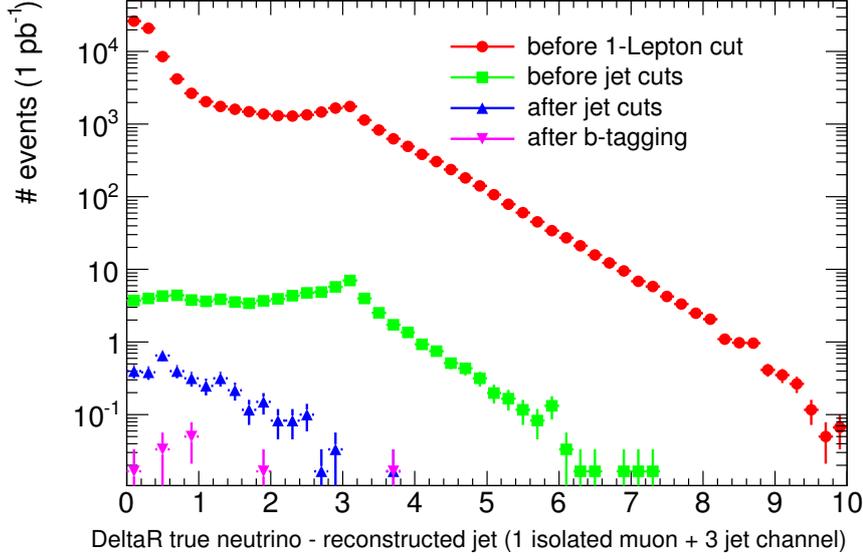


Figure 7.13: The distance between a true muon neutrino and the closest reconstructed jet in events with one selected isolated muon. The same coloring scheme is applied as in figure 7.12

With these modified cuts for the control sample the number of QCD $b\bar{b}$ events in the signal region B is estimated. This results in the numbers of events in the regions A, B, C and D for the QCD $b\bar{b}$ Pythia sample in table 7.4.

The prediction of QCD $b\bar{b}$ events in region B agrees with the number of events found in region B, if taking into account only QCD $b\bar{b}$ events. But it has been shown, that the contamination by other backgrounds is big in region A. If applying the method like this, this would result in an overestimation of the QCD $b\bar{b}$ background. One needs to suppress $t\bar{t}$ and W+jets events further.

	QCD $b\bar{b}$
A	2.05 ± 0.19
B	0.150 ± 0.050
C	11.67 ± 0.44
D	1.13 ± 0.14
A/B	13.7 ± 4.7
C/D	10.3 ± 1.3
$\frac{N_A N_D}{N_C}$	0.20 ± 0.03

Table 7.4: The population of regions A, B, C and D for applying the cuts of the modified control sample for the distribution of p_T (non-isolated muon). The prediction of QCD $b\bar{b}$ events in region B and the number of found events agree within errors.

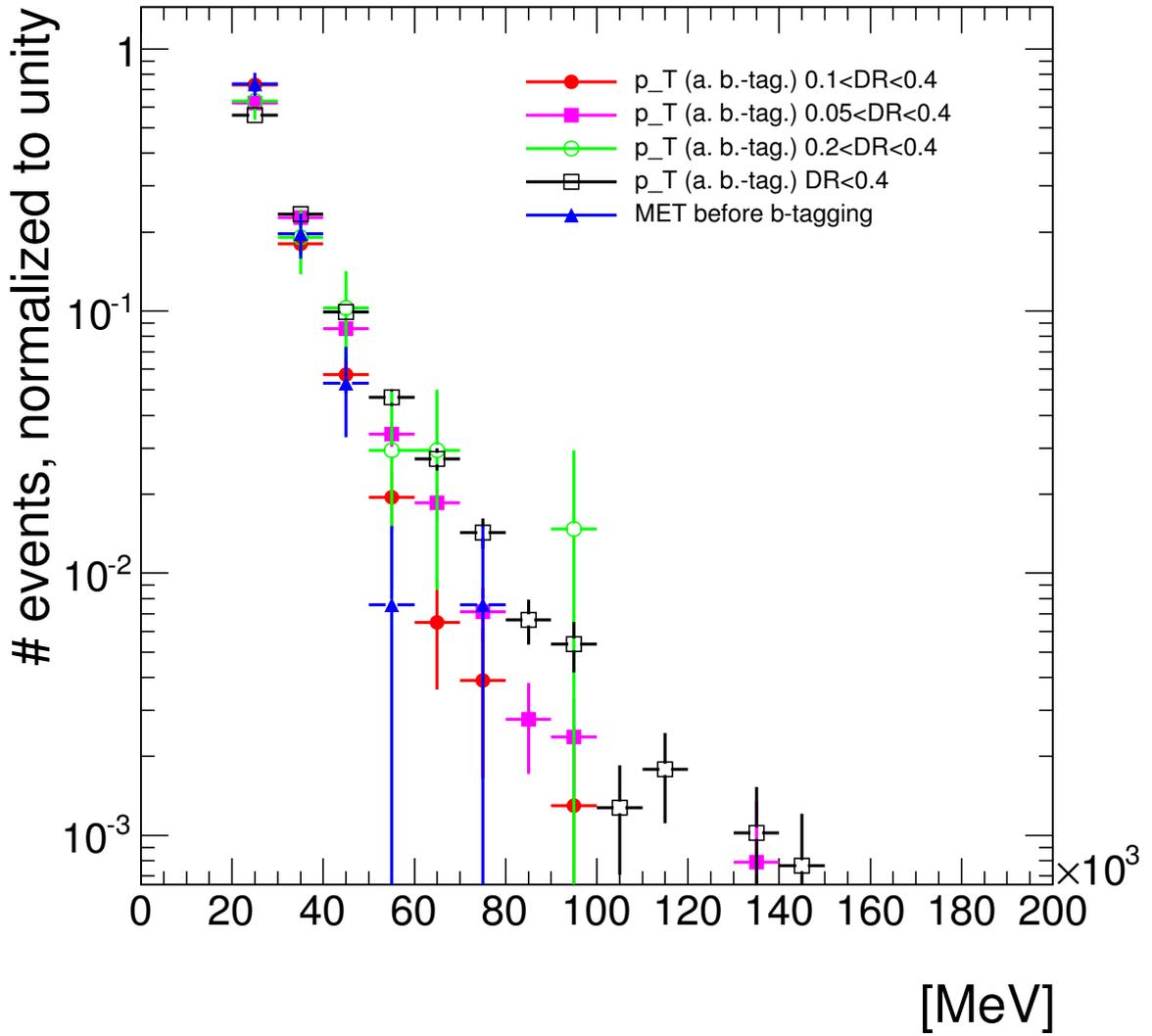


Figure 7.14: p_T of the selected non-isolated muon with different minimal distances to a reconstructed jet in comparison to MET (in events with a selected isolated muon (with $\Delta R > 0.4$) after jet cuts, in blue). A good agreement can be found between p_T with $0.1 < \Delta R < 0.4$ (in red) and MET. This agreement is worse in the case of non-isolated muons with $0.05 < \Delta R < 0.4$ (in pink), with $0.2 < \Delta R < 0.4$ (in green) or with $\Delta R < 0.4$ (in black)

7.4 Suppression of other backgrounds

7.4.1 Lowering the p_T and the MET cut

As shown in figure 7.11 $t\bar{t}$ events with one isolated muon tend to have a higher MET. Therefore, an idea to reduce the $t\bar{t}$ events in the control region A (defined so far as $20 \text{ GeV} < MET < 40 \text{ GeV}$) is to move this control region to lower MET values (Another idea would be to use an additional M_T -cut. It is preferred to avoid this cut, because MET is known to be slightly correlated with M_T). Hence, the control region A is defined as

- **A:** Events with one selected isolated muon and $15 \text{ GeV} < MET < 24 \text{ GeV}$, before b-tagging
and all other regions corresponding:
- **B:** Events with one selected isolated muon and $MET > 24 \text{ GeV}$, before b-tagging
- **C:** Events with one selected non-isolated muon (with $0.1 < \Delta R < 0.4$) and $15 \text{ GeV} < p_T < 24 \text{ GeV}$, after b-tagging
- **D:** Events with one selected non-isolated muon (with $0.1 < \Delta R < 0.4$) and $p_T > 24 \text{ GeV}$, after b-tagging

The number of events for each region and for each background is presented in table 7.5.

	QCD $b\bar{b}$	QCD $u\bar{d}sc$	$t\bar{t}$	W+bb	W+jets	Z+jets
A	10.05 ± 0.41	1.67 ± 0.42	0.912 ± 0.010	0.0289 ± 0.0037	2.075 ± 0.065	0.535 ± 0.033
B	4.94 ± 0.29	1.36 ± 0.37	9.763 ± 0.032	0.224 ± 0.010	16.17 ± 0.18	1.059 ± 0.046
C	25.48 ± 0.65	2.27 ± 0.51	0.2272 ± 0.0079	0.00136 ± 0.00078	0	0
D	12.05 ± 0.45	1.11 ± 0.33	0.1577 ± 0.0060	0.00136 ± 0.00078	0	0
A/B	2.03 ± 0.14	1.23 ± 0.45	0.0934 ± 0.0011	0.129 ± 0.017	0.1283 ± 0.0043	0.506 ± 0.058
C/D	2.115 ± 0.095	2.05 ± 0.76	1.441 ± 0.074	1.00 ± 0.82	-	-

Table 7.5: The number of events in the different regions for the various backgrounds with the modified definitions of the regions.

In region A are found in total (including all backgrounds) 15.27 ± 0.59 events, in region C 27.98 ± 0.83 events and in region D 13.32 ± 0.56 events. Considering only the QCD $b\bar{b}$ events, the number of events in the region B is estimated to

$$N_{B_{QCD \text{ } b\bar{b}}, \text{ estimated}} = \frac{N_{A_{QCD \text{ } b\bar{b}}} N_{D_{QCD \text{ } b\bar{b}}}}{N_{C_{QCD \text{ } b\bar{b}}}} = 4.75 \pm 0.29 \quad (7.2)$$

which is very close to the number of QCD $b\bar{b}$ events found in region B = 4.94 ± 0.29 . However, if one considers all background events, one will again overestimate the QCD $b\bar{b}$ events in the region B:

$$N_{B_{QCD \text{ } b\bar{b}}, \text{ estimated with all backgrounds}} = \frac{N_{A_{\text{all backgrounds}}} N_{D_{\text{all backgrounds}}}}{N_{C_{\text{all backgrounds}}}} = 7.27 \pm 0.46 \quad (7.3)$$

But region B does not correspond to the signal region here, because the MET -cut with $MET > 24 \text{ GeV}$ is too low. An usual cut would be $MET > 40 \text{ GeV}$. Therefore the definition of region B and D is modified.

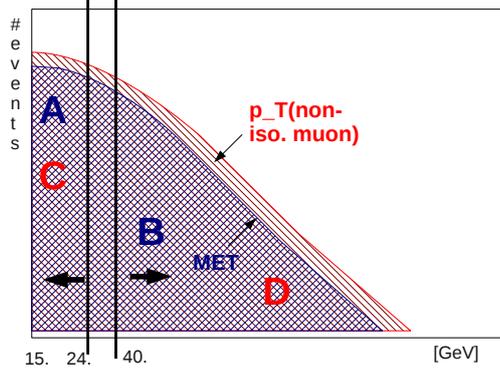


Figure 7.15: Definition of the control and signal regions schematically

	QCD $b\bar{b}$	QCD $u\bar{d}sc$	$t\bar{t}$	W+bb	W+jets	Z+jets
A	10.05 ± 0.41	1.67 ± 0.42	0.912 ± 0.010	0.0289 ± 0.0037	2.076 ± 0.065	0.535 ± 0.033
B	0.62 ± 0.10	0.51 ± 0.20	7.386 ± 0.028	0.1586 ± 0.0087	10.98 ± 0.15	0.512 ± 0.032
C	25.48 ± 0.65	2.27 ± 0.51	0.2272 ± 0.0079	0.00136 ± 0.00078	0	0
D	1.71 ± 0.17	0.15 ± 0.11	0.0510 ± 0.0031	0	0	0
A/B	16.3 ± 2.8	3.3 ± 1.6	0.1234 ± 0.0014	0.182 ± 0.025	0.1890 ± 0.0064	1.045 ± 0.091
C/D	14.9 ± 1.5	15 ± 12	4.45 ± 0.31	-	-	-

Table 7.6: The number of events in all control regions and in the signal region as defined in the text. The contribution of other backgrounds as QCD $b\bar{b}$ is considerable in control region A. This method will be referred to as method 1 in the following.

7.4.2 Estimation of the QCD $b\bar{b}$ events in the signal region (method 1)

The used regions should be rather defined as following with region B as signal region (illustrated in figure 7.15, in the following called *method 1*):

- C: p_T of selected **non-isolated** muon after b-tagging, $15 \text{ GeV} < p_T < 24 \text{ GeV}$
- D: p_T of selected **non-isolated** muon after b-tagging, $40 \text{ GeV} < p_T$
- A: MET (in events with one selected **isolated** muon before b-tagging) $15 \text{ GeV} < MET < 24 \text{ GeV}$
- B: MET (in events with one selected **isolated** muon before b-tagging) $40 \text{ GeV} < MET$

The number of events in each region is again obtained for each background and can be found in table 7.6

Considering all backgrounds (with $N_{A \text{ all backgrounds}} = 15.27 \pm 0.59$, $N_{C \text{ all backgrounds}} = 27.98 \pm 0.83$ and $N_{D \text{ all backgrounds}} = 1.91 \pm 0.20$ from table 7.6), these numbers result in an estimation of

$$\begin{aligned}
 N_{B_{QCD \text{ } b\bar{b}}, \text{ estimated with all backgrounds}} &= \frac{N_{A \text{ all backgrounds}} N_{D \text{ all backgrounds}}}{N_{C \text{ all backgrounds}}} \\
 &= \frac{15.27 \cdot 1.91}{27.98} \\
 &= 1.04 \pm 0.12
 \end{aligned} \tag{7.4}$$

QCD $b\bar{b}$ events in region B. However, only 0.62 ± 0.10 QCD $b\bar{b}$ events are found in the signal region B. This overestimation is due to the still big fraction of W+jets and Z+jets

	QCD bbbar	QCD udsc	ttbar	W+bb	W+jets	Z+jets
A	3.65 ± 0.25	0.33 ± 0.19	0.6956 ± 0.0087	0.0186 ± 0.0030	0.151 ± 0.017	0.0441 ± 0.0094
B	0.333 ± 0.074	0.49 ± 0.20	5.586 ± 0.024	0.1111 ± 0.0073	0.774 ± 0.040	0.031 ± 0.077
C	25.48 ± 0.65	2.27 ± 0.51	0.2272 ± 0.0079	0.00136 ± 0.00078	0	0
D	1.71 ± 0.17	0.15 ± 0.11	0.0510 ± 0.031	0	0	0
A/B	11.0 ± 2.6	0.69 ± 0.48	0.1245 ± 0.0016	0.168 ± 0.029	0.196 ± 0.025	0.24 ± 0.21
C/D	14.9 ± 1.5	15 ± 12	4.45 ± 0.31	-	-	-

Table 7.7: The number of events in each region after applying b-tagging also in region A and B. This method will be referred to as method 2 in the following.

events in the region A. But these W+jets and Z+jets events can be further suppressed by applying a b-tagging, because these events should not contain b-jets.

7.4.3 An additional b-tagging cut in region A (method 2)

Therefore, an additional cut for events with one selected isolated muon is added by using the SV0 b-tagging algorithm: At least one jet with a SV0-weight above 5 is required. This is useful, although the number of QCD bbbar events in region A is reduced by a factor of approximately 3. This can be seen from table 7.7.

W+jets as well as Z+jets are well suppressed by applying this additional b-tagging cut. By considering all backgrounds an estimation for the number of events in the signal region B is obtained (*method 2*, with $N_{A \text{ all backgrounds with b-tagging}} = 4.89 \pm 0.31$, $N_{C \text{ all backgrounds}} = 27.98 \pm 0.83$ and $N_{D \text{ all backgrounds}} = 1.91 \pm 0.20$ from table 7.7):

$$\begin{aligned}
N_{B_{QCD \text{ bbbar, estimated with all backgrounds}}} &= \frac{N_{A \text{ all backgrounds with b-tagging}} N_{D \text{ all backgrounds}}}{N_{C \text{ all backgrounds}}} \\
&= \frac{4.89 \cdot 1.91}{27.98} \\
&= 0.334 \pm 0.043
\end{aligned} \tag{7.5}$$

This is in very nice agreement to the number of events from QCD $b\bar{b}$ in region B = 0.333 ± 0.074 . However, region B does not correspond to the SUSY signal region here, as a b-tagging cut was applied in region A and the SUSY selection cuts does not include this b-tagging cut. Therefore, the result has to be corrected for the b-tagging applied in order to get an estimation of the number of QCD $b\bar{b}$ events in the SUSY signal region. The efficiency of the b-tagging for the QCD $b\bar{b}$ PYTHIAB mu-filtered sample can be found in figure 7.16 for events where no jet selection was applied (in green), for events with at least 3 jets with $p_T > 40, 30, 30 \text{ GeV}$ (in red) and for events with at least 3 jets with $p_T > 60, 30, 30 \text{ GeV}$ (in blue). The events contain exactly one isolated muon. Of these plots, it can be seen that the b-tagging efficiency is 0.46 for a b-tagging weight of 5 in the case of events with one selected isolated muon and at least three jets with $p_T > 40, 30, 30 \text{ GeV}$. The number of estimated events in region B by method 2 can be corrected for the applied btagging:

$$\begin{aligned}
&N_{B_{\text{estimated QCD bbbar events in signal region}}} = \\
&\frac{N_{B_{\text{estimated QCD bbbar events in region B with applied btagging}}}{\epsilon \text{ (btagging weight 5)}} = \\
&\frac{0.334 \pm 0.043}{0.46} = 0.726 \pm 0.093
\end{aligned} \tag{7.6}$$

This result is within errors in good agreement with the earlier found number of QCD $b\bar{b}$ events in the signal region $N_{QCD \text{ bbbar events in signal region}} = 0.62 \pm 0.10$.

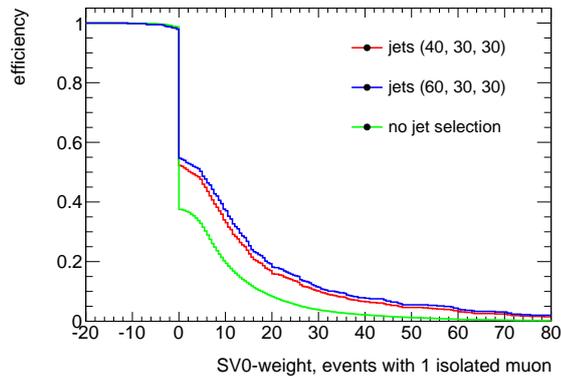


Figure 7.16: The b-tagging efficiency for events with one selected isolated muon. The distribution in green shows the b-tagging efficiency in events with no condition for the jets in this events; at least 3 jets with a $p_T > 40, 30, 30 \text{ GeV}$ must be found in events used for the red distribution and at least 3 jets with a $p_T > 60, 30, 30 \text{ GeV}$ must be found in events for the blue distribution. The efficiency for events with one selected isolated muon and three jets with $p_T > 40, 30, 30 \text{ GeV}$ at a used b-tagging weight of 5 is 0.46.

As the presented method should be data-driven, finally, the b-tagging efficiency should come from the b-tagging group and should be taken from data.

7.5 Summary

This chapter has presented two methods in order to estimate the QCD $b\bar{b}$ background by using the similarity of $p_T(\text{non-isolated muon})$ and MET. The most important steps are summarized below:

- The $p_T(\text{non-isolated muon})$ distribution in the control sample constructed in chapter 6 is compared to MET in the event. This is done for events with one isolated muon as well as for events with one non-isolated muon. It is expected, that the shape of both distributions is the same, as the kinematic of the neutrino and of the muon from the decay of a b-quark should show the same behavior (if neglecting the mass of the muon). However, after normalizing both distributions to unity, both distributions disagree slightly in each case.
- This disagreement is only due to the way of normalizing both distributions. Therefore, a further cut on MET is introduced. The cut value needs to be the same for the muon p_T cut and for the MET cut. A good agreement of both distributions is obtained if MET is taken in events with one non-isolated muon.
- Neither the contribution of further muons apart from the selected muon nor the contribution of jets to the Missing Transverse Energy have an effect on the agreement or disagreement of both distributions.
- Events with one selected isolated muon contain less true muon neutrinos which are very close to a reconstructed jet than events with one selected non-isolated muon. Therefore, the cuts of the control sample are modified again: Only non-isolated muons with a distance of $0.1 < \Delta R < 0.4$ to a reconstructed jet are considered in the construction of the control sample. Now, the distributions of MET and $p_T(\text{non-isolated muon})$ are also similar if the MET is taken in events with one selected isolated muon.

- The tail of MET in events with one selected isolated muon can be estimated by the shape of p_T (non-isolated muon) (taken in the control sample). As the tail of the MET distribution corresponds to the SUSY signal region, the number of QCD $b\bar{b}$ events in the SUSY signal region can be estimated by using the similarity of both distributions and by applying a kind of ABCD method.
- The low and the high energetic part of the p_T distribution and the low energetic part of the MET distribution are taken as control regions.
- The exact definition of the control regions needs to be chosen carefully in order to avoid and in order to reduce the contamination of the low energetic control region of the MET distribution by $t\bar{t}$ and W+jets events.
- Therefore, the following control regions are defined (**method 1**):
 - C: p_T of selected **non-isolated** muon after b-tagging, $15 \text{ GeV} < p_T < 24 \text{ GeV}$
 - D: p_T of selected **non-isolated** muon after b-tagging, $40 \text{ GeV} < p_T$
 - A: MET (in events with one selected **isolated** muon before b-tagging) $15 \text{ GeV} < MET < 24 \text{ GeV}$
and as signal region:
 - B: MET (in events with one selected **isolated** muon before b-tagging) $40 \text{ GeV} < MET$
- The QCD $b\bar{b}$ background is estimated by $\frac{(\text{Number of events in A}) \cdot (\text{Number of events in D})}{(\text{Number of events in C})}$
- However, the QCD $b\bar{b}$ background is overestimated due to a considerable contamination of control region A by W+jets events. This contamination can be reduced by applying an additional and low b-tagging cut in region A (**method 2**). However, this b-tagging also affects region B, so that region B does not correspond to the signal region any longer (as the SUSY selection cuts do not contain a b-tagging cut). Therefore, the estimation of the QCD $b\bar{b}$ background has to be corrected by the b-tagging efficiency. The QCD $b\bar{b}$ background is estimated accurately by method 2, but the b-tagging efficiency must be known and is taken preferably from a data-driven method.

Chapter 8

QCD $b\bar{b}$ background estimation in data

8.1 Analysing data

In the analysis of data some additional items have to be considered compared to an analysis based on Monte Carlo. As explained in the introduction chapter about ATLAS, it is not possible to store all collision data. A *trigger* system is used in order to reduce the amount of data. Various trigger decision chains are available. Some triggers would allow too many events to pass them. Therefore, these triggers have to be *prescaled*, so only a fraction of all events, which would pass the trigger theoretically, is recorded. In SUSY studies unprescaled triggers are preferred (if possible, difficult for control regions where often loose cuts are applied), in order to keep all interesting signal events. The trigger efficiency is defined as the number of events passing the trigger divided by the number of all events passing the Minimum Bias Trigger. The efficiency can be plotted with respect to interesting variables (in figure 8.1 a turn-on curve with respect to muon p_T is shown). Near the cut values of a trigger the efficiency is usually not at its maximum value, but increases strongly. This is the turn-on curve, which is followed by the region of maximum efficiency. Here, the trigger efficiency does not usually change largely. It is preferred to work in this region with the maximal trigger efficiency. The various triggers used are noted in table 8.1 for each data-taking period. The periods correspond to different machine (or detector) operation modes.

Apart from the trigger a *GoodRunList* is used in the analysis to select the runs and their parts, which can be used in the analysis. In the case of SUSY studies, the full functionality of the detector is asked apart from having collisions: Stable beams must have been declared, the high voltage of detector must be in place, solenoid and toroid magnets must be on and at their nominal fields. Further, the subdetector response must agree with their expected behavior. Otherwise, the precise measurement of electrons, muons, jets and MET which are used here would not be possible. The GoodRunLists which are used here were prepared by Tommaso Lari of the SUSY group with the help of official ATLAS tools [49]. The integrated luminosity of the data taking periods has been calculated with official ATLAS tools in agreement with these GoodRunLists, too [50].

Dead layers or regions in the electromagnetic calorimeter represent a problem in the reconstruction of electrons. Therefore, an event will not be used if the electron is hitting such a region (OTx-veto¹).

Cosmic muons can mimic the presence of collision muons. Such events are rejected as well by using a cut on the longitudinal impact parameter of the muon with respect to the

¹OTx are optical transmitters for the ATLAS Liquid Argon Calorimeter front-end electronics readout system. As some OTx have stopped working, these region cannot be used for a good electron reconstruction.[51]

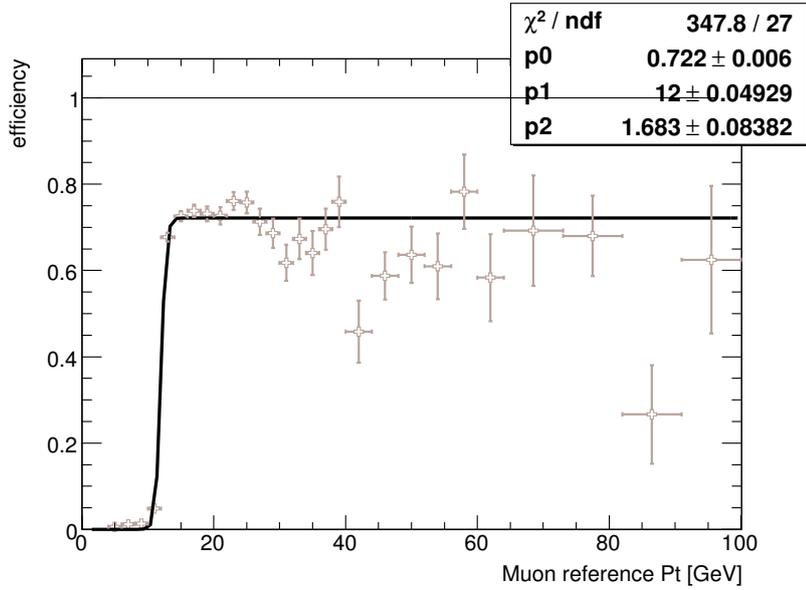


Figure 8.1: Trigger efficiency with respect to p_T of the muon for the EF_mu10 trigger in the barrel. This plot is from [53].

period	integrated luminosity	trigger
A-B	8 nb^{-1}	L1_MU6
C	8 nb^{-1}	L1_MU6
D	288 nb^{-1}	L1_MU6
E	937 nb^{-1}	EF_mu10_MSonly or EF_mu10
F	1711 nb^{-1}	EF_mu10_MSonly or EF_mu10
G	5655 nb^{-1}	EF_mu13
H	7046 nb^{-1}	EF_mu13

Table 8.1: Integrated luminosity and recommended trigger in the muon channel for each data-taking period.

primary vertex (z_0 (w.r. to primary vertex)).

The requirement of medium electron was modified to *RobustMedium*, as comparison between data and Monte Carlo showed discrepancies in shower shape distributions [28]. The cuts on the corresponding variables were loosened in the algorithm. Some other corrections from the comparison with data have also been applied (for example concerning the modeling of electronic noise in the hadronic calorimeter).

8.2 QCD $b\bar{b}$ control sample in 1 muon + 3 jets channel

In this chapter only data which was taken during period A-F has been used. The modified selection cuts are summarized in the following. The cutflow for data of period A - F can be found in table 8.2.

1. GoodRunList (for data) and trigger: L1_MU6 for muons (period A-D), EF_mu10 (period E-F)
2. Jet quality cuts (see explanation in object definition chapter 4.3.2)
3. At least one vertex with more than 4 tracks

	period A - D	period E	period F	total: period A - F
cut 1	5037160	316104	650367	6003631
cut 2	5036610	316072	650314	6002996
cut 3	4673150	316027	650247	5639424
cut 4	902	2366	4462	7730
cut 5	901	2364	4458	7723
cut 6	899	2357	4452	7708
cut 7	288	790	1506	2584
cut 8	63	230	446	739
cut 9	25	97	195	374
cut 10	7	27	49	83
MC	10.67 ± 0.23	32.88 ± 0.72	60.0 ± 1.3	103.6 ± 2.3

Table 8.2: Cutflow for period A - F. The last row indicates the number of events expected from MC.

This cut selects good collision candidates.

4. One non-isolated muon² with $p_T > 15. \text{ GeV}$
5. Veto events with crack electrons (see object definition chapter 4.1.1), OTx-veto (explained in the last section)
6. $|z_0(\text{w.r. to primary vertex})| < 10 \text{ mm}$
7. $MET > 15. \text{ GeV}$
The MET definition recommended in August by the JetEtmisss group was used: $MET = MET_LocHadTopo + MET_MuonBoy - MET_RefMuonTrack$.
8. 3 jets with $p_T > 40., 30., 30. \text{ GeV}$
9. At least one jet with a SV0 weight above 8
10. At least one more jet with a SV0 weight above 5

83 events in data (corresponds to regions C + D + range in p_T between both regions) are available in the control sample already with an integrated luminosity of 2.9 pb^{-1} (period A-F). As the control sample should be dominated by QCD $b\bar{b}$ events, the number of events after all cuts in data should agree with the number of events expected from QCD $b\bar{b}$ Monte Carlo samples. The comparison of the last two rows in 8.2 indicates that the data is not described perfectly by Monte Carlo, but that an additional scaling factor has to be applied to Monte Carlo QCD Pythia datasets. This behavior is known from other recent studies [54]. In particular, it is not expected that the PYTHIA QCD Monte Carlo samples describe perfectly the absolute normalization factor of the cross section to data, because these Monte Carlo samples are simulated at leading order in the strong coupling constant. This scaling factor is usually obtained in QCD dominated control regions (like low MET and low M_T). The numbers in table 8.2 indicate a scaling factor of $\frac{83}{103.6} = 0.80$ between data and QCD $b\bar{b}$ PythiaB Monte Carlo. Further, the distributions in the control sample are expected to show an agreement between data and QCD $b\bar{b}$ Monte Carlo. After scaling Monte Carlo to data a general agreement can be seen between data and Monte Carlo in the distributions of MET, M_T , $p_T^{\text{first jet}}$, $p_T^{\text{second jet}}$ and $p_T(\text{selected muon})$ in figures 8.2, 8.3. Slight disagreements are

²*etcone20* $> 10. \text{ GeV}$, $0.1 < \Delta R < 0.4$

	period A - D	period E	period F	total: period A - F
cut 1	5037160	316104	650367	6003631
cut 2	5036610	316072	650314	6002996
cut 3	4673150	316027	650247	5639424
cut 4	4918	13412	25690	44020
cut 5	4918	13412	25689	44019
cut 6	4890	13386	25625	43901
cut 7	1829	4976	9908	16715
cut 8	21	32	83	136

Table 8.3: Cutflow for period A - F for events with one selected isolated muon.

visible mainly in the tails, but there the statistics in data is low and therefore, the shape of the distributions is not reliable.

8.3 QCD $b\bar{b}$ background in data

The number of QCD $b\bar{b}$ events in the signal region is estimated by using method 1 and method 2 as defined in the last chapter. Hereby, the signal region is defined as in the last chapter with $MET > 40 GeV$. Although cuts on a signal region with a higher MET cut (for example $MET > 60 GeV$) and additional cuts on M_T and M_{eff} are now used by the SUSY group, the signal region used here is still of interest. First, it gives an upper limit on the number of QCD $b\bar{b}$ events in the signal region with tighter definitions. Second, it is difficult to verify the method presented with respect to Monte Carlo with higher MET cut or more cuts, as then the Monte Carlo statistics is poor even if PYTHIA Monte Carlo datasets are used. As explained in the previous chapter, the shape of the p_T distribution in the control sample is compared to the shape of the MET distribution (in events with one selected isolated muon). In order to obtain this MET distribution, the same cuts as for the control sample, but without the two last b-tagging cuts, are applied and an isolated muon is used in place of the non-isolated muon in cut 4. The number of events which survive these cuts, is summarized in table 8.3. It can be seen that a large statistics of 136 events after all cuts is available already for period A-F.

8.3.1 Method 1 (without b-tagging for isolated muons):

The distributions of muon p_T (in events with selected muon after all cuts in the control sample) and of MET (in events with one selected isolated muon) are used to obtain the number of events in regions A, B, C and D:

- Control region A: $15 GeV < MET < 24 GeV$
- Control region C: $15 GeV < p_T < 24 GeV$
- Control region D: $p_T > 40 GeV$
- Signal region B: $MET > 40 GeV$

B is the signal region. The numbers are shown together with the Monte Carlo expectation in table 8.4.

Regions C and D in table 8.4 indicate again a scaling factor of approximately 0.71 (region D) - 0.78 (region C) between data and Monte Carlo. The number of QCD $b\bar{b}$ events in the signal region can be estimated by applying formula 8.1 of the previous chapter.

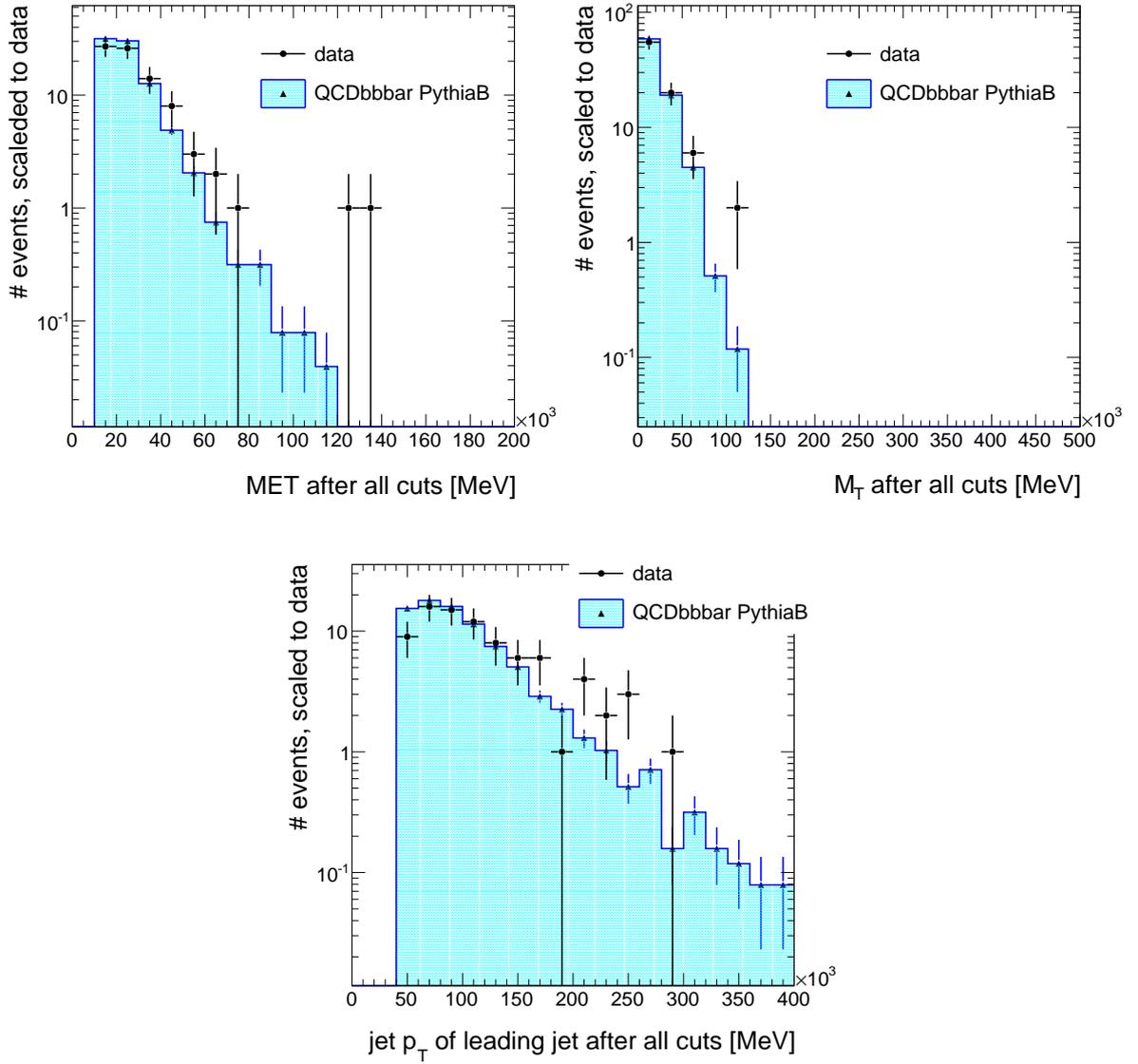


Figure 8.2: Comparisons of the distributions of MET , M_T , $p_T^{first\ jet}$ between data and Monte Carlo. Only the QCD $b\bar{b}$ PythiaB Monte Carlo dataset is used. Monte Carlo is scaled to data.

	A	B	C	D
data	43 ± 6.6	50 ± 7.1	64 ± 8.0	4 ± 2
MC (all backgrounds)	45.0 ± 1.7	59.49 ± 0.80	82.5 ± 2.4	5.64 ± 0.60

Table 8.4: The number of events in region A, B, C and D for period A-F. Top: data, bottom: MC scaled to luminosity of period A-F ($2.9\ pb^{-1}$).

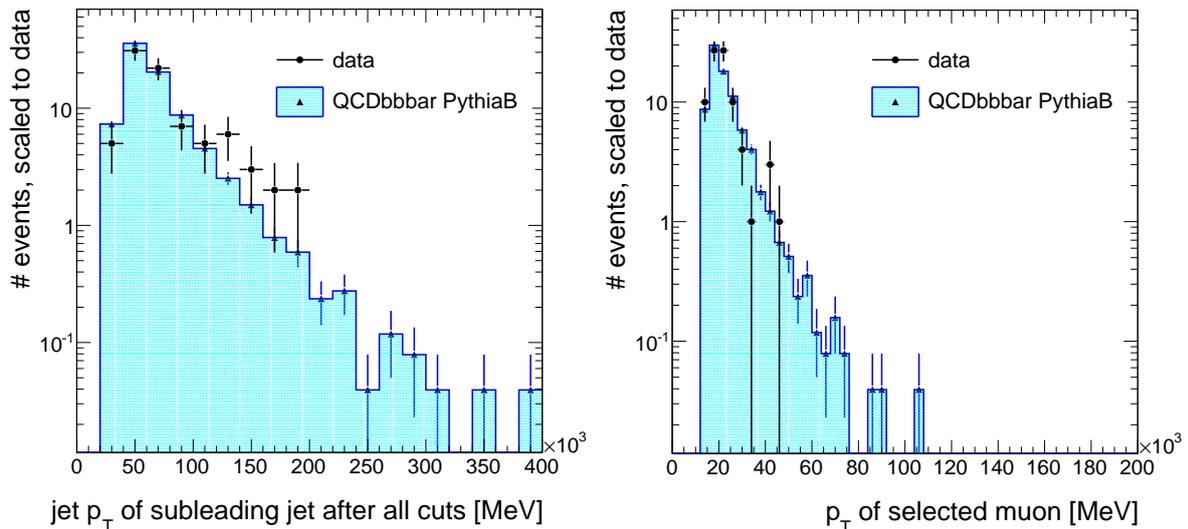


Figure 8.3: Comparisons of the distributions of $p_T^{second\ jet}$ and $p_T(\text{selected muon})$ between data and Monte Carlo. Only the QCD $b\bar{b}$ PythiaB Monte Carlo dataset is used. Monte Carlo is scaled to data.

	A	B	C	D
data	11 ± 3.3	15 ± 3.9	64 ± 8	4 ± 2
MC (all backgrounds)	14.43 ± 0.91	21.61 ± 0.68	82.5 ± 2.4	5.64 ± 0.60

Table 8.5: The number of events in region A, B, C and D for period A-F and method 2. Top: data, bottom: MC scaled to luminosity of period A-F ($2.9\ pb^{-1}$)

$$\begin{aligned}
 \frac{N_A \cdot N_D}{N_C} &= \frac{(Number\ of\ events\ in\ A) \cdot (Number\ of\ events\ in\ D)}{(Number\ of\ events\ in\ D)} \\
 &= 2.7 \pm 1.4
 \end{aligned}
 \tag{8.1}$$

From Monte Carlo, the estimated QCD $b\bar{b}$ background is

$$\frac{N_{A_{MC}} \cdot N_{D_{MC}}}{N_{C_{MC}}} = 3.08 \pm 0.34
 \tag{8.2}$$

whereas the true QCD $b\bar{b}$ background in Monte Carlo datasets in the signal region is:

$$N_{B_{QCD\ b\bar{b}ar,MC}} = 1.83 \pm 0.30
 \tag{8.3}$$

As pointed out in the last chapter, the contamination from other backgrounds (in particular W+jets) others than QCD $b\bar{b}$ is expected to be big in the control region A. A possibility is to apply an additional b-tagging cut in region A, for events with one isolated muon.

8.3.2 Method 2 (with b-tagging for isolated muons), period A - F

The numbers in table 8.5 are found if one applies an additional b-tagging in region A. The b-tagging is carried out as explained in the last chapter: Only events with at least one b-tagged jet with a b-tagging weight larger than 5 are selected.

The statistics is reduced by 74 % in region A in data by this additional b-tagging cut. $N_{B_{QCD\ b\bar{b}ar,MC}} = 0.98 \pm 0.22$ QCD $b\bar{b}$ events are expected in the region B from Monte Carlo. The estimation in data yields:

$$\frac{N_A \cdot N_D}{N_C} = 0.69 \pm 0.41 \quad (8.4)$$

These 0.69 ± 0.41 estimated QCD $b\bar{b}$ events in data have to be compared to number of estimated QCD $b\bar{b}$ events of Monte Carlo:

$$\frac{N_{AMC} \cdot N_{DMC}}{N_{CMC}} = 0.99 \pm 0.13 \quad (8.5)$$

These numbers are - within their errors and considering a QCD scaling factor of 0.71 or 0.78 - in good agreement. However, this results has to be corrected again for the applied b-tagging in region A. Assuming a b-tagging efficiency of 50 %, 1.38 ± 0.82 QCD $b\bar{b}$ are estimated in the signal region in data. This 50 % is a rough estimate for a b-tagging weight of 5 from the efficiency plots in chapter 7. Finally, however, the b-tagging efficiency with errors need to be taken from the b-tagging group.

8.4 Update of the jet cuts

In order to be in agreement with the latest SUSY selection cuts, the jet cuts are modified to

- At least three jets
- $p_T^{first\ jet} > 60\ GeV, p_T^{third\ jet} > 30\ GeV$

As shown in chapter 7 about the construction of a QCD $b\bar{b}$ control sample, modifications to the jet p_T requirements influence the statistics significantly. Therefore, both methods, presented in the last chapter, are applied to data in detail with modified jet cuts. These jet cuts correspond to the latest recommendation of the SUSY signal region as defined by the ATLAS SUSY working group. In table 8.6 and in table 8.7 the number of events in all regions for all backgrounds and for both methods can be found. The numbers for Monte Carlo are scaled to $2.95\ pb^{-1}$. In the last column, the number of events in data in each region is noted and next to it the corresponding number of Monte Carlo events including all backgrounds. The scaling factor between data and QCD $b\bar{b}$ Monte Carlo can be obtained in the control regions C and D and it is found to be 0.706 ± 0.095 for region C and 0.71 ± 0.36 for region D (this combines to a scaling factor of 0.71 ± 0.19). The contribution of W+jets is not negligible in region A for method 1.

- **Method 1:**

The number of QCD $b\bar{b}$ events in region B is estimated by using the number of data events in region A, C and D:

$$N_{QCD\ BG, estimated} = \frac{N_{events\ in\ A} \cdot N_{events\ in\ D}}{N_{events\ in\ C}} = 2.1 \pm 1.1 \quad (8.6)$$

This is in agreement with the estimated number of QCD $b\bar{b}$ events in Monte Carlo.

$$N_{QCD\ BG, estimated} \stackrel{MC+ scaling\ factor}{=} 1.61 \pm 0.47 \quad (8.7)$$

1.22 ± 0.39 QCD $b\bar{b}$ events are expected by using the number of QCD $b\bar{b}$ events in region B and applying the scaling factor in Monte Carlo.

- **Method 2:**

An additional b-tagging cut is applied in region A. This reduces the number of events in data to 6 ± 2.5 events found. This results in an estimation of the QCD $b\bar{b}$ background with applied b-tagging:

$$N_{QCD\ BG, estimated} = \frac{N_{events\ in\ A} \cdot N_{events\ in\ D}}{N_{events\ in\ C}} = 0.4 \pm 0.3 \quad (8.8)$$

By using Monte Carlo, it is estimated:

$$N_{QCD\ BG, estimated} \stackrel{MC}{=} 0.75 \pm 0.10 \quad (8.9)$$

This is in agreement with data if again the scaling factor of 0.71 ± 0.19 , which was found in the control sample, is applied:

$$N_{QCD\ BG, estimated} \stackrel{MC + scaling\ factor}{=} 0.53 \pm 0.16 \quad (8.10)$$

0.66 ± 0.17 QCD $b\bar{b}$ events should be estimated (according to the number of QCD $b\bar{b}$ events in region B multiplied by the scaling factor of the control sample (0.71 ± 0.19)). These results agree within errors. The results can be corrected again for the b-tagging applied. Then, the estimation of the QCD $b\bar{b}$ background in data yields:

$$N_{QCD\ BG, estimated, corrected\ for\ b\text{-tagging}} = \frac{0.4 \pm 0.3}{0.5} = 0.8 \pm 0.6 \quad (8.11)$$

and in Monte Carlo:

$$N_{QCD\ BG, estimated, corrected\ for\ b\text{-tagging}} \stackrel{MC + scaling\ factor}{=} \frac{0.53 \pm 0.16}{0.5} = 1.06 \pm 0.32 \quad (8.12)$$

Both data and Monte Carlo estimations are in agreement with the 1.22 ± 0.39 QCD $b\bar{b}$ events in the signal region B which are expected in the signal region by Monte Carlo.

- **Plots in the control sample:**

In the control sample it is expected that the data is well described by the simulated QCD $b\bar{b}$ Monte Carlo samples. This is verified in some important distributions (figures 8.4,8.5). The distributions of the Missing Transverse Energy and of p_T (selected muon) are used both in the estimation of the number of QCD $b\bar{b}$ events. A good agreement is found in the distribution of MET. Although a little excess of data seems to be visible in the tail, the statistics is poor in that region. The distribution of p_T (selected muon) shows a good agreement as well. In general, the agreement of data and Monte Carlo is good in the distributions of M_T , $p_T^{first\ jet}$, $p_T^{second\ jet}$ and the number of muons in the event, too. However, the distribution of the number of jets in the event indicates that more jets are present in the event in data than in Monte Carlo. Pile-up in data can cause this. Another explanation could be that PYTHIAB simulates only particle interactions with two incoming particles and two outgoing particles for QCD events.

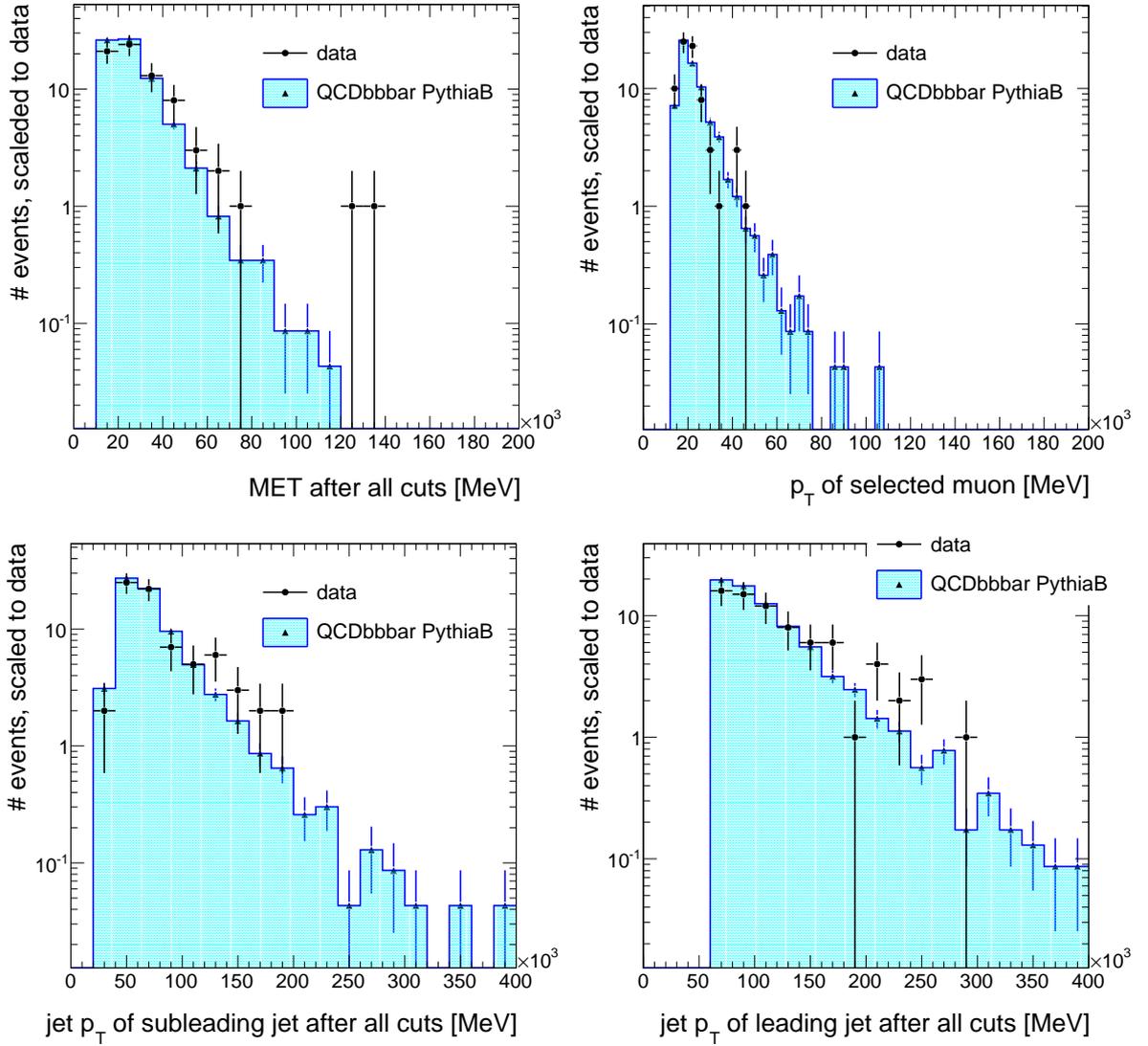


Figure 8.4: Comparison between data and Monte Carlo. Monte Carlo is scaled to data. Top left: Missing Transverse Energy, top right: $p_T(\text{selected muon})$, bottom left: $p_T^{\text{second jet}}$, bottom right: $p_T^{\text{first jet}}$

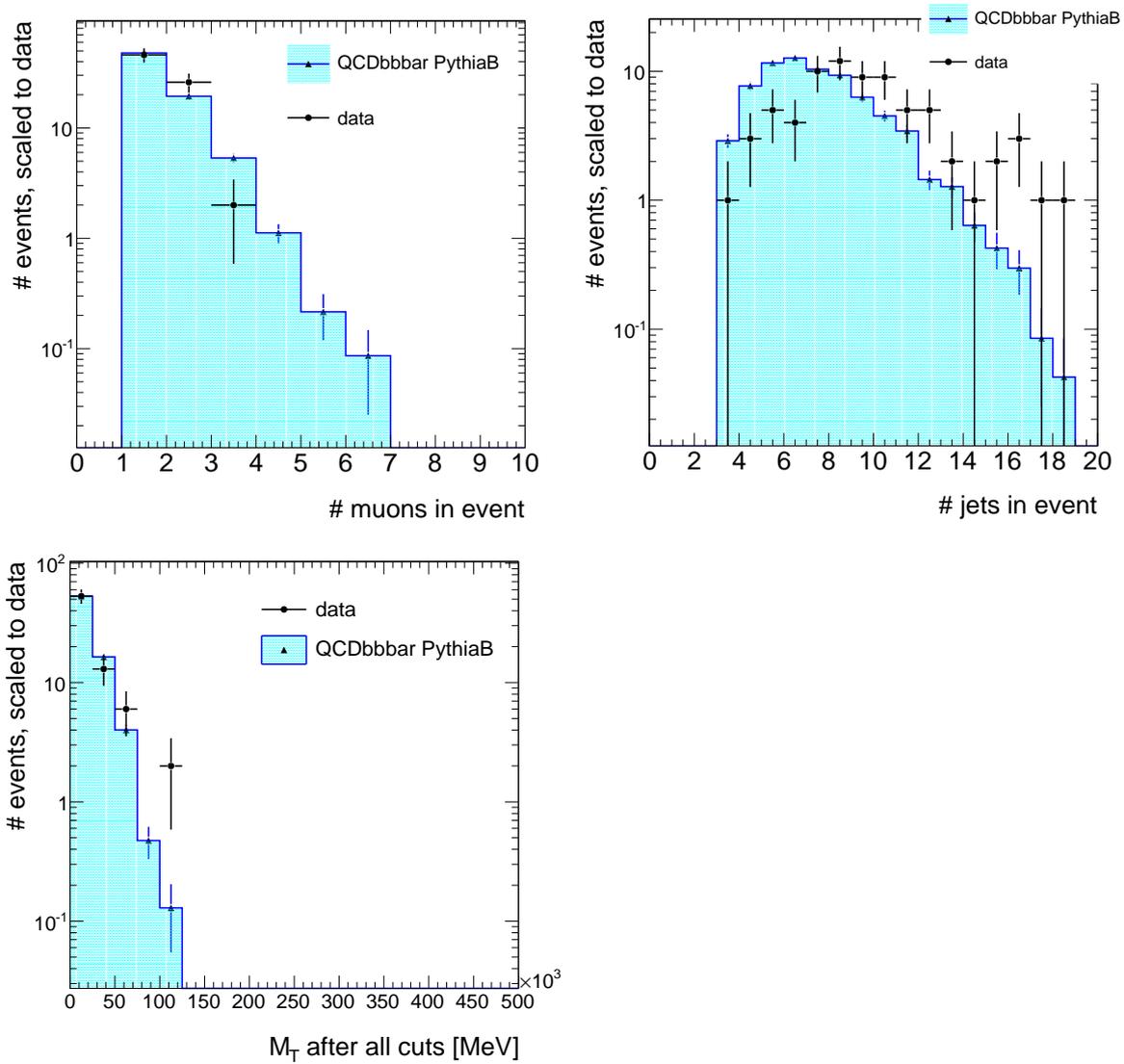


Figure 8.5: Comparison between data and Monte Carlo. Monte Carlo is scaled to data. Top left: Number of muons in the event, Top right: Number of jets in the event, bottom: M_T

	QCD bbbar	QCD udsc	ttbar	W+jets	Z+jets	Wbb	total MC ($2.95 pb^{-1}$)	data(A-F, $2.95 pb^{-1}$)
A	20.7 ± 1.0	3.9 ± 1.1	2.43 ± 0.028	4.78 ± 0.17	1.251 ± 0.088	0.079 ± 0.010	33.1 ± 1.5	30 ± 5.48
B	1.72 ± 0.30	0.97 ± 0.47	20.057 ± 0.078	27.43 ± 0.40	1.325 ± 0.088	0.401 ± 0.024	51.90 ± 0.70	44 ± 6.63
C	75.2 ± 1.9	6.4 ± 1.5	0.625 ± 0.023	0	0	$4.0 \cdot 10^{-3} \pm 2.3 \cdot 10^{-3}$	82.2 ± 2.4	58 ± 7.62
D	5.06 ± 0.50	0.44 ± 0.33	0.1482 ± 0.0091	0	0	0	5.65 ± 0.60	4 ± 2
A/B	12.0 ± 2.1	4.1 ± 2.3	0.1212 ± 0.0015	0.1742 ± 0.0067	0.944 ± 0.090	0.198 ± 0.029	-	0.68 ± 0.16
C/D	14.9 ± 1.5	15 ± 12	4.22 ± 0.30	-	-	-	-	14.5 ± 7.5

Table 8.6: Number of events in all regions as defined in the previous chapter and for all different backgrounds ($2.95 pb^{-1}$) for method 1. The last column contains results of data of period A-F. The column next to the last column includes all different backgrounds.

79

	QCD bbbar	QCD udsc	ttbar	W+jets	Z+jets	Wbb	total MC ($2.95 pb^{-1}$)	data(A-F, $2.95 pb^{-1}$)
A	7.71 ± 0.62	0.72 ± 0.49	1.870 ± 0.024	0.382 ± 0.047	0.106 ± 0.025	0.0549 ± 0.0087	10.84 ± 0.79	6 ± 2.45
B	0.93 ± 0.21	0.90 ± 0.47	15.279 ± 0.068	2.00 ± 0.11	0.0306 ± 0.0077	0.284 ± 0.020	19.42 ± 0.53	14 ± 3.74
C	75.2 ± 1.9	6.4 ± 1.5	0.625 ± 0.023	0	0	$4.0 \cdot 10^{-3} \pm 2.3 \cdot 10^{-3}$	82.2 ± 2.4	58 ± 7.62
D	5.06 ± 0.50	0.44 ± 0.33	0.1482 ± 0.0091	0	0	0	5.65 ± 0.60	4 ± 2
A/B	8.3 ± 2.0	0.79 ± 0.68	0.1224 ± 0.0017	0.191 ± 0.026	1.17 ± 0.40	0.193 ± 0.034	-	0.43 ± 0.21
C/D	14.9 ± 1.5	15 ± 12	4.22 ± 0.30	-	-	-	-	14.5 ± 7.5

Table 8.7: Number of events in all regions as defined in the previous chapter and for all different backgrounds ($2.95 pb^{-1}$). The structure is the same as in table 8.6, but in region A and B an additional b-tagging cut at SV0 weight 5 was applied.

8.5 Outlook

Both methods of the last chapter in order to estimate the QCD $b\bar{b}$ background have been applied to data in this chapter. An good agreement between data and expectations from Monte Carlo is found. The QCD $b\bar{b}$ background in the SUSY signal region is estimated to be low with 0.8 ± 0.6 events in the SUSY signal region (1-muon + 3 jets channel with $p_T^{first\ jet} > 60$ and $p_T^{third\ jet} > 30$ and for an integrated luminosity of $2.95\ pb^{-1}$). While both methods work well, two points still have to be mentioned:

- The results of both methods - in particular in this chapter - have to be understood as an upper limit for the QCD $b\bar{b}$ background in the SUSY signal region, because not all SUSY selection cuts were used in the estimation. The application of all SUSY selection cuts would not be possible, as this would reduce the statistics in the QCD PYTHIAB $b\bar{b}$ Monte Carlo samples too much, so that both methods could not be verified in Monte Carlo.
- Both methods need low energetic regions (in the sense of SUSY studies) in MET and in p_T (non-isolated muon). However, we have already seen that the turn-on curve of unrescaled triggers was in the region of 15 GeV in 2010. Only triggers with turn-on curves at even higher p_T values will stay unrescaled in the next year because of rate limitations. All control regions used in both methods must be in the saturation region of the trigger used. Most likely, this will be the case only for rescaled triggers. The results of both methods will need to be corrected for the rescaling of the trigger in order to get a correct estimation of the number of QCD $b\bar{b}$ events in the SUSY signal region.

Chapter 9

An alternative method in order to estimate the QCD background in the 1-muon channel

In the last chapters a method to estimate the QCD $b\bar{b}$ background in the signal region was presented. In this chapter more methods, relying on a cut on $|d_0^{sign}| = |\frac{d_0}{\sigma(d_0)}|$, are studied. d_0 is the transverse impact parameter with respect to the primary vertex of the selected muon. The transverse impact parameter is the distance in the x-y plane of the point of the closest approach to the primary vertex. This cut was studied in the SUSY group recently, as this cut is found to be efficient in rejecting the QCD background with respect to the SUSY signal. In this section the selection cuts in the 1-muon channel are the following [55]:

1. Good run list and trigger
2. Reject events with bad jet(s)
3. One isolated muon with $p_T > 20 \text{ GeV}$, no further isolated leptons with $p_T > 10 \text{ GeV}$
4. At least 3 jets, $p_T^{first \text{ jet}} > 60 \text{ GeV}$, $p_T^{third \text{ jet}} > 30 \text{ GeV}$
5. $|d_0^{sign}| < 5$
6. $MET > \max(60 \text{ GeV}, 0.25 M_{eff})$
7. $M_T > 100 \text{ GeV}$
8. $M_{eff} > 300, 500 \text{ GeV}$

All results for Monte Carlo samples in this chapter are for 1 pb^{-1} .

The efficiency of the $|d_0^{sign}| < 5$ cut (number of events after this cut divided by the number of events before this cut) was tested for the SU4 point as well as for the various backgrounds: $t\bar{t}$, W+jets, Z+jets, Wbb and the background from various QCD generators: QCD Alpgen udsc, QCD Alpgen $b\bar{b}$, QCD Pythia dijet, QCD PythiaB $b\bar{b}$. In table 9.1 the efficiencies are listed for the various physical processes, for events with isolated muons¹ as used in the selection cuts above and on the bottom for events with one selected muon without any isolation criteria² applied. The efficiency is near 100 % for all different physical processes in events with one isolated muon. However, the efficiency for QCD processes is near 50 % in contrast to SU4 or $t\bar{t}$ in events, where no isolation criteria was asked to be satisfied for muons.

¹isolation is defined here as $ptcone20 < 1.8 \text{ GeV}$ and $\Delta R(\text{muon} - \text{jet}) > 0.4$

²so **no** $ptcone20$ and **no** ΔR criteria applied

efficiency of $ d_0^{sign} $ -cut, isolation applied			
SU4	ttbar	QCD Pythia dijet	QCD Alpgen
98.5 %	98.7 %	94.5 %	100 %
efficiency of $ d_0^{sign} $ -cut, no isolation applied			
SU4	ttbar	QCD Pythia dijet	QCD Alpgen
86.7 %	78.0 %	55.4 %	56.4 %

Table 9.1: Efficiencies for passing a $|d_0^{sign}| < 5$ cut for SU4, ttbar, QCD Pythia dijet and QCD Alpgen. Events with one selected isolated muon on the top, events with no isolation criteria applied for the selected muon on the bottom. Whereas the efficiency is near 100 % for all physical processes if isolation criteria are applied, the efficiency is only near 55 % for QCD if no isolation criteria were applied. This information can be used in the construction of an enriched QCD control sample.

The $|d_0^{sign}|$ distribution (as it can be seen after applying the jet cuts) is shown in figure 9.1 for SU4, $t\bar{t}$ and the various QCD generators, on the left for events with one isolated muon, on the right for events with one muon (can be isolated or not). Whereas $t\bar{t}$ and SU4 peak very sharply at low values and decrease faster, the QCD Pythia dijet samples have more events with high $|d_0^{sign}|$ values. It can be seen, too, that the statistics of QCD Alpgen samples is very limited at high $|d_0^{sign}|$ values. Therefore, the Alpgen samples are not used further for this study.

A very simple method to estimate the QCD background in the signal region is just to use an ABCD-method with the MET versus $|d_0^{sign}|$ ³. However, in comparison to data, it was seen that the contamination of other backgrounds than QCD is rather big in one control region (with low MET and low $|d_0^{sign}|$). So, this method does not seem promising. Therefore, it is not developed and not shown in the following ⁴. In contrast, only a method which is used successfully by the top-group in their note [56] is studied in more detail and presented in the following.

9.1 QCD estimation with the matrix method

The top-group has developed and tested successfully an advanced matrix method to estimate the QCD background in their top signal region [56]. The order of magnitude of the QCD background is estimated by solving a system of linear equations, where all variables needed are measurable or known. By modifying this method slightly it is possible to use this method for the estimation of the QCD background in the SUSY 1 muon + 3 jets channel. This method is described in detail below. In the following two sections the method is first tested with MC samples and thereafter applied to data.

Two different samples are defined for applying this method:

- **Tight:** Applying the usual signal selection (as defined above, only up to cut 5).

³Define for example regions like this:

- signal region: $MET > 60 \text{ GeV}$ and $|d_0^{sign}| < 5$
- control region 1: $20 < MET < 60 \text{ GeV}$ and $|d_0^{sign}| < 5$
- control region 2: $20 < MET < 60 \text{ GeV}$ and $|d_0^{sign}| > 5$
- control region 3: $MET > 60 \text{ GeV}$ and $|d_0^{sign}| > 5$

Contamination in control region 1 by W+jets.

⁴However, it might be still possible to use such a method if correcting the contributions of W+jets and Z+jets in the contaminated control region by using the measured cross section of W+jets and Z+jets.

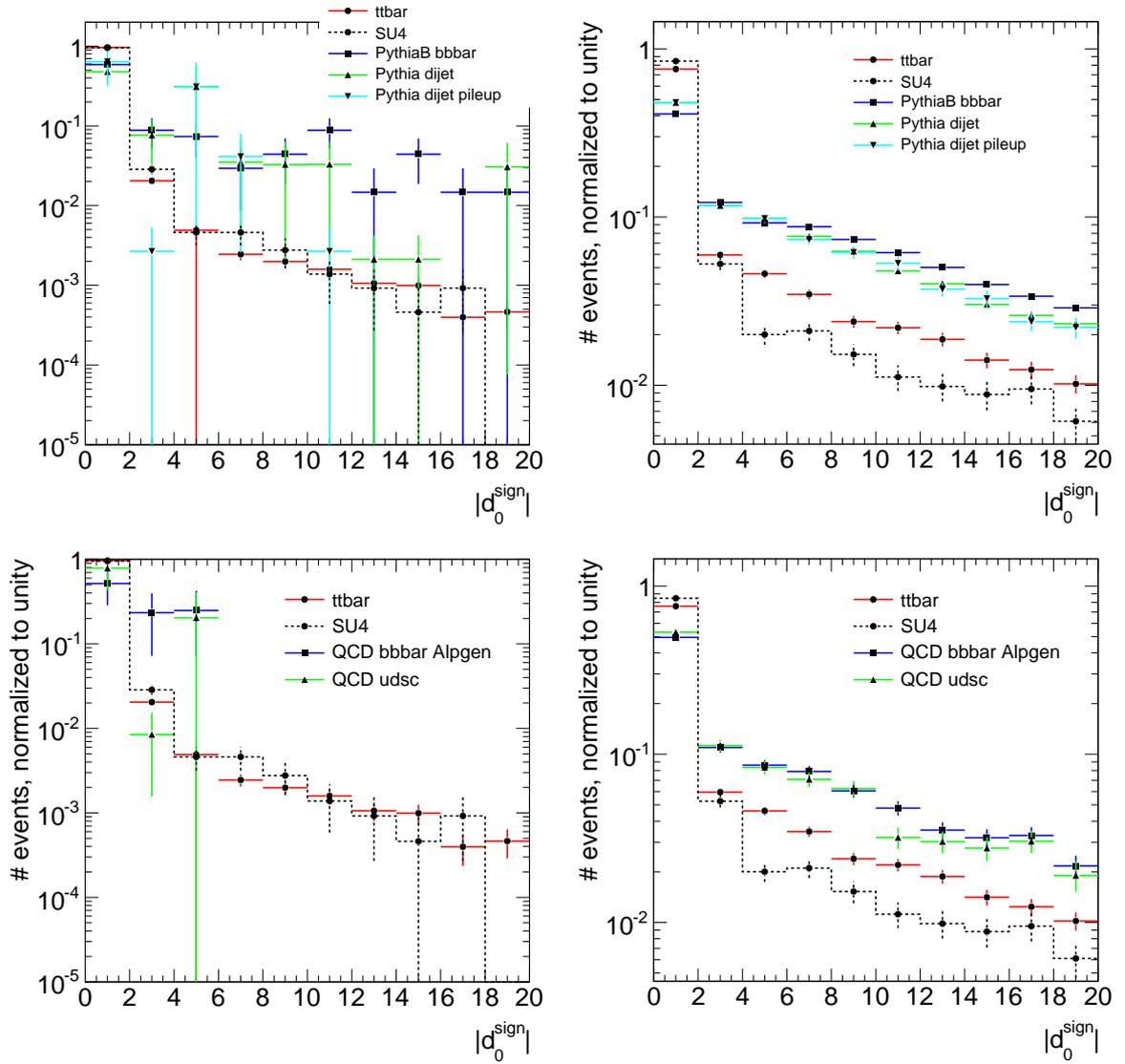


Figure 9.1: $|d_0^{sign}|$ distribution for $t\bar{t}$, SU4 and the various QCD MC samples. On the left for events with one isolated muon and on the right for events with one muon without any isolation criteria applied. The distributions of SU4 and $t\bar{t}$ peak more sharply near 0 than the QCD distributions.

- Loose: The signal selection cuts (up to cut 5), but not applying any isolation criteria for the selected muon (so **no** $ptcone20$ cut, **no** $|d_0^{sign}|$ -cut and **no** ΔR cut).

The number of events passing the loose selection cuts (N^{loose}) can be expressed as a sum of signal and background (without QCD) events (SUSY + BG events, $N_{SUSY+BG}^{loose}$) and of QCD events (N_{QCD}^{loose}):

$$N^{loose} = N_{SUSY+BG}^{loose} + N_{QCD}^{loose} \quad (9.1)$$

Similarly, the number of events passing the tight selection cuts (N^{tight}) can be expressed as the sum of the signal events and background events passing the tight selection cuts. N^{tight} can be expressed as a function of $N_{SUSY+BG}^{loose}$ and N_{QCD}^{loose} , if one introduces two efficiencies:

$$\begin{aligned} \epsilon_{SUSY+BG} &= \frac{N_{SUSY+BG}^{tight}}{N_{SUSY+BG}^{loose}} \\ &= \frac{\text{Number of signal and BG (no QCD) events passing tight selection cuts}}{\text{Number of signal and BG (no QCD) events passing loose selection cuts}} \end{aligned} \quad (9.2)$$

$$\epsilon_{QCD} = \frac{N_{QCD}^{tight}}{N_{QCD}^{loose}} = \frac{\text{Number of QCD events passing tight selection cuts}}{\text{Number of QCD events events passing loose selection cuts}} \quad (9.3)$$

$$N^{tight} = \epsilon_{SUSY+BG} N_{SUSY+BG}^{loose} + \epsilon_{QCD} N_{QCD}^{loose} \quad (9.4)$$

Equations 9.1 and 9.4 can be solved to get an expression of N_{QCD}^{tight} in measurable quantities:

$$N_{QCD}^{tight} = \frac{\epsilon_{QCD}}{\epsilon_{SUSY+BG} - \epsilon_{QCD}} (N^{loose} \epsilon_{SUSY+BG} - N^{tight}) \quad (9.5)$$

ϵ_{QCD} and $\epsilon_{SUSY+BG}$ will be determined in control samples. Therefore, it has to be assumed here that the efficiencies do not dependent on the event topology, so that the efficiency is the same in the control sample as with the signal selection cuts. Equation 9.5 will not work if the number of events passing loose selection cuts is too similar to the number of events passing tight selection cuts. Therefore, it is not possible to choose the loose selection criteria too tight. In the two following sections the different quantities will be obtained step by step. $\epsilon_{SUSY+BG}$ is taken from MC for the time being, but as will be shown later, its value can vary without affecting the results.

9.2 Check matrix method with MC samples

9.2.1 Determination of N^{tight} and N^{loose}

N^{tight} and N^{loose} are determined in two different regions with the event selection cuts as defined above and the modifications for the loose sample:

- W control region (W CR)⁵: $30 < MET < 60 \text{ GeV}$ and $40 < M_T < 80 \text{ GeV}$
- Signal region (SR): $MET > 60 \text{ GeV}$ and $M_T > 100 \text{ GeV}$

The numbers for SU4 and the various backgrounds, namely $t\bar{t}$, W+jets, Z+jets, Wbb and QCD Pythia dijet can be found in table 9.2 and in table 9.3, respectively. The MET versus M_T distributions are shown in figure 9.2 for the tight selection cuts and in figure 9.3 for the loose selection cuts.

⁵This region is expected to be dominated by $t\bar{t}$ and by W+jets events. Therefore, it can be used for cross checks with other physics groups.

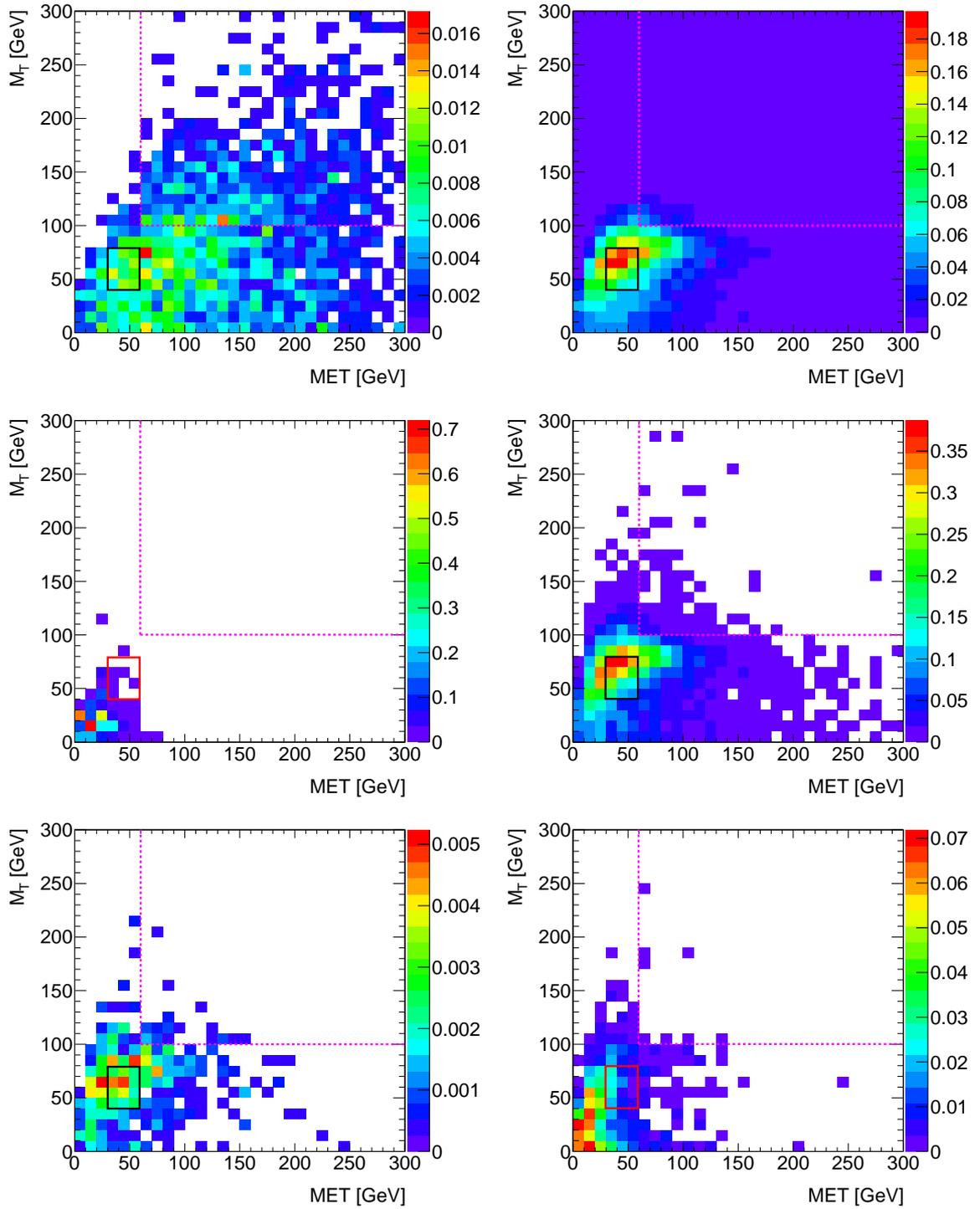


Figure 9.2: N_{tight} for the various physical processes. From top to bottom and from left to right: SU4, $t\bar{t}$, QCD Pythia dijet, W +jets, W bb, Z +jets. The W control region is indicated by the black or red box. The signal region at high MET and M_T values is indicated by magenta dashed lines.

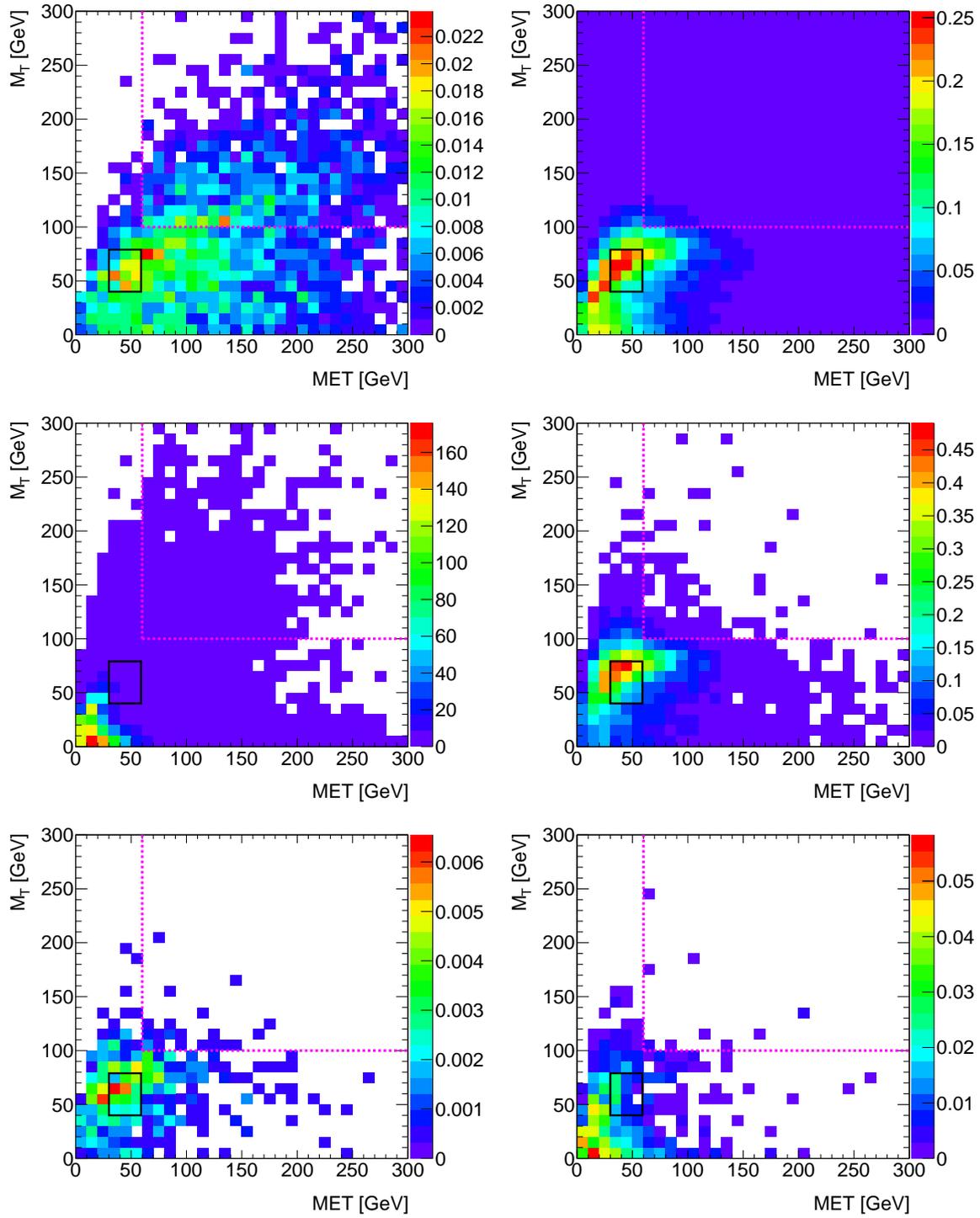


Figure 9.3: N_{loose} for the various physical processes. From top to bottom and from left to right: SU4, $t\bar{t}$, QCD Pythia dijet, W+jets, Wbb, Z+jets. The W control region is again indicated by a black box and the signal region with magenta dashed lines.

	SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	total
W CR	0.108 ± 0.010	0.029 ± 0.010	1.632 ± 0.029	2.811 ± 0.076	0.0354 ± 0.0041	0.151 ± 0.017	4.766 ± 0.084
SR	0.720 ± 0.026	0 ± 0	0.473 ± 0.017	0.269 ± 0.023	0.0102 ± 0.0022	0.016 ± 0.005	1.488 ± 0.039

Table 9.2: N^{tight} for various physical processes

	SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	
W CR	0.171 ± 0.013	47.4 ± 2.0	2.199 ± 0.034	3.489 ± 0.085	0.0414 ± 0.0044	0.145 ± 0.017	53.4 ± 2.0
SR	1.135 ± 0.033	1.25 ± 0.14	0.600 ± 0.020	0.353 ± 0.026	0.0120 ± 0.0024	0.0251 ± 0.0079	3.38 ± 0.15

Table 9.3: N^{loose} for various physical processes

9.2.2 Determination of ϵ_{QCD}

ϵ_{QCD} as defined above will be obtained in two QCD-enriched control samples. For this N_{tight} and N_{loose} as defined above will be used.

- Control sample 1 (low MET):
 - Events with at least one jet with $p_T > 60 \text{ GeV}$
 - $MET < 10 \text{ GeV}$
- Control sample 2 (high $|d_0^{sign}|$):
 - Events with at least one jet with $p_T > 60 \text{ GeV}$
 - $|d_0^{sign}| > 5$

Control sample 1:

The numbers in table 9.4 and in table 9.5 are found for SU4 and the various backgrounds with the definitions above for loose and tight selection cuts .

The contamination of W+jets for the tight event selection cuts is large with respect to QCD Pythia, but it is negligible for loose event selection criteria. In total, including signal and all backgrounds, $N^{tight} = 19.9 \pm 2.0$ and $N^{loose} = 1665 \pm 18$. According to the definition of ϵ_{QCD} and taking into account all BG and signal events, this results in:

$$\epsilon_{QCD} = \frac{N^{tight}}{N^{loose}} = 12.0 \cdot 10^{-3} \pm 1.2 \cdot 10^{-3} \quad (9.6)$$

and has to be compared to

$$\epsilon_{QCD} = \frac{N^{tight}(only \text{ QCD})}{N^{loose}(only \text{ QCD})} = 6.8 \cdot 10^{-3} \pm 1.2 \cdot 10^{-3} \quad (9.7)$$

for only taking into account QCD events.

Control sample 2:

The second control sample, as defined above, profits from inverting the $|d_0^{sign}|$ -cut. Like this, a strongly dominated QCD control sample is obtained, where also W+jets and Z+jets are well suppressed. However, the full isolation criteria as defined in the tight selection cuts are not applied, because this is contradictory to the applied $|d_0^{sign}|$ -cut. The numbers in table

SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	total
0.0162 ± 0.0039	11.3 ± 2.0	0.239 ± 0.011	5.37 ± 0.13	0.0196 ± 0.0031	2.996 ± 0.079	19.9 ± 2.0

Table 9.4: N_{tight} in SU4 and the various backgrounds for control sample 1.

SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	total
0.0210 ± 0.0045	1656 ± 18	0.5507 ± 0.017	6.27 ± 0.15	0.0235 ± 0.0034	1.935 ± 0.064	1665 ± 18

Table 9.5: N_{loose} in SU4 and the various backgrounds for control sample 1.

SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	total
0.0334 ± 0.0056	11.2 ± 2.0	0.1330 ± 0.0083	1.380 ± 0.068	0.0059 ± 0.0017	0.313 ± 0.025	13.1 ± 2.0

Table 9.6: N^{tight} for the second control sample.

9.6 and in table 9.7 are obtained. With $N^{tight} = 13.1 \pm 2.0$ and $N^{loose} = 2550 \pm 21$ (including all signal and backgrounds events) ϵ_{QCD} is calculated.

$$\epsilon_{QCD} = \frac{N^{tight}}{N^{loose}} = 5.14 \cdot 10^{-3} \pm 0.79 \cdot 10^{-3} \quad (9.8)$$

and has to be compared to

$$\epsilon_{QCD} = \frac{N^{tight}(only\ QCD)}{N^{loose}(only\ QCD)} = 4.40 \cdot 10^{-3} \pm 0.71 \cdot 10^{-3} \quad (9.9)$$

In the second control sample the agreement between ϵ_{QCD} which is estimated with all events from all backgrounds and signal and which is only estimated from QCD is better than in control sample 1. This is explained by the large contamination by W+jets events in control sample 1. The distributions of MET versus M_T for tight and loose criteria are shown in figure 9.4 and 9.5, respectively. These distributions show that it is not possible to introduce additional cuts on high MET or M_T in the control sample without losing too much statistics in the QCD events. ⁶

9.2.3 Estimation of the number of QCD events in the W control region and in the signal region

The following values were obtained in the last subsections:

- In W CR:

- $N^{tight} = 4.766 \pm 0.084$
- $N^{loose} = 53.4 \pm 2.0$

- In SR:

- $N^{tight} = 1.488 \pm 0.039$
- $N^{loose} = 3.38 \pm 0.15$

- As ϵ_{QCD} it is taken the value from the second control sample as the W+jets contamination is lower: $\epsilon_{QCD} = 5.14 \cdot 10^{-3} \pm 0.79 \cdot 10^{-3}$

⁶Cuts on MET or M_T in the control sample might be useful if ϵ_{QCD} would depend on MET or M_T in order to be closer to the definitions of the signal regions.

SU4	QCD Pythia dijet	$t\bar{t}$	W+jets	Wbb	Z+jets	total
0.437 ± 0.020	2546 ± 21	1.594 ± 0.029	1.675 ± 0.074	0.0265 ± 0.0036	0.385 ± 0.030	2550 ± 21

Table 9.7: N^{loose} for the second control sample.

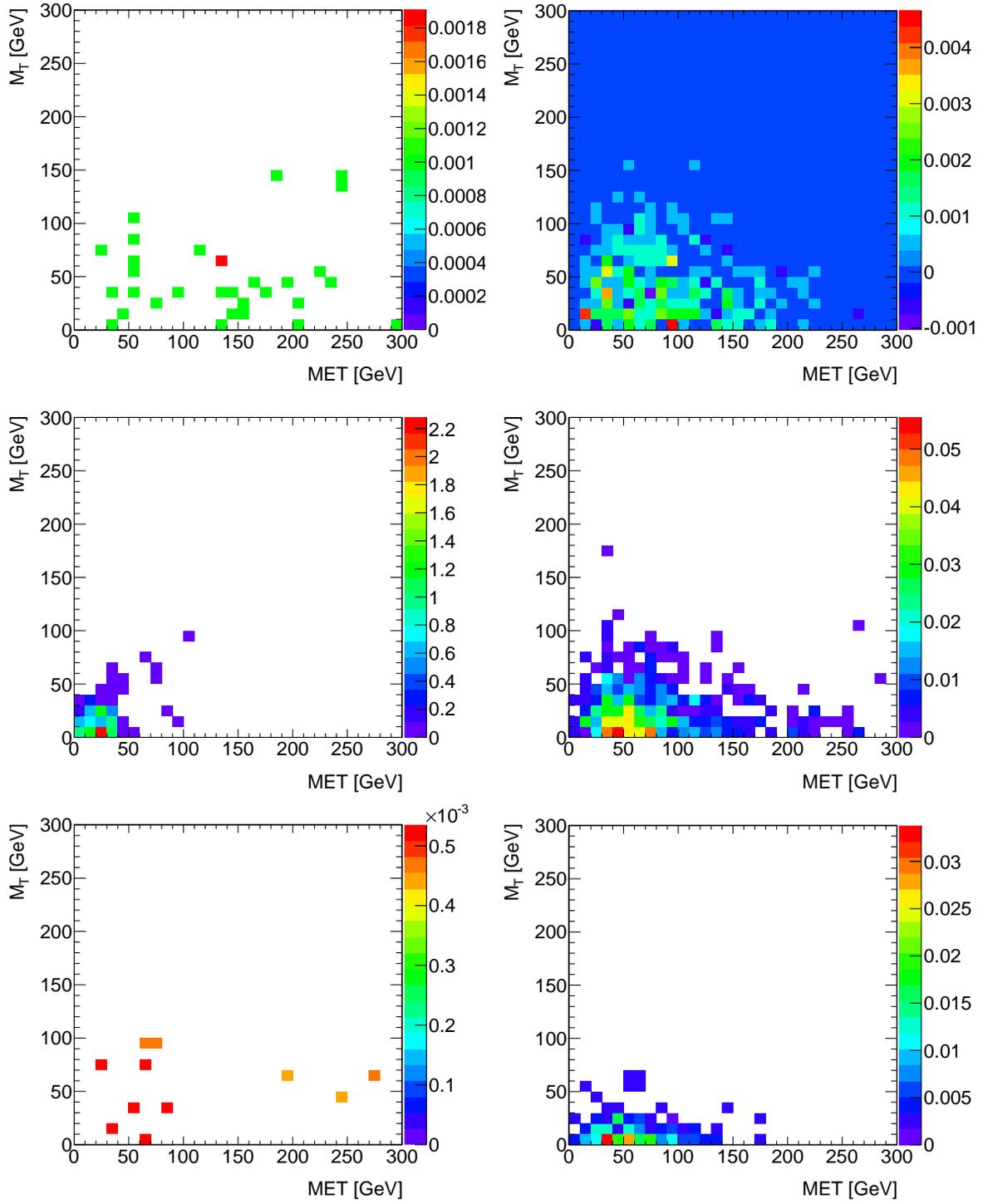


Figure 9.4: M_T versus MET for tight selection cuts. From top to bottom and from left to right: SU4, $t\bar{t}$, QCD Pythia dijet, W+jets, Wbb, Z+jets

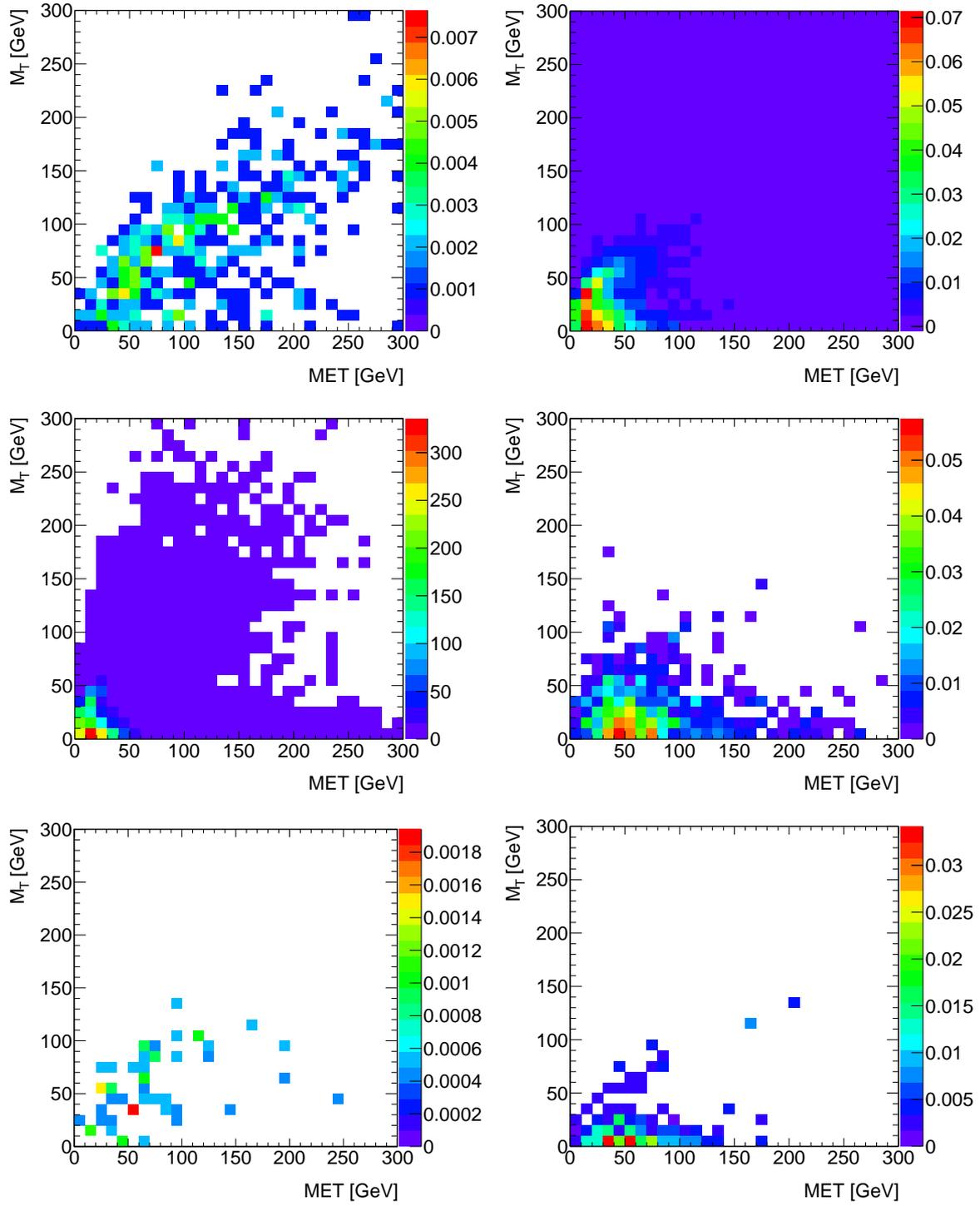


Figure 9.5: M_T versus MET for loose selection cuts. From top to bottom and from left to right: SU4, $t\bar{t}$, QCD Pythia dijet, W+jets, Wbb, Z+jets

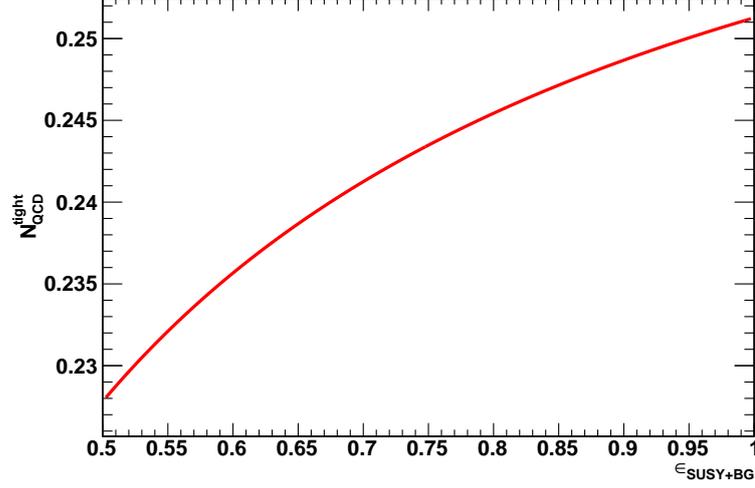


Figure 9.6: N_{QCD}^{tight} is shown as function of $\epsilon_{SUSY+BG}$ in the W control region.

- $\epsilon_{SUSY+BG} = 0.63 \pm 0.03$ (this is calculated by comparing the number of SU4 events passing the tight and the loose selection cuts. It will be shown below, that this value can be varied between 0.5 and 1.0 without changing the results strongly.)

With these values equation 9.5 gives for the W CR:

$$\begin{aligned}
 N_{QCD}^{tight} &= \frac{\epsilon_{QCD}}{\epsilon_{SUSY+BG} - \epsilon_{QCD}} (N^{loose} \epsilon_{SUSY+BG} - N^{tight}) \\
 &= \frac{5.14 \cdot 10^{-3}}{0.63 - 5.14 \cdot 10^{-3}} (53.4 \cdot 0.63 - 4.766) \\
 &= 0.238 \pm 0.038
 \end{aligned} \tag{9.10}$$

In the signal region one obtains:

$$\begin{aligned}
 N_{QCD}^{tight} &= \frac{\epsilon_{QCD}}{\epsilon_{SUSY+BG} - \epsilon_{QCD}} (N^{loose} \epsilon_{SUSY+BG} - N^{tight}) \\
 &= \frac{5.14 \cdot 10^{-3}}{0.63 - 5.14 \cdot 10^{-3}} (3.38 \cdot 0.63 - 1.488) \\
 &= 5.3 \cdot 10^{-3} \pm 1.3 \cdot 10^{-3}
 \end{aligned} \tag{9.11}$$

The matrix method predicts a negligible QCD background in the signal region. This in agreement with the zero QCD events in the signal region in table 9.2. However, the number of QCD events in the W control region is overestimated by a factor of nearly 10. The estimation in the signal region and in the W control region does not vary strongly with $\epsilon_{SUSY+BG}$ as shown in figure 9.7 for the signal region and in figure 9.6 for the W control region. As some cuts for the selection of the signal region were not applied (for example M_{eff})⁷ and the definition of the signal region is rather loose compared to common definitions of the SUSY working group, the result in the signal region should be understood as upper limit of the QCD background in the signal region.

⁷Applying the whole SUSY signal selection cuts would reduce the available statistics in the QCD PYTHIA dijet samples dramatically and complicate the verification of the matrix method.

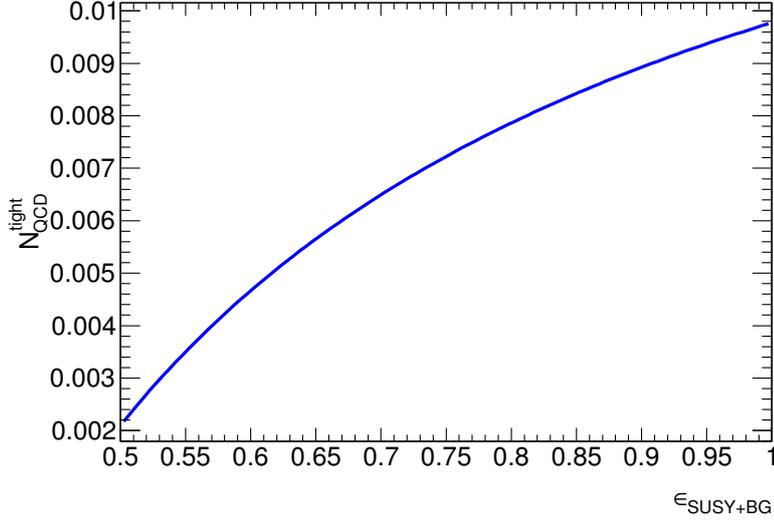


Figure 9.7: N_{QCD}^{tight} is shown as function of $\epsilon_{SUSY+BG}$ in the signal region.

9.3 Applying the matrix method to data

The method presented in detail in the last section is applied to data. So far, only data of period G has been considered (5.655 pb^{-1}). In the same order as before all quantities are obtained from data.

N^{tight} and N^{loose} :

MET versus M_T distributions can be seen in figure 9.8 on the left for the tight selection cuts as defined above and on the right for the loose selection cuts. Of these plots the number of events in the W control region (W CR) and in the signal region (SR) is obtained:

- W CR:

- $N^{tight} = 34 \pm 6$
- $N^{loose} = 300 \pm 17$

- In SR:

- $N^{tight} = 4 \pm 2$
- $N^{loose} = 18 \pm 4$

ϵ_{QCD} obtained from control sample 1 - low MET :

In the control sample defined in the last section it is found with the tight selection cuts $N^{tight} = 89 \pm 9$ and with the loose selection cuts $N^{loose} = 6298 \pm 79$. Thereof, ϵ_{QCD} is calculated:

$$\epsilon_{QCD} = \frac{N^{tight}}{N^{loose}} = 0.014 \pm 0.002 \quad (9.12)$$

This is in good agreement with the expectation from Monte Carlo:

$$\epsilon_{QCD} \stackrel{MC}{=} 0.012 \pm 0.001 \quad (9.13)$$

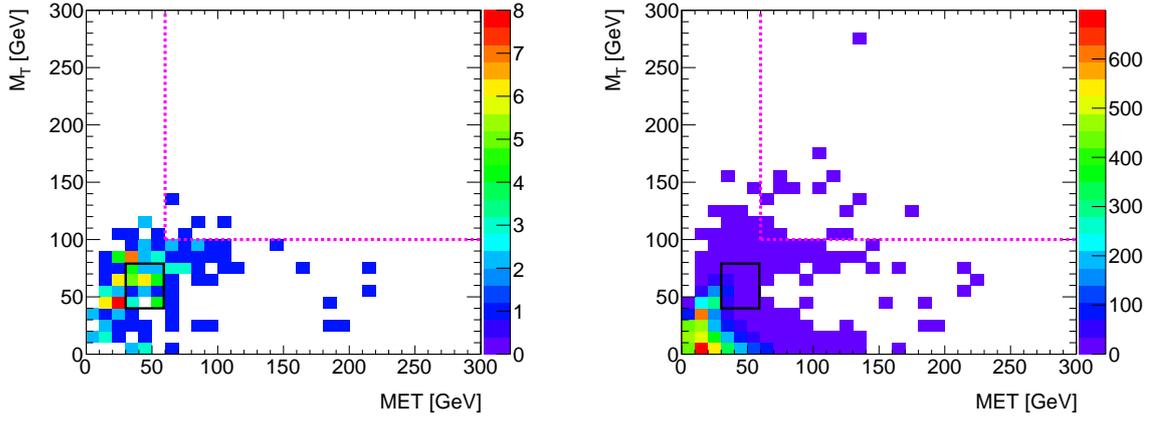


Figure 9.8: MET versus M_T for tight selection cuts on the left and for loose selection cuts on the right.

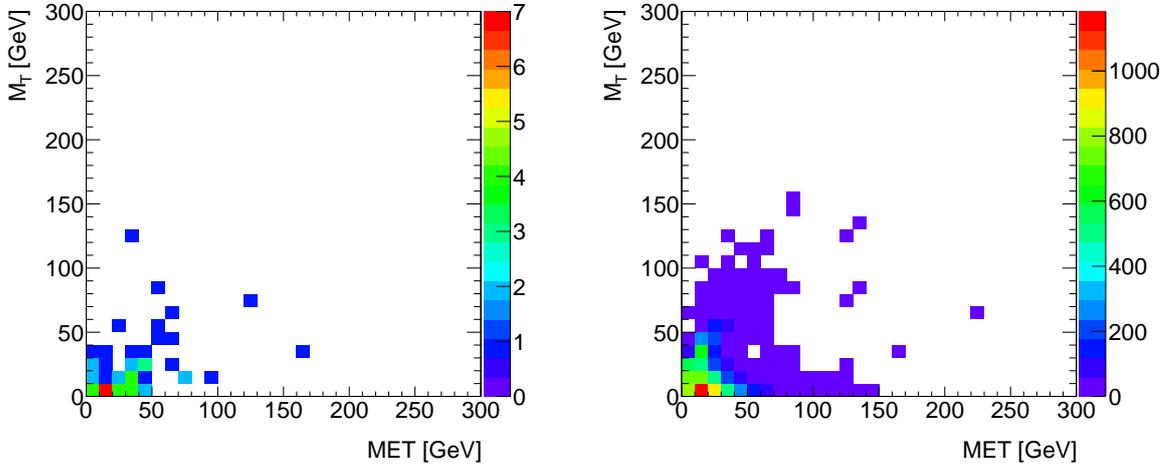


Figure 9.9: MET versus M_T in the second control sample. Tight selection criteria were applied on the left and loose selection criteria were applied on the right.

ϵ_{QCD} obtained from control sample 2 - high $|d_0^{sign}|$:

In the second control sample defined above $N^{tight} = 56 \pm 8$ events pass the tight selection cuts and $N^{loose} = 9650 \pm 98$ events pass the loose selection cuts. (MET versus M_T distributions can be found in figure 9.9.)

$$\epsilon_{QCD} = \frac{N^{tight}}{N^{loose}} = 5.8 \cdot 10^{-3} \pm 0.1 \cdot 10^{-3} \quad (9.14)$$

This result is compared to the expectation from Monte Carlo.

$$\epsilon_{QCD} \stackrel{MC}{=} 5.1 \cdot 10^{-3} \pm 0.8 \cdot 10^{-3} \quad (9.15)$$

Results from data are in agreement with Monte Carlo expectations for both control samples.

The number of QCD events in the W CR is estimated by 9.5 to:

$$\begin{aligned}
N_{QCD}^{tight} &= \frac{\epsilon_{QCD}}{\epsilon_{SUSY+BG} - \epsilon_{QCD}} (N^{loose} \epsilon_{SUSY+BG} - N^{tight}) \\
&= \frac{5.8 \cdot 10^{-3}}{0.63 - 5.8 \cdot 10^{-3}} (300 \cdot 0.63 - 34) \\
&= 1.4 \pm 0.1
\end{aligned} \tag{9.16}$$

This is in a very good agreement to expectations from Monte Carlo which predict 1.3 ± 0.2 events.

And in the SR:

$$\begin{aligned}
N_{QCD}^{tight} &= \frac{\epsilon_{QCD}}{\epsilon_{SUSY+BG} - \epsilon_{QCD}} (N^{loose} \epsilon_{SUSY+BG} - N^{tight}) \\
&= \frac{5.8 \cdot 10^{-3}}{0.63 - 5.8 \cdot 10^{-3}} (18 \cdot 0.63 - 4) \\
&= 6.8 \cdot 10^{-2} \pm 3.1 \cdot 10^{-2}
\end{aligned} \tag{9.17}$$

Whereas Monte Carlo predicts 0.030 ± 0.007 events. The results indicate that the QCD background is negligible in the signal region.

Chapter 10

Summary

The QCD background in SUSY searches in the 1-lepton (and in particular 1-muon) channel is expected to be very small, because QCD events should be suppressed by the requirement of one isolated lepton in the SUSY selection cuts. This small order of magnitude of QCD events in the SUSY signal region was also indicated by studies at a center-of-mass energy of 10 TeV in the past. As the LHC is running at a center-of-mass energy of 7 TeV, the QCD background has to be estimated again with the modified conditions. Data-driven methods are preferred in order to avoid dependencies on Monte Carlo simulations.

In particular, QCD $b\bar{b}$ events can survive the SUSY selection cuts, because these events can have a high Missing Transverse Energy due to the neutrino escaping the detector from the decay of the b-quark (B-hadron). Therefore, the work focused on the QCD $b\bar{b}$ background in its first part.

An important step in the estimation of the QCD $b\bar{b}$ background is the construction of a control sample, consisting of nearly purely QCD $b\bar{b}$ events. The selection cuts for the control sample are close to the usual SUSY selection cuts, although no cut on MET, M_{eff} or M_T is used. In contrast to the usual SUSY selection cuts, however, only events with one non-isolated muon are used (this suppresses $t\bar{t}$ events which contain mainly isolated muons). In addition to the usual SUSY selection cuts, the SV0 b-tagging algorithm is applied two times (the SV0 algorithm is one of the b-tagging algorithms recommended for early data). Both b-tagging cuts combined select events which include most likely two b-jets and so they suppress QCD events with light flavors.

From this control sample it is possible to estimate the QCD $b\bar{b}$ background in the 1-muon channel. For this purpose the selection cuts of the control sample are improved by introducing an additional cut on the Missing Transverse Energy at the same cut value as for the muon p_T . This is necessary due to normalization issues. Further, the definition of non-isolated muons is modified by not using muons which are too close to a reconstructed jet. By comparing the p_T distribution of the selected muon in the control sample with the MET distribution in events with one selected isolated muon (which corresponds to the first part of the usual selection cuts for the SUSY signal region), it is possible to predict the shape of the tail of the MET distribution. Using this feature, the QCD $b\bar{b}$ background in a signal region defined by $MET > 40 \text{ GeV}$ is estimated to be 1.04 ± 0.12 events (selection cuts for jets: At least 3 jets with $p_T^{first \text{ jet}} > 40 \text{ GeV}$ and $p_T^{third \text{ jet}} > 30 \text{ GeV}$) for an integrated luminosity of 1 pb^{-1} .

However, this proves to be an overestimation due to contamination from other backgrounds, in particular $t\bar{t}$ and W+jets events, in one of the control regions. This contamination can be reduced successfully by introducing an additional b-tagging cut on top of the selection cuts of the signal region. Due to this cut, it is necessary to correct the result obtained by the b-tagging efficiency. Up to this point, the presented method was purely data-driven and not

depending on Monte Carlo. At this point however, the b-tagging efficiency for QCD $b\bar{b}$ has to be taken from Monte Carlo. It is expected, that after final calibration of the SV0 b-tagging algorithm by the ATLAS b-tagging group this efficiency will be provided by the b-tagging group and will come from data-driven methods as well.

By taking the efficiency from Monte Carlo for the time being this results in an estimation of 0.726 ± 0.093 QCD $b\bar{b}$ events in the SUSY signal region (again defined by $MET > 40 GeV$, all results are quoted for $1 pb^{-1}$ in this section). By comparing this result to the expected number of QCD $b\bar{b}$ events from Monte Carlo, this is found to be an excellent estimation.

The validity of the method can be tested in data by using the QCD $b\bar{b}$ control sample. A very good agreement between data and Monte Carlo is found if considering a scaling factor which reflects the simulation of the QCD PYTHIA Monte Carlo samples at leading order in the strong coupling constant. That a scaling factor has to be applied is known from various SUSY studies of summer 2010. Therefore, the QCD $b\bar{b}$ background can be estimated with the method.

However, the results have to be understood as an upper limit for the QCD $b\bar{b}$ background in the signal region. This is mainly due to the very limited statistics in the Monte Carlo samples, so that it is not possible to verify the method with all selection cuts as usually used for the selection of the SUSY signal region. In particular, a cut on the Missing Transverse Energy at $MET > 125 GeV$ is usually applied in the definition of the signal region. However, already with a cut of $MET > 60 GeV$ the method cannot be verified in Monte Carlo any longer, because the statistic in the used QCD Monte Carlo samples is not sufficient. But an upper limit on the QCD $b\bar{b}$ background is sufficient for SUSY 1-lepton searches because of the smallness of the QCD background. A more serious problem exists though. As outlined in detail, it is essential to use very low MET and p_T values for two of the three control regions. However, the triggers, which are allowing these low energy regimes, are going to be heavily prescaled with the increasing luminosity. This will be a serious problem in the next year, as already at the end of the proton-proton collisions in 2010, the turn-on curves of the triggers barely allowed the application of the method presented. However, the method might still be applicable by using prescaled triggers if correcting for this prescaling.

Nevertheless, the application of a more robust method is desirable, whose systematic uncertainties can be calculated more easily. The matrix method, as introduced by the top group, has the additional advantage of estimating the total QCD background in the signal region. It works by considering the total number of events in the signal region (or another control region) and by separating this total number of events in various categories depending on their origin. By the definition of looser selection cuts it is possible to obtain a quantitative expression of how many QCD events survive the signal selection cuts (the tight selection cuts). This efficiency, ϵ_{QCD} , can be obtained in control sample enriched in QCD events. The result has again to be understood as an upper limit on the expected QCD background, because it is again not possible to apply the full SUSY selection cuts. Finally, the efficiency of any signal events surviving the signal selection cuts with respect to the looser selection cuts needs to be obtained from data as well. Here, a method has not been developed yet, although simple calculations indicate that the final estimation of QCD events in the signal region does not depend strongly on this efficiency. So, an upper limit of predicted QCD events in the signal region (definition of the signal region here: $MET > 60 GeV$ and $M_T > 100 GeV$) can be given to be $5.3 \cdot 10^{-3} \pm 1.3 \cdot 10^{-3}$ events for $1 pb^{-1}$ (for Monte Carlo). As the definition of the signal region is significantly stricter than in the studies of the number of QCD $b\bar{b}$ events in a signal region, this result is consistent with earlier results. This method shows a good agreement between the results from data and from Monte Carlo, too.

All together, both methods predict a tiny QCD background in the signal region.

Appendix A

Monte Carlo samples

The Monte Carlo datasets used in these work are detailed here with their cross sections and number of events. These numbers are available by Ami [57]¹. The p_T (leading jet) slices are explained in table A.1.

A.1 QCD $b\bar{b}$

Number	Comments	σ (nb^{-1})	Generator efficiency ϵ	Number of events
107310	AlpGen J4	$6.9127 \cdot 10^{-2}$	1	1000
107311	AlpGen J4	$4.6517 \cdot 10^{-1}$	1	4500
107312	AlpGen J4	$5.6561 \cdot 10^{-1}$	1	6000
107313	AlpGen J4	$3.7662 \cdot 10^{-1}$	1	4000
107314	AlpGen J4	$2.2743 \cdot 10^{-1}$	1	2500
107315	AlpGen J5	$1.2921 \cdot 10^{-3}$	1	500
107316	AlpGen J5	$8.9419 \cdot 10^{-3}$	1	3000
107317	AlpGen J5	$1.5576 \cdot 10^{-2}$	1	5000
107318	AlpGen J5	$1.4651 \cdot 10^{-2}$	1	4500
107319	AlpGen J5	$1.4009 \cdot 10^{-2}$	1	4500
107340	AlpGen J2 mu-filtered	$7.6144 \cdot 10^1$	$6.7295 \cdot 10^{-2}$	51000
107341	AlpGen J2 mu-filtered	$1.0263 \cdot 10^2$	$3.940421 \cdot 10^{-2}$	40500
107342	AlpGen J2 mu-filtered	$2.5605 \cdot 10^1$	$4.645977 \cdot 10^{-2}$	11500
107343	AlpGen J2 mu-filtered	5.0575	$4.864286 \cdot 10^{-2}$	2500
107344	AlpGen J2 mu-filtered	$9.2641 \cdot 10^{-1}$	$4.552490 \cdot 10^{-2}$	500
107345	AlpGen J3 mu-filtered	2.7560	$1.494768 \cdot 10^{-1}$	4000
107346	AlpGen J3 mu-filtered	$1.3144 \cdot 10^1$	$8.895214 \cdot 10^{-2}$	12000
107347	AlpGen J3 mu-filtered	8.3547	$8.622176 \cdot 10^{-2}$	7000
107348	AlpGen J3 mu-filtered	3.0452	$8.014105 \cdot 10^{-2}$	2500
107349	AlpGen J3 mu-filtered	1.0428	$8.282259 \cdot 10^{-2}$	1000

Number	Comments	σ (nb^{-1})	Generator efficiency ϵ	Number of events
108405	PYTHIAB μ -filtered ($p_T > 15$ GeV)	73.9	1	4440000
108326	PYTHIAB e -filtered ($p_T > 15$ GeV)	77.4	0.9	4420000

¹The number of events cited here is the number of events generated. These do not need to agree with the number of events in the final Monte Carlo samples.

Name	$p_T(\text{min})$	$p_T(\text{max})$
J0	8	17
J1	17	35
J2	35	70
J3	70	140
J4	140	280
J5	280	560
J6	560	1120
J7	1120	2240
J8	2240	∞
J5+	280	∞

Table A.1: The Monte Carlo samples are divided into slices according to the p_T of the leading jet. This table is from [19].

A.2 QCD $udsc$

Number	Comments	$\sigma (nb^{-1})$	Generator efficiency ϵ	Number of events
108818	AlpGen J2 mu-filtered	17305	$1.4425 \cdot 10^{-3}$	280000
108819	AlpGen J2 mu-filtered	5213.1	$1.4813 \cdot 10^{-3}$	86500
108820	AlpGen J2 mu-filtered	702.75	$1.9507 \cdot 10^{-3}$	12500
108821	AlpGen J2 mu-filtered	93.361	$2.2025 \cdot 10^{-3}$	1500
108822	AlpGen J2 mu-filtered	12.189	$3.0136 \cdot 10^{-3}$	500
108823	AlpGen J3 mu-filtered	597.45	$5.1629 \cdot 10^{-3}$	30000
108824	AlpGen J3 mu-filtered	734.65	$4.5275 \cdot 10^{-3}$	37500
108825	AlpGen J3 mu-filtered	2.5490	$4.8504 \cdot 10^{-3}$	12500
108826	AlpGen J3 mu-filtered	61.863	$5.4530 \cdot 10^{-3}$	4000
108827	AlpGen J3 mu-filtered	14.379	$6.4851 \cdot 10^{-3}$	500
108362	AlpGen J4	15.582	1	120500
108363	AlpGen J4	29.297	1	241000
108364	AlpGen J4	20.111	1	172000
108365	AlpGen J4	8.6647	1	78000
108366	AlpGen J4	3.4622	1	31000
108367	AlpGen J5	0.32166	1	81000
108368	AlpGen J5	0.75013	1	192000
108369	AlpGen J5	0.71049	1	187500
108370	AlpGen J5	0.41984	1	115000
108371	AlpGen J5	0.24135	1	65500

A.3 QCD dijet

Number	Comments	$\sigma (nb^{-1})$	Generator efficiency ϵ	Number of events
109276	PYTHIA J0	$9.8605 \cdot 10^6$	$8.5572 \cdot 10^{-5}$	497500
109277	PYTHIA J1	$6.7804 \cdot 10^5$	$1.2111 \cdot 10^{-3}$	500000
109278	PYTHIA J2	$4.0968 \cdot 10^4$	$5.4310 \cdot 10^{-3}$	490000
109279	PYTHIA J3	$2.1929 \cdot 10^3$	$1.3051 \cdot 10^{-2}$	500000
109280	PYTHIA J4	$8.7665 \cdot 10^1$	$2.2357 \cdot 10^{-2}$	495000
109281	PYTHIA J5	2.3488	$3.0230 \cdot 10^{-2}$	500000
109281	PYTHIA J6	0.033616	$3.4024 \cdot 10^{-2}$	10000

A.4 $t\bar{t}$

The number of events includes here the negative weight of some events.

Number	Comments	$\sigma (nb^{-1})$	Generator efficiency ϵ	Number of events
105200	semi-leptonic	$1.4412E - 01$	$5.5620E - 01$	773167
105204	full hadronic	$1.4428E - 01$	$4.360353e - 01$	116047

A.5 W+jets

Number	Comments	$\sigma (nb^{-1})$	Generator efficiency ϵ	Number of events
107680	AlpGen+Jimmy W+e ν	6.8705	1	1382000
107681	AlpGen+Jimmy W+e ν	1.2930	1	258500
107682	AlpGen+Jimmy W+e ν	$3.7660 \cdot 10^{-1}$	1	189000
107683	AlpGen+Jimmy W+e ν	$1.0129 \cdot 10^{-1}$	1	50500
107684	AlpGen+Jimmy W+e ν	$2.5246 \cdot 10^{-2}$	1	13000
107685	AlpGen+Jimmy W+e ν	$7.1239 \cdot 10^{-3}$	1	3500
107690	AlpGen+Jimmy W+ $\mu\nu$	6.8711	1	1387000
107691	AlpGen+Jimmy W+ $\mu\nu$	1.2947	1	256000
107692	AlpGen+Jimmy W+ $\mu\nu$	$3.7608 \cdot 10^{-1}$	1	188000
107693	AlpGen+Jimmy W+ $\mu\nu$	$1.0072 \cdot 10^{-1}$	1	51000
107694	AlpGen+Jimmy W+ $\mu\nu$	$2.5993 \cdot 10^{-2}$	1	13000
107695	AlpGen+Jimmy W+ $\mu\nu$	$7.1300 \cdot 10^{-3}$	1	3500
107700	AlpGen+Jimmy W+ $\tau\nu$	6.8733	1	1367000
107701	AlpGen+Jimmy W+ $\tau\nu$	1.2952	1	255000
107702	AlpGen+Jimmy W+ $\tau\nu$	$3.7507 \cdot 10^{-1}$	1	188500
107703	AlpGen+Jimmy W+ $\tau\nu$	$1.0177 \cdot 10^{-1}$	1	50500
107704	AlpGen+Jimmy W+ $\tau\nu$	$2.5756 \cdot 10^{-2}$	1	13000
107705	AlpGen+Jimmy W+ $\tau\nu$	$7.0016 \cdot 10^{-3}$	1	4000

A.6 W+bb

Number	Comments	$\sigma (nb^{-1})$	Generator efficiency ϵ	Number of events
106280	AlpGen+Jimmy	$3.3072 \cdot 10^{-3}$	1	6500
106281	AlpGen+Jimmy	$2.6756 \cdot 10^{-3}$	1	5000
106282	AlpGen+Jimmy	$1.3763 \cdot 10^{-3}$	1	3000
106283	AlpGen+Jimmy	$6.5689 \cdot 10^{-4}$	1	1500

A.7 Z+jets

Number	Comments	σ (nb^{-1})	Generator efficiency ϵ	Number of events
107650	AlpGen+Jimmy Z+ee	$6.6410 \cdot 10^{-1}$	1	304500
107651	AlpGen+Jimmy Z+ee	$1.3299 \cdot 10^{-1}$	1	63500
107652	AlpGen+Jimmy Z+ee	$4.0226 \cdot 10^{-2}$	1	19500
107653	AlpGen+Jimmy Z+ee	$1.1138 \cdot 10^{-2}$	1	5500
107654	AlpGen+Jimmy Z+ee	$2.8925 \cdot 10^{-3}$	1	1500
107655	AlpGen+Jimmy Z+ee	$7.5343 \cdot 10^{-4}$	1	500
107660	AlpGen+Jimmy Z+ $\mu\mu$	$6.6379 \cdot 10^{-1}$	1	304000
107661	AlpGen+Jimmy Z+ $\mu\mu$	$1.3295 \cdot 10^{-1}$	1	63000
107662	AlpGen+Jimmy Z+ $\mu\mu$	$4.0375 \cdot 10^{-2}$	1	19000
107663	AlpGen+Jimmy Z+ $\mu\mu$	$1.1161 \cdot 10^{-2}$	1	5500
107664	AlpGen+Jimmy Z+ $\mu\mu$	$2.8987 \cdot 10^{-3}$	1	1500
107665	AlpGen+Jimmy Z+ $\mu\mu$	$7.5662 \cdot 10^{-4}$	1	500
107670	AlpGen+Jimmy Z+ $\tau\tau$	$6.6250 \cdot 10^{-1}$	1	303500
107671	AlpGen+Jimmy Z+ $\tau\tau$	$1.3394 \cdot 10^{-1}$	1	63500
107672	AlpGen+Jimmy Z+ $\tau\tau$	$4.0295 \cdot 10^{-2}$	1	19500
107673	AlpGen+Jimmy Z+ $\tau\tau$	$1.1029 \cdot 10^{-2}$	1	5500
107674	AlpGen+Jimmy Z+ $\tau\tau$	$2.8040 \cdot 10^{-3}$	1	1500
107675	AlpGen+Jimmy Z+ $\tau\tau$	$7.8054 \cdot 10^{-4}$	1	500
107710	AlpGen+Jimmy Z+ $\nu\nu$	3.5388	$7.7231 \cdot 10^{-3}$	3000
107711	AlpGen+Jimmy Z+ $\nu\nu$	$7.3135 \cdot 10^{-1}$	$6.0738 \cdot 10^{-1}$	44500
107712	AlpGen+Jimmy Z+ $\nu\nu$	$2.2250 \cdot 10^{-1}$	$8.7764 \cdot 10^{-1}$	39500
107713	AlpGen+Jimmy Z+ $\nu\nu$	$6.2001 \cdot 10^{-2}$	$9.6592 \cdot 10^{-1}$	12000
107714	AlpGen+Jimmy Z+ $\nu\nu$	$1.5866 \cdot 10^{-2}$	$9.9012 \cdot 10^{-1}$	8000
107715	AlpGen+Jimmy Z+ $\nu\nu$	$4.3781 \cdot 10^{-3}$	$9.9920 \cdot 10^{-1}$	2500

Appendix B

Quality criteria for tracks used in SV0-algorithm

The tracks entering the SV0 algorithm as explained in chapter 6 need to satisfy basic quality criteria. These are summarized in table B.1.

p_T	> 0.5 GeV
$d_0(w.r. PV)$	< 2 mm
$z_0(w.r. PV) \sin \theta$	< 2 mm
$\sigma(d_0(w.r. PV))$	< 1 mm
$\sigma(z_0(w.r. PV))$	< 5 mm
$\chi^2/ndof$	< 3
Number of Pixel hits	≥ 2
Number of SCT hits	≥ 4
Number of Pixel+SCT hits	≥ 7

Table B.1: Track selection cuts as used by the SV0 algorithm. [\[44\]](#)

Appendix C

Event displays in the 1-Lepton + 3 jet channel

The cuts of the control sample which were defined in chapter 6 are applied to data of period A -D3. As the integrated luminosity is small with approximately 300 nb^{-1} , only 13 events survive the selection cuts of the control sample. The following cuts are used as cuts for the control sample:

1. good run list (for data) and trigger: L1_MU6 for muons
2. jet quality cuts
3. at least one vertex with more than 4 tracks
4. one non-isolated muon with $p_T > 20 \text{ GeV}$
5. 3 jets with $p_T > 40., 30., 30. \text{ GeV}$
6. at least one jet with a SV0-weight above 8
7. at least one more jet with a SV0-weight above 5

As definition for MET was used: $MET_{x,y} = MET_Topo_etx,y + p_{x,y}(selected\ muon)$ (corresponds to the recommendation of the JetEtmiss group in early summer.)

Applying these cuts leads to the cutflow in table C.1.

The surviving events were found in the runs 155112 (1 event), 155697 (2 events), 159041 (2 events), 159086 (6 events) and 158582 (2 events). In the following, event displays for a selection of surviving events are shown which were created with the help of the ATLAS tool

cut	data (304.7 nb^{-1})
1	2632780
2	2632440
3	2371010
4	744
5	122
6	49
7	13

Table C.1: Cutflow for period A - D with the MET definition $MET_{x,y} = MET_Topo_etx,y + p_{x,y}(selected\ muon)$

Atlantis [58]. In the event display in figure C.1 the b-jets are marked with blue color, the muon with magenta, all other jets with a dark red. MET is indicated by an orange arrow. These event displays indicate that indeed events with a non-isolated muon from a b-jet are found. MET can be aligned with the muon and the jet, but does not need to be. In general, the selection cuts select events as wished.

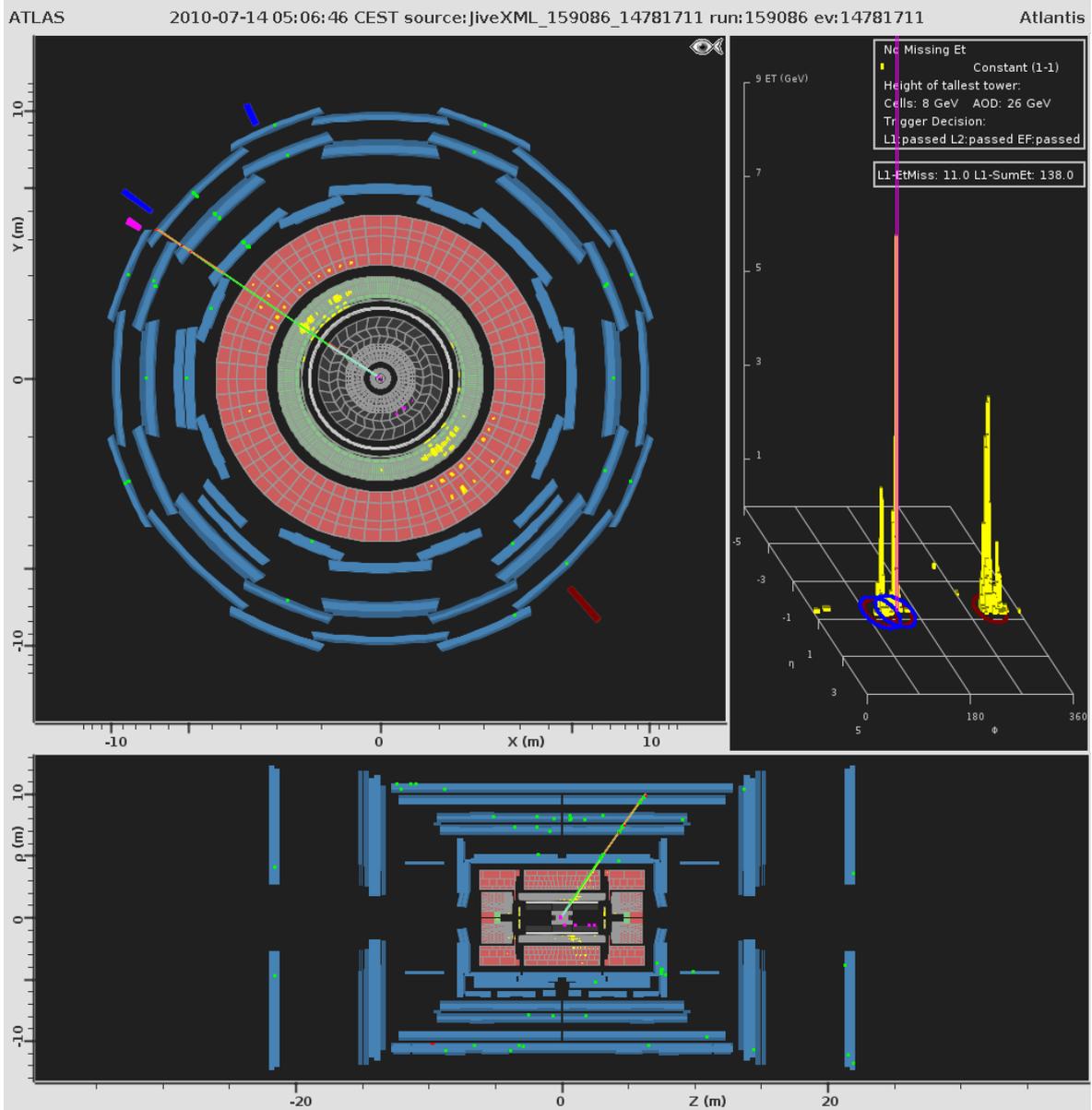
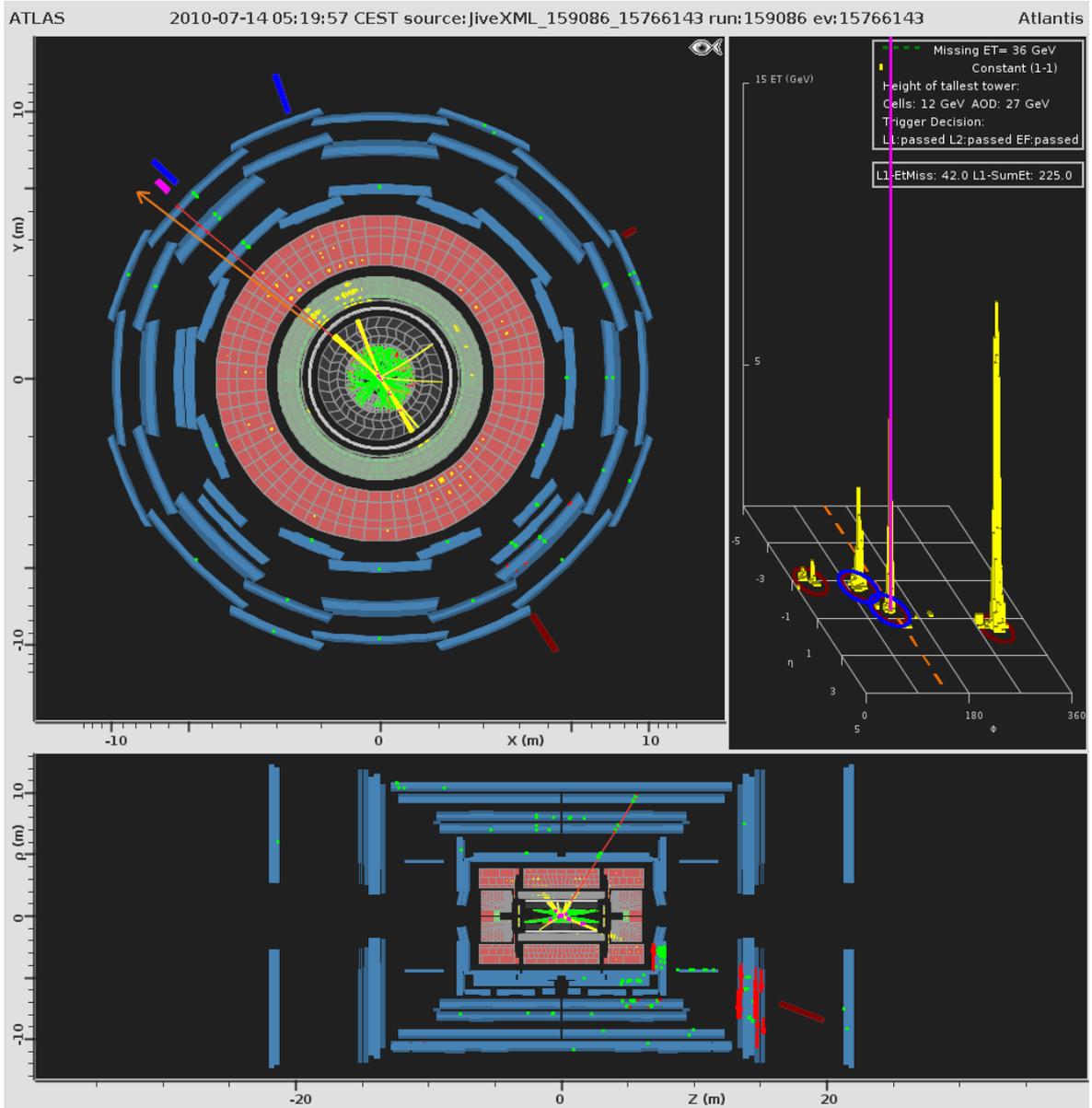
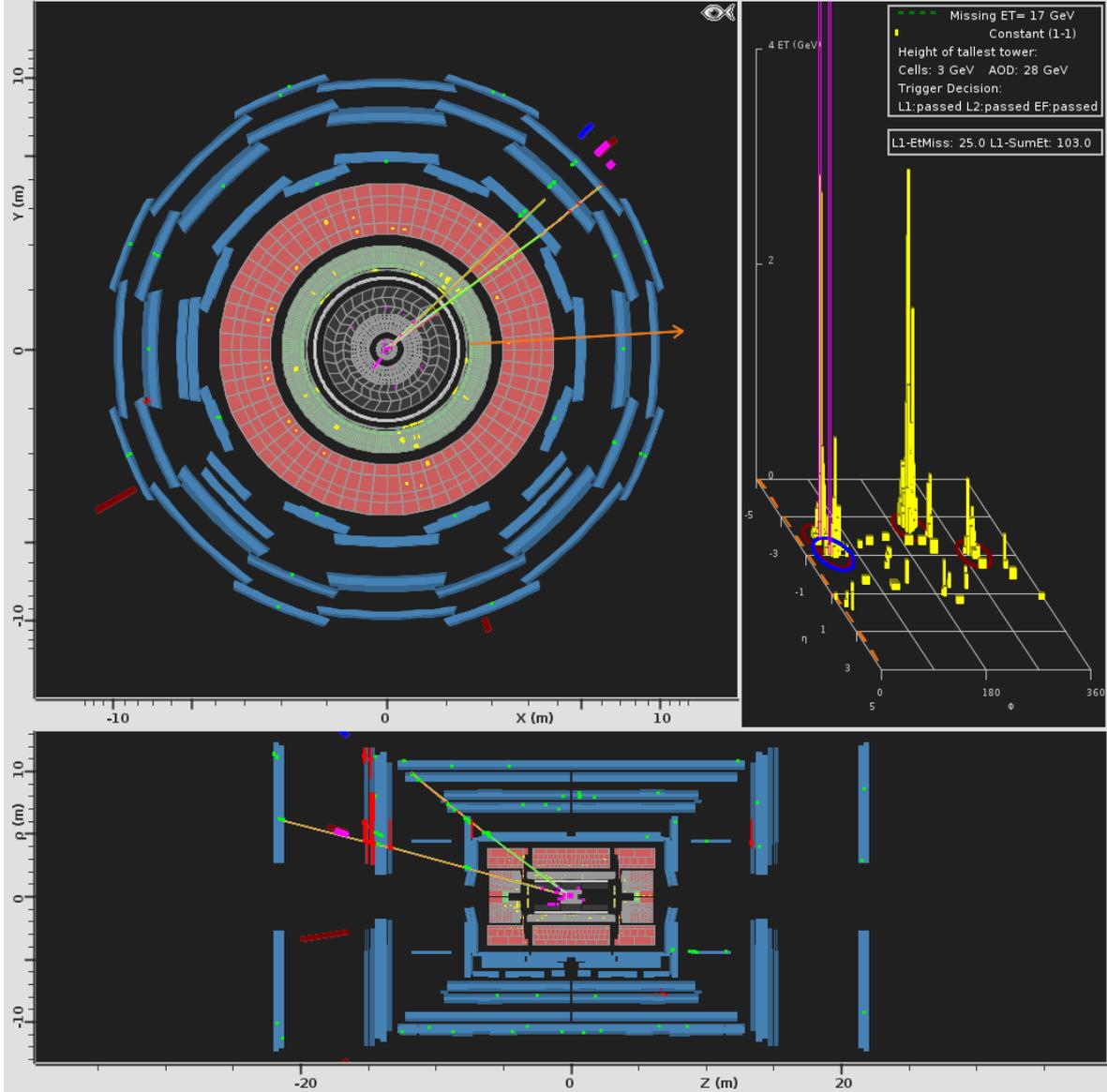


Figure C.1: Event display of one surviving event after all cuts.





Bibliography

- [1] The ATLAS Collaboration. Expected Performance of the ATLAS Experiment. Detector, Trigger and Physics. 2008. CERN-OPEN-2008-020.
- [2] D. Griffiths. Introduction to Elementary Particles. Second, revised Edition. Wiley-VCH 2008.
- [3] I. J. R. Aitchison, Anthony J. G. Hey. Gauge Theories in Particle Physics. Volume II.: QCD and the Electroweak Theory. Third Edition. Institute of Physics Publishing 2004.
- [4] J. L. Rosner. The Standard Model in 2001. 2002. arXiv:hep-ph/0108195v6
- [5] B. Povh. K. Rith. C. Scholz. F. Zetsche. Teilchen und Kerne. 7. Auflage. Springer Verlag 2006.
- [6] M. E. Peskin. Daniel V. Schroeder. An Introduction to Quantum Field Theory. Westview Press 1995.
- [7] Particle Data Group. Review of Particle Physics. Published in Journal of Physics G. Nuclear and Particle Physics. Volume 37. Number 7A. July 2010. Article 075021. <http://pdg.lbl.gov/>
- [8] S. P. Martin. A Supersymmetry Primer. Version 5. 2008. arXiv:hep-ph/9709356v5.
- [9] I. J. R. Aitchison. Supersymmetry and the MSSM: An Elementary Introduction. 2005. arXiv:hep-ph/0505105v1.
- [10] K. Müller. Einführung Supersymmetrie. 2002. <http://www.physik.uzh.ch/~kmueller/text/vorlesung/susy.pdf>
- [11] T. Sjöstrand. S. Mrenna. P. Skands. PYTHIA 6.4. Physics and Manual. 2006. arXiv:hep-ph/0603175.
- [12] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. JHEP 0307:001,2003, arXiv:hep-ph/0206293.
- [13] S. Frixione, P. Nason, B.R. Webber. Matching NLO QCD and parton showers in heavy flavour production. JHEP 0308 (2003) 007. 2003. arXiv:hep-ph/0204244. and <http://www.hep.phy.cam.ac.uk/theory/webber/MCatNLO/>
- [14] The ATLAS Collaboration. Computing Technical Design Report. 2005. <http://atlas-proj-computing-tdr.web.cern.ch/atlas-proj-computing-tdr/Html/contents.html>
- [15] Rene Brun et al. ROOT. A Data Analysis Framework. <http://root.cern.ch>

- [16] The SUSYD3PDMaker. Juli 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SUSYD3PDMaker>
- [17] M.A. Dobbs et al. Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics. 2004. arXiv:hep-ph/0403045v2
- [18] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytrcek, G. Cooperman et al. G4—a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 506, Issue 3, 1 July 2003, Pages 250-303 and <http://geant4.cern.ch/>
- [19] M. Bosman et al. Understanding Monte Carlo Generators for Top Physics. ATLAS note. 2009. ATL-COM-PHYS-2009-334.
- [20] P. Bryant, L. Evans. LHC machine. 2008. Jinst 3 S08001.
- [21] S. Redaelli. The operation of the LHC accelerator complex. Part 1 & Part 2. LHC Physics Centre at CERN - Student lectures. 7th an 9th April 2010.
- [22] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. 2008. Jinst 3 S08001.
- [23] The ATLAS Collaboration. Reconstruction and Identification of Electrons. 2008. In [1]
- [24] The ATLAS Collaboration. Electron and photon reconstruction and identification in ATLAS: Expected performance of high energy and results at 900 GeV. ATLAS note. ATLAS-CONF-2010-005.
- [25] The ATLAS Collaboration. Observation of $W \rightarrow l\nu$ production and search for $Z \rightarrow ll$ in proton-proton collisions at 7 TeV with the ATLAS detector. ATLAS note. ATLAS-CONF-2010-051.
- [26] Twiki-page of the ATLAS EGamma group about electron reconstruction. Last accessed October 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronReconstruction>
- [27] Twiki-page of the ATLAS EGamma group about electron identification. Last accessed October 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronIdentification>
- [28] The ATLAS Collaboration. Expected electron performance in the ATLAS experiment. 2010. ATLAS note. ATL-PHYS-INT-2010-126.
- [29] The ATLAS Collaboration. Muon Reconstruction and Identification: Studies with Simulated Monte Carlo Samples. 2008. In [1].
- [30] Twiki-page of the ATLAS MuonPerformance Group. October 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MuonPerformance>
- [31] B. Lenzi, D. L. Mateos. Muon Reconstruction: Algorithms and Analysis with Muons. In Offline Software tutorial of May 2010.
- [32] The ATLAS Collaboration. Jet reconstruction performance. 2008. In [1].
- [33] P. Loch. Introduction to HadronicFinal State Reconstruction in Collider Experiments. 2010. http://atlas.physics.arizona.edu/~loch/HFSL_spring2010.html

- [34] The ATLAS Collaboration. Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. 2010. ATLAS-CONF-2010-053.
- [35] M. Cacciari. G. P. Salam. The anti- k_t jet clustering algorithm. 2008. arXiv:hep-ph/0802.1189v2
- [36] Recommendations of the JetEtmis ATLAS group concerning the jet cleaning. Last accessed October 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HowToCleanJets>
- [37] Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s} = 7$ TeV and measured with the ATLAS detector. ATLAS note. 2010. ATLAS-CONF-2010-056.
- [38] P. Loch. Brief introduction on Missing Transverse Energy Reconstruction at LHC. http://atlas.physics.arizona.edu/~loch/material/lectures/pdf/Extra_MET.pdf
- [39] The ATLAS Collaboration. Performance of Missing Transverse Energy Reconstruction and Calibration in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV with the ATLAS Detector. ATLAS note. ATLAS-CONF-2010-057.
- [40] The ATLAS Collaboration. Measurement of Missing Transverse Energy. 2008. In [1].
- [41] Twiki of MET supgroup of ATLAS JetEtmis group. July and August 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EtMiss>
- [42] The ATLAS Collaboration. b-Tagging Performance. 2008. In [1].
- [43] The ATLAS Collaboration. Early supersymmetry searches in events with missing transverse energy and b-jets with the ATLAS detector. ATLAS note. 2010. ATLAS-CONF-2010-079.
- [44] The ATLAS Collaboration. Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data. ATLAS note. 2010. ATLAS-CONF-2010-042.
- [45] Recommendations of the ATLAS b-tagging group concerning the early taggers. Last accessed October 2010. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingForEarlyData#TrackConting2D_Tagger
- [46] The ATLAS Collaboration. Prospects for Supersymmetry Discovery Based on Inclusive Searches. 2008. In [1].
- [47] The ATLAS Collaboration. Supersymmetry Searches. 2008. In [1].
- [48] Public CERN Web page. September 2010. <http://www.cern.ch>
- [49] GoodRunLists provided for the SUSY working group. Corresponding twiki-page: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/GoodRunListSummer2010>.
- [50] Corresponding integrated luminosity to the linked GoodRunLists. Twiki-page: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/LuminositySummer2010>
- [51] Liu, Tiankuan. Optical Links for ATLAS Liquid Argon Calorimeter Front-end Electronics Readout. 2010. ATL-LARG-PROC-2010-016
- [52] Official object definitions by the SUSY working group. Last accessed December 2010. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SusyObjectDefintions>

- [53] The ATLAS Collaboration. Search for Supersymmetry with jets and missing transverse momentum and one lepton at $\sqrt{s} = 7$ TeV (supporting INT note). ATLAS note. In preparation.
- [54] The ATLAS Collaboration. Early supersymmetry searches with jets, missing transverse momentum and one or more leptons with the ATLAS Detector. ATLAS note. 2010. ATLAS-COM-CONF-2010-067.
- [55] A.J. Barr et al. Search for supersymmetry using final states with jets, one isolated electron or muon, and missing transverse momentum with the ATLAS detector in $\sqrt{s} = 7$ TeV proton-proton collisions at the LHC. In preparation.
- [56] B. Abi et al. Mis-identified lepton backgrounds to top quark pair production. ATLAS-note. In preparation.
- [57] ATLAS Metadata Interface. Last accessed December 2010. <http://ami.in2p3.fr/opencms/opencms/AMI/www>
- [58] ATLANTIS. Event display for ATLAS. July 2010. <http://www.hep.ucl.ac.uk/atlas/atlantis/>

Acknowledgements

Zum Gelingen dieser Arbeit haben viele Leute beigetragen, sei es durch die gute Arbeitsatmosphäre in Garching und am CERN oder durch die zahlreichen fachlichen Impulse in SUSY Meetings. Ihnen möchte ich dafür sehr danken. Ganz besonders möchte ich mich bei den folgenden Personen bedanken:

- Prof. Dr. Dorothee Schaile möchte ich für die Stellung des Themas danken und für die damit verbundene Möglichkeit in der SUSY Gruppe sowohl am Lehrstuhl als auch am CERN zu arbeiten. Außerdem möchte ich ihr für die ausgezeichnete Betreuung dieser Arbeit und für die Ermöglichung längerer Aufenthalte am CERN danken.
- Prof. Dr. Gerhard Buchalla danke ich für die Übernahme des Zweitgutachtens.
- Dr. Marie-Helene Genest möchte ich ebenfalls für die ausgezeichnete Betreuung dieser Arbeit danken.
- Tapas Sarangi von der Universität von Wisconsin möchte ich für einige Diskussionen und Inspirationen das QCD $b\bar{b}$ Kontroll Sample und die QCD $b\bar{b}$ Untergrundbestimmung betreffend danken.
- George Redlinger danke ich für einige hilfreiche Kommentare das QCD $b\bar{b}$ Kontroll Sample betreffend.
- Dr. Renaud Brunelieri von der Universität Freiburg danke ich für Diskussionen und einige Inspirationen bezüglich der Matrixmethode.
- Dr. Felix Rauscher möchte ich für die zahlreichen Tipps das Programmieren betreffend danken und für die Einführung in das Grid.
- Dr. Jörg Stelzer vom DESY möchte ich für die zahlreichen Hinweise zum Programmieren in Python danken.
- Frau Franz möchte ich sehr für die exzellente Abwicklung administrativer Fragen dieses Jahr danken.
- Bei meiner Familie möchte ich mich sehr für die Unterstützung dieses Jahr bedanken und besonders bei meinen Eltern für die stetige Förderung meiner Interessen.

Erklärung

Ich erkläre hiermit, dass ich meine Diplomarbeit mit dem Titel
**Study of Missing Transverse Energy in Heavy
Flavor Decays**

selbstständig verfasst sowie keine anderen als die angegebenen
Hilfsmittel verwendet habe.

München, den 14. Dezember 2010

(Jeanette Lorenz)