

Department of Physics and Astronomy

University of Heidelberg

Master Thesis in Physics

submitted by

Mathis Kolb

born in Heidelberg

2014



**Top quark reconstruction using the  
“Buckets of tops”  
method in the ATLAS experiment**

This Master thesis has been carried out by Mathis Kolb

at the

Physikalisches Institut Heidelberg

under the supervision of

Prof. Dr. André Schöning



## Abstract

The “Buckets of tops” method is used to reconstruct hadronically decaying top quark pairs in the ATLAS experiment at the LHC. The method targets a moderate transverse momentum regime of the top quarks in the range  $p_T^{\text{top}} = 100 - 400$  GeV and complements existing methods of resolved or substructure based reconstruction of top quarks. The method assigns Anti- $k_T$  ( $R = 0.4$ ) jets to three “Buckets”. These correspond to the two top quarks and the extra hadronic activity of the proton-proton collision. The performance and applicability of the method for the ATLAS experiment is investigated in this thesis. The method is further validated in a Monte-Carlo to data comparison. In this study the reconstruction efficiency, the power to suppress the background and the dependence on pile-up is investigated. The performance of the method is illustrated by a simulation study of scalar top quark partner pair production with a decay to top quark pairs. The reconstruction of the top quark four momenta achieved by the “Buckets of tops” method allows an improved background suppression.

## Zusammenfassung

Die “Buckets of tops” Methode wird angewandt um hadronisch zerfallende Top-Quark Paare im ATLAS Experiments am LHC zu rekonstruieren. Die Methode eignet sich insbesondere für moderate transversale Impulse der Top-Quarks im Bereich  $p_T^{\text{top}} = 100-400$  GeV und ergänzt sowohl traditionelle als auch substruktur-basierte Methoden der Top-Quark Rekonstruktion. Es werden Anti- $k_T$  ( $R=0.4$ ) Jets in drei “Buckets” aufgeteilt. Diese entsprechen den beiden Top-Quarks und der zusätzlichen hadronischen Aktivität der Proton-Proton Kollision. Die Arbeit behandelt die Leistungsfähigkeit und Anwendungsmöglichkeiten der Methode innerhalb des ATLAS Experiments. Die Methode wird des weiteren in einem Monte-Carlo zu Daten Vergleich überprüft. In dieser Studie wird die Rekonstruktionseffizienz, die Fähigkeit zur Unterdrückung des Untergrunds, sowie die Abhängigkeit von Pile-up untersucht. Die Leistungsfähigkeit der Methode wird anhand einer auf Simulation basierten Studie skalarer Top-Quark Partner, die in Top-Quark Paare zerfallen, verdeutlicht. Die Rekonstruktion des Viererimpuls der Top-Quarks durch die “Buckets of tops” Methode ermöglicht eine verbesserte Unterdrückung des Untergrunds.



## Preface

First of all, I would like to thank my supervisor Prof. Dr. André Schöning for his continuous motivating guidance and for introducing the interesting topic of top quark reconstruction to me which eventually resulted in this master thesis. I would also like to express my sincere gratitude to Prof. Dr. Tilman Plehn for doing the second correction. Further, I am thankful to Dr. Christoph Anders for his very capable and ever-patient support during the course of this thesis. I benefited greatly from the wonderful working atmosphere provided by the former and current members of the ATLAS group in Heidelberg, especially Maddalena Giulini, Dr. Gregor Kasieczka, Dr. Tatsiana Klimkovich, Rohin Narayan, Priv.-Doz. Dr. Sebastian Schätzel, Dr. Sebastian Schmitt and David Sosa. Finally, I thank my family and friends for their support.



# Contents

<b>1</b>	<b>Motivation and outline</b>	<b>11</b>
<b>2</b>	<b>Theory</b>	<b>13</b>
2.1	Standard Model of Particle Physics . . . . .	13
2.2	Hadron Collider Physics . . . . .	14
2.2.1	Proton-Proton Collisions and Monte Carlo . . . . .	15
2.2.2	Jet Clustering . . . . .	16
2.2.3	Pile-up . . . . .	18
2.3	Top quark physics . . . . .	19
2.4	Beyond Standard Model Physics . . . . .	20
2.4.1	Supersymmetry . . . . .	20
2.4.2	Heavy resonances . . . . .	21
<b>3</b>	<b>Experimental setup</b>	<b>23</b>
3.1	LHC . . . . .	23
3.2	ATLAS Detector . . . . .	27
3.2.1	Inner Detector . . . . .	28
3.2.2	The Calorimeters . . . . .	28
3.2.3	The Muon System . . . . .	30
3.2.4	Trigger System . . . . .	30
3.3	Data and Monte-Carlo samples . . . . .	32
3.4	Reconstruction and selection of physics objects and events . . . . .	33
3.4.1	Electrons . . . . .	33
3.4.2	Muons . . . . .	34
3.4.3	Tracks and vertices . . . . .	34
3.4.4	Topological clusters . . . . .	35
3.4.5	Jets . . . . .	35
3.4.6	Bottom-jets . . . . .	37
3.4.7	Overlap removal . . . . .	38
3.4.8	Event selection . . . . .	38
3.4.9	Missing transverse energy . . . . .	39
3.4.10	Trigger . . . . .	39
<b>4</b>	<b>Description of the Bucket Algorithm</b>	<b>41</b>

<b>5</b>	<b>Performance of the Bucket Algorithm</b>	<b>48</b>
5.1	Performance of simulated Standard Model events . . . . .	49
5.2	Validation with $Z'$ boson samples . . . . .	58
<b>6</b>	<b>Model to Data comparison for the Bucket Algorithm</b>	<b>62</b>
6.1	Trigger study . . . . .	62
6.2	Data to Monte-Carlo comparison . . . . .	65
6.3	Background estimate from data . . . . .	65
6.3.1	Construction and validation of background template . . . . .	68
6.3.2	Data to Monte-Carlo plus background model comparison in the signal region . . . . .	72
<b>7</b>	<b>Study of scalar top partner decays</b>	<b>76</b>
7.1	MT2 . . . . .	77
7.2	Monte-Carlo samples . . . . .	79
7.3	Signal Selection and Comparison . . . . .	79
<b>8</b>	<b>Discussion and outlook</b>	<b>85</b>
	<b>Bibliography</b>	<b>88</b>

# 1 Motivation and outline

The Standard Model of Particle Physics (SM) describes all known elementary particles and their interactions. This includes the strong, weak and electromagnetic interactions. The SM successfully predicted many experimental discoveries over the last decades. It agrees to high precision with almost all current experimental data. Shortcomings of the SM involve e.g. the cosmological evidence for dark matter, a missing description of gravity and the baryon asymmetry of the universe or the fine-tuning problem. The observation of the Higgs boson [1, 2] at relatively low mass makes the fine-tuning problem even more pronounced.

This discovery is one of the main achievements of the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) in Geneva. The LHC accelerates protons and brings them to collision at energies never reached before in an experiment. Hence, it provides the power to test models for physics beyond the standard model (BSM), existing at higher scales as well as to investigate the electroweak symmetry breaking in more detail.

The Higgs boson discovery completed the SM and proved that it is indeed a renormalizable theory. Renormalizability allows to extend the SM from small to higher energy scales. The top quark is the heaviest quark with a mass around the electroweak scale and a width so large that it decays before hadronization. Due to the high center of mass energy of the LHC large numbers of top quark pairs are produced.

The past years of LHC running have already shown that the study of top quarks can be an important tool to investigate the above mentioned research areas. This requires a good understanding of the top quark decay and techniques to identify and reconstruct them. Traditional resolved reconstruction [3] suffers from relatively small efficiencies at higher transverse momentum of the top quark when the decay products measured as jets begin to merge. In contrast, substructure based reconstruction [4, 5, 6, 7, 8] requires large jets (fatjets) as a starting point. As a consequence a threshold at around  $p_T^{\text{top}} \approx 200$  GeV where the fatjets can be identified is introduced. In addition, top quark pairs decay in 46% of the events into complete hadronic final states. This decay channel is particularly challenging as it has to compete with large background due to QCD multijet production.

In Reference [9] a novel approach for top quark pair reconstruction was proposed

targeting a moderate transverse momentum regime of the top quarks:

$$p_T^{\text{top}} = 100 - 400 \text{ GeV}$$

It aims at sorting relatively small sized jets into “buckets” which look like top quarks while also considering extra hadronic activity like initial state radiation in an event. In this thesis first simulation tests of this algorithm in experimental conditions as present in the ATLAS detector are performed and further validated in a data to Monte-Carlo comparison. Due to the focused momentum regime and the high efficiency, the algorithm should be able to complement existing top quark reconstruction techniques.

The thesis is structured as follows: A brief introduction into the basic theoretical concepts is given in Chapter 2, followed by an overview of the LHC and the ATLAS detector in Chapter 3. In Chapter 4 the bucket algorithm for the reconstruction of top quark pairs in the fully hadronic decay mode is introduced. In Chapter 5, the performance of the algorithm is investigated in a simulation study taking into account realistic experimental conditions. The quality of the signal and background modeling is further validated in a Monte-Carlo to data comparison study in Chapter 6. The algorithm is then applied to scalar top quark partner searches in a simulation study in Chapter 7.

# 2 Theory

Some basic concepts of the Standard model are summarized with emphasis on the top quark as it appears at hadron colliders. In the following two scenarios for physics beyond the standard model are sketched.

## 2.1 Standard Model of Particle Physics

The Standard Model of Particle Physics is the currently most successful theory to describe elementary particles and their interactions, except for gravity. The SM is a quantum field theory based on the gauge group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . Due to the discovery of the Higgs boson at the LHC the structure of the SM is completed.  $SU(3)_C$  describes the strong interaction and is a nonabelian gauge group. The electroweak part is described by the  $SU(2)_L \times U(1)_Y$  component. The underlying spacetime symmetry is characterized by the Lorentz group.

Fermions contained in the SM can be classified in quarks and leptons. All of them come in three generations called flavors. Leptons do not interact via the strong force. Quantum numbers of the fermions and the Higgs boson are summarized in Table 2.1.

The SM Lagrangian is composed of four parts.

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermions}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Scalar}}$$

It consists of all possible gauge invariant and renormalizable terms in concordance with the assignment of quantum numbers in Table 2.1. The gauge sector  $\mathcal{L}_{\text{Gauge}}$  contains the gauge fields  $G_\mu^i$ ,  $W_\mu^i$  and  $B_\mu^i$  of the  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge group. Its corresponding fermion interactions are described in  $\mathcal{L}_{\text{Fermions}}$ . Fermions acquire their mass through interactions to the Higgs field described in the Yukawa sector  $\mathcal{L}_{\text{Yukawa}}$ . The scalar part is contained in  $\mathcal{L}_{\text{Scalar}}$ . A detailed description of the structure of each term and the electroweak symmetry breaking can be found in e.g. Reference [10]. According to Noether's theorem each symmetry can be associated to a conserved current and therefore conserved charges. Eight massless gluons and the massless photon comprise the bosonic particles. Furthermore, the electroweak gauge bosons give rise to massive  $W^\pm$  bosons mediating charged currents and a  $Z$  boson mediating neutral currents.

Each fermion generation has a charged lepton (electron, muon, tau) and a neutral lepton called neutrino ( $\nu_e, \nu_\mu, \nu_\tau$ ). In addition, there are two massive quarks labeled

field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	<b>3</b>	<b>2</b>	1/3
$u_R$	<b>3</b>	<b>1</b>	4/3
$d_R$	<b>3</b>	<b>1</b>	-2/3
$l_L = \begin{pmatrix} \nu_L \\ l_L \end{pmatrix}$	<b>1</b>	<b>2</b>	-1
$l_R$	<b>1</b>	<b>1</b>	-2
$\phi$	<b>1</b>	<b>2</b>	1

Table 2.1: Representations and quantum numbers of SM fermions and the Higgs field  $\phi$ . The hypercharge definition uses the convention  $Y = 2(Q - I_3)$  where  $Q$  is the electromagnetic quantum number and  $I_3$  is the third isospin component. Furthermore, left ( $L$ ) and right ( $R$ ) handed fields are distinguished.

up- or down-type quark. Due to their mass, quarks can interact between different generations. This mixing of the quarks is described by the  $CKM$  matrix which is a matrix close to unity.

$$U_{CKM} \simeq \begin{pmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 1 & \epsilon \\ \epsilon & \epsilon & 1 \end{pmatrix} \quad (2.1.1)$$

The values of the entries of the  $CKM$  matrix influence the strength of charged current interactions between the quarks mediated by the  $W^\pm$  bosons. Considering the couplings, the parameters of the scalar sector, the nine fermion masses and four  $CKM$  parameters there are 18 parameters describing the SM. There is only one dimensionfull parameter  $v_0 = 246$  GeV characterizing the corresponding electroweak scale. The observation of neutrino oscillations implies that neutrinos are massive as well. Describing these masses and the associated mixing would introduce further parameters into the model.

## 2.2 Hadron Collider Physics

The Large Hadron Collider (LHC) accelerates protons to high energies and brings them to collision. Protons are hadrons consisting of elementary particles like quarks and gluons. The interactions of quarks and gluons can be described by Quantum Chromodynamics (QCD). It is the  $SU(3)$  component of the  $SU(3) \times SU(2) \times U(1)$  Standard Model of Particle Physics (SM) and therefore a non-abelian gauge theory of the strong force. The running of the associated strong coupling  $\alpha_s$  can

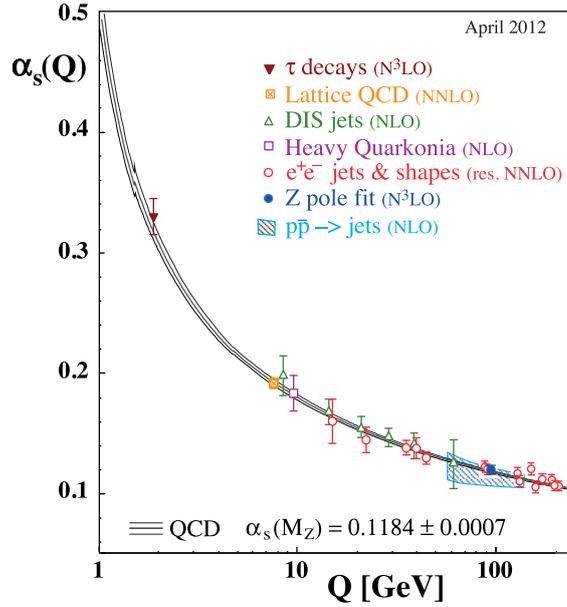


Figure 2.1: Measurements of  $\alpha_s$  at different energy scales  $Q$ , from [11].

be described in a simplified way by

$$\alpha_s(q^2) = \frac{1}{b_0 \log \frac{q^2}{\Lambda_{\text{QCD}}^2}} \quad (2.2.1)$$

where  $b_0$  is the 1-loop beta-function coefficient and  $\Lambda_{\text{QCD}}$  the corresponding Landau pole of the strong coupling. The measured energy scale dependence is shown in Figure 2.1 together with the corresponding value at the  $Z$  boson mass  $\alpha_s(m_Z) = 0.1184 \pm 0.0007$  [11]. The coupling becomes large for small energy scales and the quarks and gluons are not free anymore but form hadrons as baryons or mesons. This phenomenon is called confinement. In contrast, the strong coupling becomes small for so called hard processes involving large energy scales. Therefore the process can be calculated perturbatively.

## 2.2.1 Proton-Proton Collisions and Monte Carlo

For describing  $pp$  collisions  $pp \rightarrow X$  with a final state  $X$  the factorization theorem of QCD allows to separate the cross section of the process ( $\sigma$ ) in a perturbative and non-perturbative part. The hard process is described at parton level ( $\hat{\sigma}$ ) whereas the non-perturbative part is characterized by parton distribution functions (pdf)  $f_i(x)$  describing the fraction  $x$  of the proton momentum carried by an incoming

parton  $i$ :

$$\sigma = \int_0^1 dx_1 \int_0^1 dx_2 \sum_{ij} f_i(x_1) f_j(x_2) \hat{\sigma}_{ij}(x_1 x_2 s) \quad (2.2.2)$$

here  $s$  is the center of mass energy of the two colliding protons. The calculation of pdfs at arbitrary scales is not possible. But measurements at lower energies can be extrapolated to higher scales like the ones at the LHC by evolution functions.

As the final states like e.g. top quark pairs decay they loose energy and hadronize. Therefore they can no longer be described perturbatively. This hadronization process is described by various parton shower (PS) models like e.g. Lund string fragmentation [12].

Monte-Carlo (MC) Generators are generally used to simulate physics processes at the LHC. General purpose Monte-Carlo generators interface leading order matrix elements with parton shower. The parton showers can describe processes at arbitrary high orders in  $\alpha_s$  but only in the leading logarithmic approximation of transverse momentum. Therefore Monte-Carlo simulations can also describe the underlying event and initial or final state radiation.

In Figure 2.2 production rates for several important SM processes at different center of mass energies are shown. The majority of events involves soft processes with small transverse energy of the final state particles. Signals involving rare physics can come at rates many orders of magnitude smaller than the total cross section.

## 2.2.2 Jet Clustering

In QCD all freely existing particles are color neutral. The partons involved in the hard scattering like quarks and gluons are colored particles. After the fragmentation and hadronization process they are combined to form hadrons. These generally collimated sprays of hadrons are observed in the detector by the tracks and energy deposits they produce. These observed objects e.g. topological clusters in the ATLAS detector (see Chapter 3.4.4) are referred to as particles in the following. The task of a jet algorithm is to reconstruct a jet consisting of these particles which can than be associated to the initial parton. The procedure of the grouping of the particles should not depend on soft gluon radiation from a parton or on collinear parton splitting. These requirements are called the infrared or collinear safety of the algorithm. One class of jet algorithms providing infrared and collinear safety are sequential recombination algorithms. They take all particles as a starting point and than sequentially combine pairs of two neighboring objects according to two distance measures  $d_{ij}$  and  $d_{iB}$ . The  $d_{ij}$  is the distance between two objects whereas

proton - (anti)proton cross sections

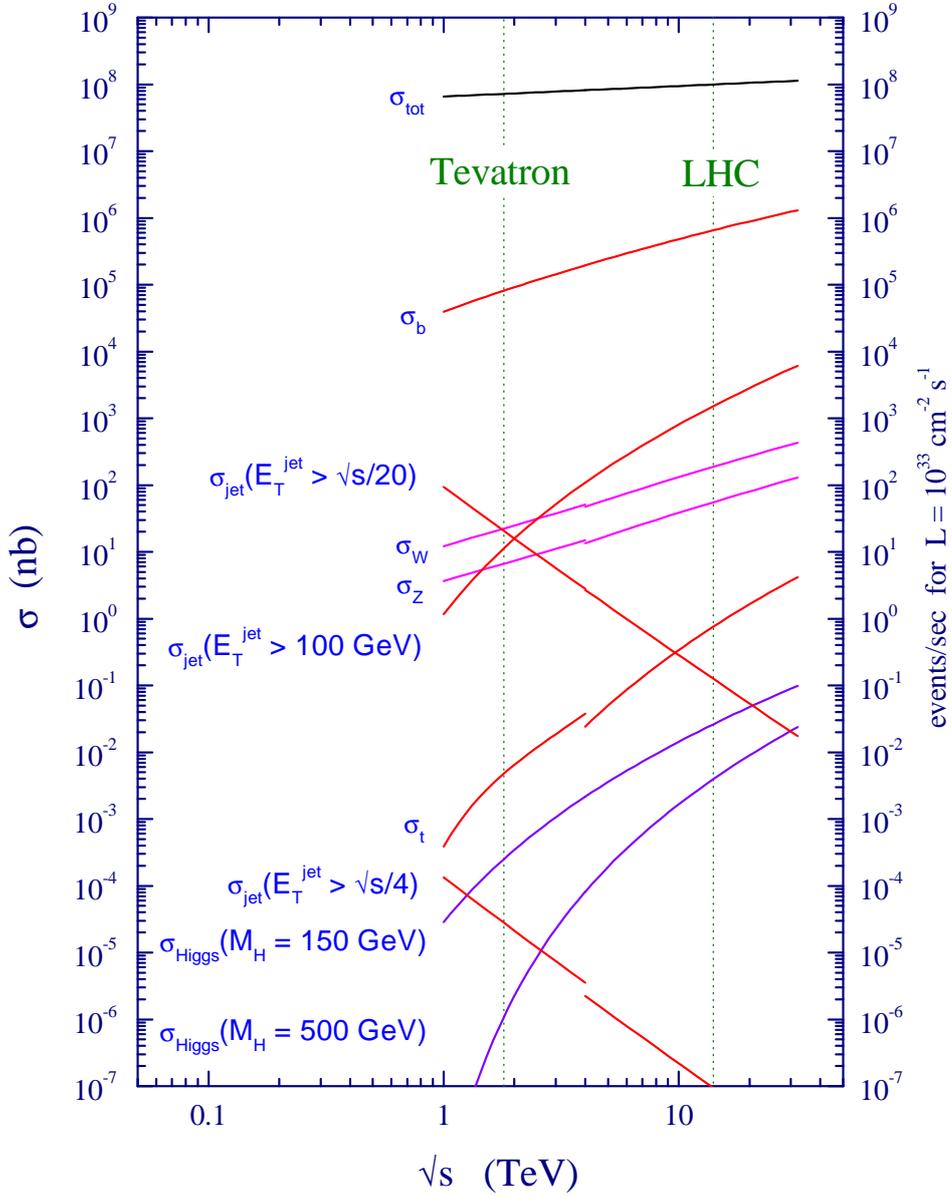


Figure 2.2: Standard Model cross sections at Tevatron and LHC collider for different center of mass energies. The indicated  $\sqrt{s}$  for the LHC corresponds to its design energy, from [13].

$d_{iB}$  is the distance of an object to the beam:

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta R_{ij}}{R} \quad (2.2.3)$$

$$d_{iB} = p_{Ti}^{2n} \quad (2.2.4)$$

The parameter  $R$  defines the “jet radius”. Here  $\Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta y_{ij}^2}$  is the geometrical distance of two objects in the  $y - \phi$  plane<sup>1</sup> and  $p_{Ti}$  is the transverse momentum of the object. The integer  $n$  defines whether the Anti- $k_T$  ( $n = -1$ ), the Cambridge/Aachen ( $n = 0$ ) or the  $k_T$  ( $n = 1$ ) algorithm is considered. Two objects are combined by adding their four-momenta if for all combinations of two objects  $i$  and  $j$  the smallest  $d_{ij}$  is less than the smallest beam distance  $d_{iB}$ . If in contrast the smallest beam distance  $d_{iB}$  is less than the smallest  $d_{ij}$  the object is removed from the list of objects and labeled as a final jet. When all objects are assigned to jets the procedure is completed.

The jet radius parameter  $R$  is by construction the minimal geometrical distance between the final jets. However, the jets constructed this way can have a quite irregular shape. One measure of the area of a jet is the “jet area” [14] as implemented in **FastJet** [15, 16]. A uniform and dense distribution of soft “ghost” particles is introduced. The number of ghosts associated to each jet then defines its area. In Figure 2.3 the jet area normalized to a circular area  $\pi R^2$  is shown for Anti- $k_T$  ( $R = 0.4$ ), Anti- $k_T$  ( $R = 0.6$ ) and Cambridge/Aachen ( $R = 1.2$ ) jets. The Anti- $k_T$  jets are more circular shaped. As harder objects are clustered first in the Anti- $k_T$  algorithm the hard jets ( $p_T^{\text{jet}} \geq 40$  GeV) tend to be insensitive to soft radiation around the boundaries and to form a circular shape with an area close to  $\pi R^2$ . However, non-isolated low  $p_T$  jets will show a more irregular (often crescent) shape [17].

### 2.2.3 Pile-up

Pile-up is denoting the minimum bias (MB) and underlying event (UE) present in a collider like the LHC. The MB is the non-single diffractive part of the inelastic cross section. It comprises roughly speaking all accepted inelastic collisions of two protons with the only requirement that there is some activity in the detector. Therefore, in principal it has some overlap with high  $p_T$  events. The pp cross-sections for the LHC at centre-of-mass energy of 14 TeV are estimated [18] using PYTHIA. The total pp cross-section is  $\sigma_{\text{total}} = 102$  mb. The inelastic part is given by  $\sigma_{\text{inel}} = 79$  mb. The MB cross-section with  $\sigma_{\text{MB}} = 65$  mb contributes the most to it.

In any hard proton scattering process additional present activity is referred to as underlying event. The UE receives contributions from interactions of partons in the

---

<sup>1</sup>A right-handed coordinate system is used. The polar angle is measured with respect to the  $z$ -axis. The azimuthal angle is measured with respect to the  $x$ -axis. The rapidity  $y$  is defined as  $y = 0.5 \times \ln[(E + p_z)/(E - p_z)]$ , where  $E$  represents the energy and  $p_z$  is the momentum component in the  $z$ -direction. The transverse objects are in the  $x - y$  plane with  $p_T = p \times \sin \theta$ .

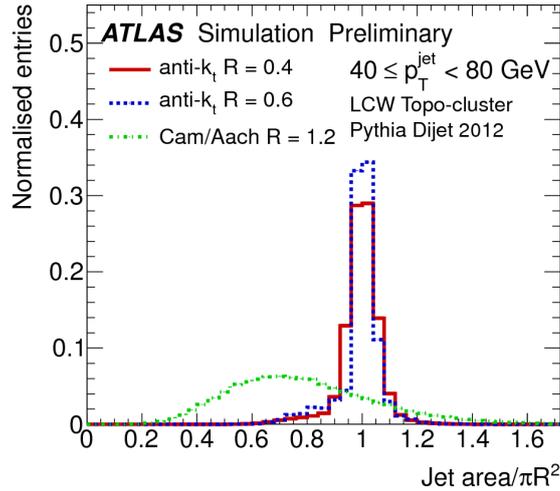


Figure 2.3: Normalized jet area distribution for Cambridge-Aachen and Anti- $k_T$  algorithm, from [17].

proton not involved in the hard process as well as final and initial-state radiation. In contrast to the MB the UE is not independent of the hard scatter part.

Protons in hadron colliders like the LHC are produced in successive bunches containing several protons. Hence, given a hard scattering process there will be further single MB events superimposed. This situation is denoted as in-time pile-up. As the speed of response of the subdetectors is normally longer than the bunch crossing rate also previous and following bunch crossings will affect the hard scattering process under investigation. The influence of other bunch crossings leads to so-called out-of-time pile-up.

## 2.3 Top quark physics

The top quark is the heaviest known quark with a mass at the electroweak scale measured at the Tevatron to be

$$m_t = 173.07 \pm 0.52(\text{stat}) \pm 0.72(\text{syst}) [11, 19] \quad (2.3.1)$$

Its relatively high mass requires large center of mass energies to produce a  $t\bar{t}$  pair at rest. The momentum fraction  $x_i$  of the partons in the hard process must be larger than

$$\sqrt{x_i x_j} \geq \frac{2m_t}{\sqrt{s}} \quad (2.3.2)$$

Assuming  $x = x_i \approx x_j$  this results in  $x \approx 0.04$  at  $\sqrt{s} = 8$  TeV. Therefore at the LHC which is a proton-proton collider top quark pairs are predominantly produced in gluon fusion with a fraction of around 80% (90%) at  $\sqrt{s} = 7(14)$  TeV [20].

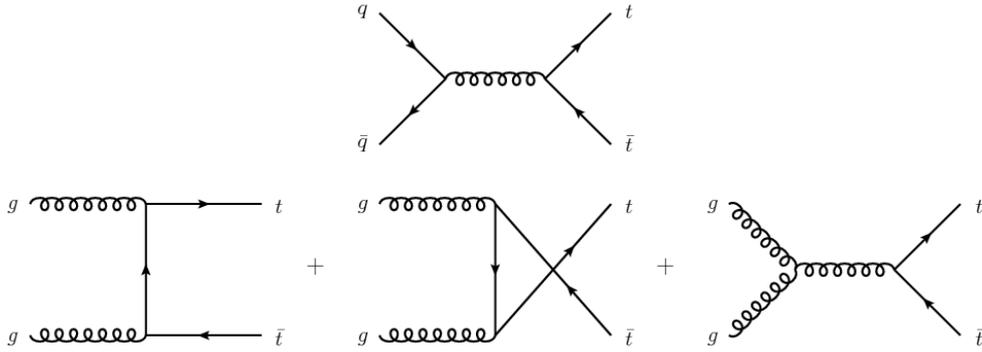


Figure 2.4: Leading order Feynman Diagrams for  $t\bar{t}$  production, from [21].

Feynman diagrams describing  $t\bar{t}$  production at leading order in QCD are shown in Figure 2.4. In 2012 around five million top quark pair events were produced by the LHC. Even though the top quark is produced in large number it has to compete with many other processes at a pp collider as illustrated in Figure 2.2.

The large width of the top quark  $\Gamma \approx 2 \text{ GeV}$  [11, 19] allows that it decays before hadronization which is characterized by the scale  $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}$ . This unique property distinguishes the top quark from other quarks. By studying its decay products it is possible to infer the properties of the top quark. Due to the CKM-matrix element  $V_{tb}$  being close to unity it nearly uniquely decays into a  $W$  boson and a bottom quark. The  $W$  boson can decay leptonically or hadronically resulting in three different decay channels for a top quark pair. At leading order 2/3 of the  $W$  decays are hadronic. At next-to-leading order the branching ratio for one flavour  $\text{BR}(W \rightarrow l\nu) = 0.108$  is used. Hence, the “all-jets”, “allhadronic” or “fully hadronic” channel has the largest branching fraction for top quark pairs with approximately 46%. The single lepton channel has a BR of 44% whereas the dileptonic channel has a BR of 11%.

The top quark also provides the largest Yukawa coupling to the Higgs boson. This large coupling to the Higgs boson makes the top quark an important tool for studying electroweak symmetry breaking. Top quark pairs can be a large background or even signal in physics beyond the standard model. Therefore, the understanding and reconstruction of the top quark plays a crucial role in corresponding searches

## 2.4 Beyond Standard Model Physics

### 2.4.1 Supersymmetry

Supersymmetry (SUSY) is a widely considered model for physics beyond the standard model. It is a spacetime symmetry relating fermionic and bosonic fields. There exists a plethora of models based on supersymmetry [22]. The most experimental investigated one is the minimal supersymmetric standard model (MSSM). It is characterized by a reduced number of fields and couplings. As a result of su-

persymmetry each standard model field has a supersymmetric partner that has the same quantum numbers except for spin. The so called “superpartners” have a spin differing by  $1/2$ . The spin-0 fields are complex scalars. The superpartners of the quarks (squarks) come in a left and right handed version corresponding to the SM quark partner. This classification is not their chirality but indicates whether they are singlets or doublets under  $SU(2)_L$  i.e. whether they couple to the  $W$  boson.

The main theoretical motivations for supersymmetry are that it provides a solution to the hierarchy problem and the unification of forces at higher scales. A problem of theories involving scalars like the SM with the Higgs field is that loop corrections to its mass involve quadratic divergences. The observation of a light Higgs mass would require an extreme fine-tuning. In contrast supersymmetry prevents divergences that are worse than logarithmic as contributions from the superpartners cancel the quadratic quantum corrections. In addition, supersymmetry with its enlarged field content causes the coupling constants of the electromagnetic, weak and strong interaction to be unified at an energy scale of around  $10^{16}$  GeV. An unbroken supersymmetry model would predict easy detectable particles with low mass. As they are not observed it must be a broken symmetry.

Furthermore, supersymmetry is generally equipped with a discrete  $C_2$  symmetry called R-parity. It ensures that supersymmetric particles can only be produced in pairs. It is defined as a multiplicative quantity  $R$  depending on the spin  $S$ , baryon number  $B$  and lepton number  $L$  of a particle.

$$R = (-1)^{2S+3B+L} \tag{2.4.1}$$

It is  $R = -1$  for supersymmetric particles and  $R = 1$  for SM particles. Hence, the lightest supersymmetric particle (LSP) cannot further decay. It is massive, neutral and only interacts via the weak force. Therefore it could provide a candidate for dark matter. As a consequence of the measured relic density of the universe the mass of this weakly interacting massive particle (WIMP) should be around the electroweak scale. R-parity also affects the form of signatures needed to search for in a collider. Such a signature is illustrated in Figure 2.5 where a stop pair is produced in a proton-proton (pp) collision. The stop can then decay to a top quark and the lightest neutralino  $\chi_1^0$  which is stable. The  $\chi_1^0$  could escape the detector and produce large missing transverse energy.

## 2.4.2 Heavy resonances

Various extensions of the gauge symmetry of the standard model like e.g. left-right symmetric models predict new gauge bosons like  $Z'$  boson. As the name indicates the  $Z'$  boson is characterized by its similarity to the SM  $Z$  boson. Namely this is its property to mediate neutral current interactions. In this thesis first of all the fact that it couples to a  $t\bar{t}$  pair is exploited. In many cases the  $Z'$  models under consideration also predict relatively small resonances of the order  $\Gamma_{Z'}/m_{Z'} = \mathcal{O}(1\%)$ . The mass of the  $Z'$  is generally larger than the SM  $Z$  boson

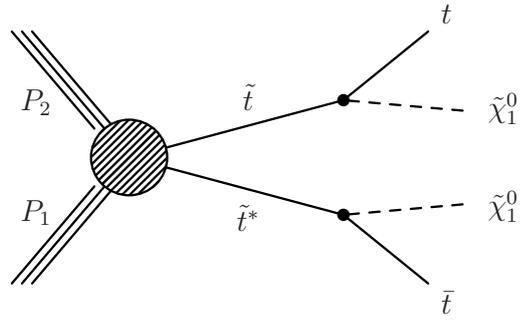


Figure 2.5: Diagram for the production of scalar top partner pair and decay to top quark and lightest neutralino, from [23].

mass resulting in the production of relatively boosted top quarks.

# 3 Experimental setup

## 3.1 LHC

The Large Hadron Collider [24] (LHC) is a particle accelerator for high energy physics searches. It is located at the European Organization for Nuclear Research (CERN) near Geneva in the former tunnel of the Large Electron–Positron Collider (LEP). The LHC is a circular collider designed to collide two counter-rotating beams of either protons or heavy ions. It is developed to reach a center of mass energy of 14 TeV for proton beams and a luminosity of  $10^{34}\text{cm}^{-2}\text{s}^{-1}$ . It achieves the highest collision energies and luminosities compared to other accelerators. The LHC proved its potential by the discovery of the Higgs boson [1, 2] in 2012.

The tunnel has a circumference of approximately 27 km and is situated about 100 m underground. Around the accelerator ring experiments are installed at four interaction points as shown in Figure 3.1.

The goals of the multi-purpose detectors ATLAS (A Toroidal LHC ApparatuS) [25] and CMS (Compact Muon Solenoid) [26] are measurements in the Higgs sector and search for new heavy particles but equally offering possibilities to perform high-accuracy measurements of known objects. The flavor sector of the Standard Model is targeted by LHCb (Large Hadron Collider beauty) [27] in order to investigate CP violation and to perform tests of the Standard Model. The geometry of LHCb allows measurements in the forward direction. ALICE (A Large Ion Collider Experiment) [28] is optimized to study heavy-ion collisions. It is designed to produce and measure the quark gluon plasma [29].

The CERN accelerator chain provides the proton beams injected at an energy of 450 GeV into the LHC beam pipe. The whole accelerator complex is shown in Figure 3.2. The protons are produced by ionizing hydrogen gas. In a first step they are accelerated by a linear collider (LINAC2) up to 50 MeV. Next the Proton Synchrotron Booster (PSB) which is composed of four superimposed synchrotron rings accelerates the protons from the LINAC2 up to 1.4 GeV before injecting them into the Proton Synchrotron (PS). The purpose of the PSB is to enhance the number of protons which can be accepted by the PS. The PS is a synchrotron with a circumference of 628 m and room-temperature magnets bending the beam. The protons are accelerated to 25 GeV in the PS. The Super Proton Synchrotron (SPS) delivers the proton beams for the LHC. It has a circumference of about 7 km again with room-temperature magnets.

The 450 GeV protons are then accelerated in the LHC during a “fill” to the

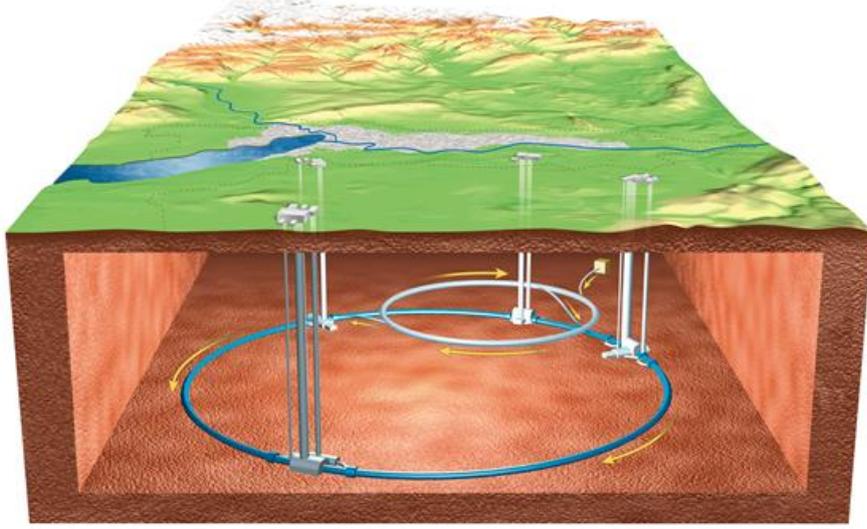


Figure 3.1: Overview of the LHC accelerator ring and its interaction points, from [30].

center of mass energy of 7 TeV in 2011 and 8 TeV in 2012 data taking period. The beam is bent on a circular path by 1232 superconducting dipole magnets cooled down to 1.8° K. The provided magnetic field has a strength of 8.33 T. The rate of pp interactions is maximized by focusing the beam in the straight sections of the accelerator ring with 502 quadrupole magnets. The acceleration inside the LHC is achieved by radio frequency cavities while the protons circulate in two separate vacuum tubes.

The proton beams are separated in 2808 bunches with about  $1.15 \times 10^{11}$  protons. Collisions at the interaction points can take place up to every 25 ns. The operated spacing in 2011 and 2012 was 50 ns. The number of produced events  $N_{\text{events}}$  in a process with cross section  $\sigma$  is depending on the instantaneous luminosity  $L$  provided by the collider in the following way

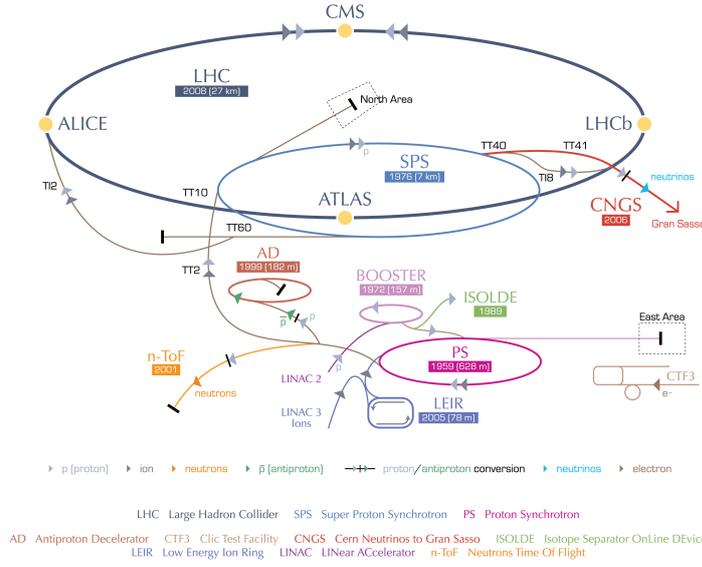
$$\frac{dN}{dt} = \sigma \times L \quad (3.1.1)$$

The maximal instantaneous luminosity of the LHC is  $10^{34} \text{cm}^{-2} \text{s}^{-1}$ . The instantaneous luminosity is determined by the crossing rate of bunches ( $f$ ), the number of protons in each bunch ( $n_1, n_2$ ) and the size ( $\sigma_1, \sigma_2$ ) of the beam.

$$L = f \frac{n_1 n_2}{4\pi\sigma_1\sigma_2} \quad (3.1.2)$$

The integrated luminosity  $\mathcal{L}$  is the integral over time of the delivered instantaneous

## CERN's accelerator complex



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

Figure 3.2: Overview of the LHC accelerator complex, from [31].

luminosity.

$$\mathcal{L} = \int L dt \tag{3.1.3}$$

The integrated luminosity quantifies the productivity of the collider. The integrated luminosity recorded at the LHC as well as the maximal delivered luminosity in the years 2011 and 2012 are shown in Figure 3.3. A total integrated luminosity of  $5.25\text{fb}^{-1}$  was recorded by ATLAS in the 2011 run and  $21.7\text{fb}^{-1}$  in the 2012 run.

The high luminosities reached especially at the end of the 2012 run (see Figure 3.3) also raise the impact of pile-up. It is determined by the average number of inelastic pp interactions per bunch crossing  $\mu$ . The variable  $\mu$  is calculated from the inelastic cross section  $\sigma_{\text{inel}}$ , the LHC revolution frequency  $f_r$ , the number of colliding bunches  $n_{\text{bunch}}$  and  $\mathcal{L}$ .

$$\mu = \sigma_{\text{inel}} \frac{\mathcal{L}}{n_{\text{bunch}} f_r} \tag{3.1.4}$$

The average of  $\mu$  in the respective run was significantly higher in 2012 with  $\langle \mu \rangle = 20.7$  compared to  $\langle \mu \rangle = 9.1$  in 2011 as shown in Figure 3.4.

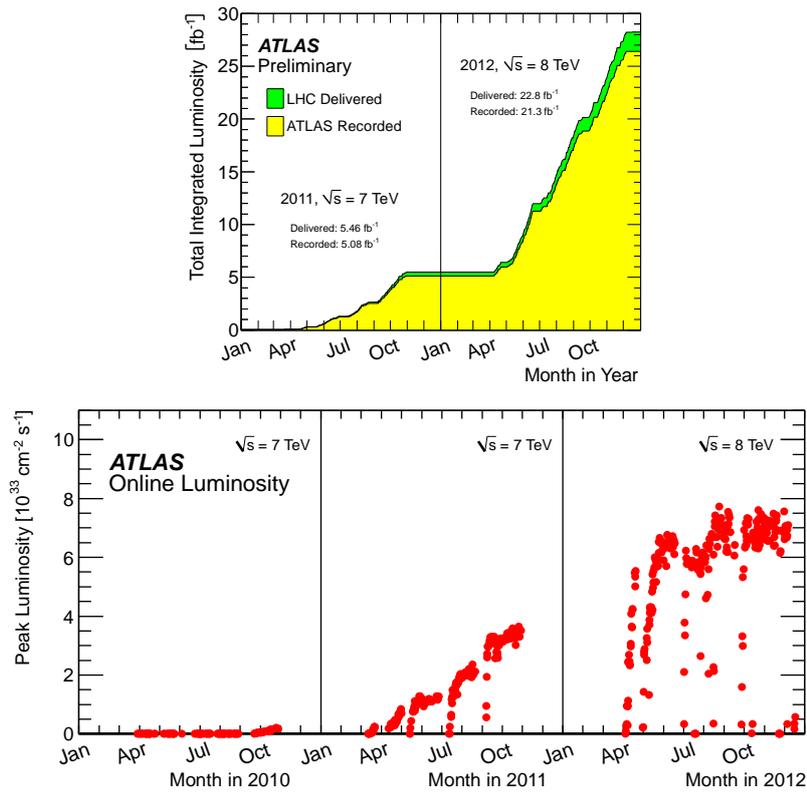


Figure 3.3: Total integrated luminosity in 2011 and 2012 and peak luminosity, from [32].

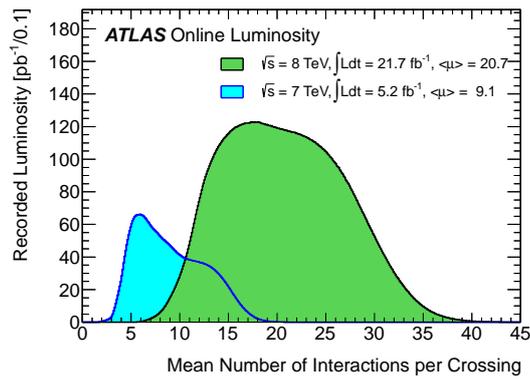


Figure 3.4: Luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 ATLAS data, from [32].

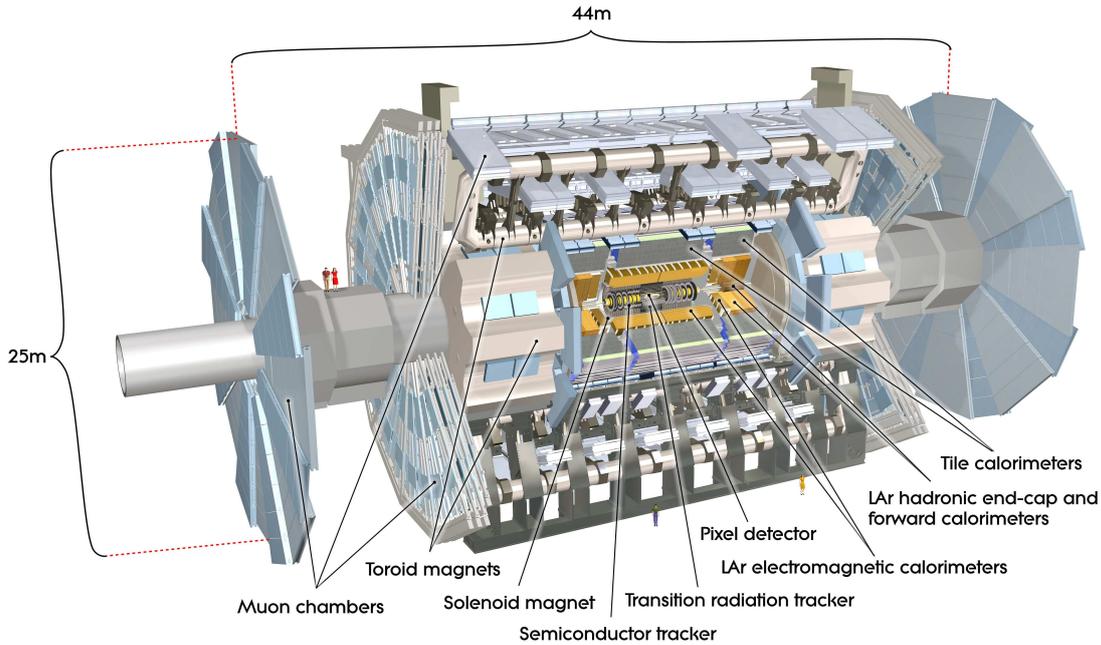


Figure 3.5: Overview of the ATLAS detector, from [33].

## 3.2 ATLAS Detector

In this section a brief overview of the ATLAS subdetectors [33] is given. The subsystems constituting the ATLAS detector are shown in Figure 3.5.

The coordinate system of ATLAS is a right-handed coordinate system with its origin corresponding to the nominal interaction point in the center of the detector. The beam direction defines the  $z$ -axis, the  $x$ -axis points towards the center of the LHC ring and the  $y$ -axis points upwards. Due to the geometry cylindrical coordinates are used, where  $\phi$  is the azimuth angle around the beam pipe in the  $x$ - $y$  plane. The pseudorapidity is defined as  $\eta = -\ln \tan \theta/2$ , with  $\theta$  being the polar angle from the beam axis. Conventional “transverse” quantities like transverse momentum are defined in the  $x$ - $y$  plane.

The general design of ATLAS is that of a multi-purpose detector as illustrated in the wedge profile in Figure 3.6. It is optimized to provide a large possible selection of physics objects. This large selection is required for redundant measurements and internal-cross-checks.

The superconducting magnetic system has a central solenoid (CS) component around the inner detector (ID) and two end-cap toroids (ECT) outside the calorimeters.

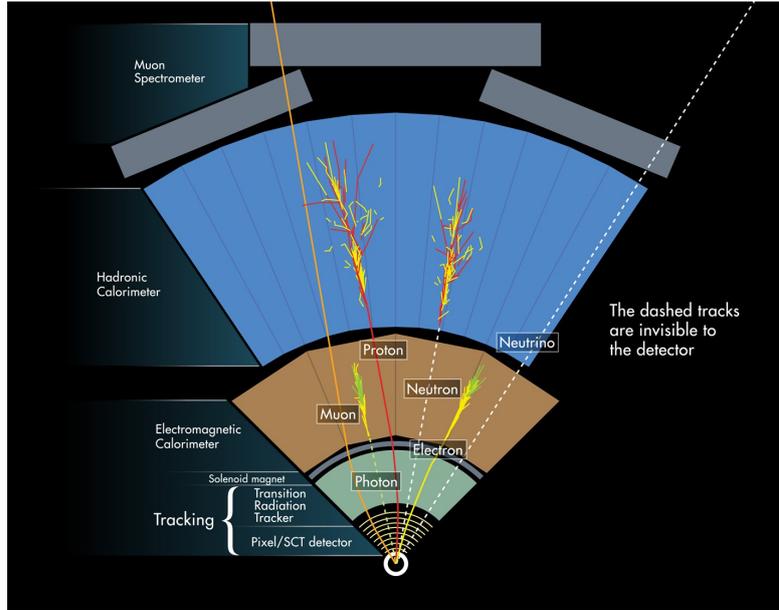


Figure 3.6: Schematic overview of the ATLAS detector, from [34].

### 3.2.1 Inner Detector

The inner detector layout is shown in Figure 3.7. It consists of three individual subsystems situated in a 2 T magnetic field of the CS. The components are, with increasing distance from the interaction point, the pixel detector (PD), the silicon strip detector (SCT) and the transition radiation tracker (TRT). The subsystems cover the pseudorapidity range  $|\eta| < 2.5$ . The ID allows to reconstruct charged particle tracks, as well as primary and secondary vertices.

It has to handle a high track density with up to about 1000 particles from the interaction point every 25 ns. This task is achieved by a high granularity of the semiconductor tracking detectors. The PD consists of three cylinders of silicon sensors at radii  $r = 50.5$  mm,  $r = 88.5$  mm and 122.5 mm in the barrel region and end-cap disks at  $|z| = 495$  mm,  $|z| = 580$  mm and  $|z| = 650$  mm. The spatial resolution of the PD is  $10 \mu\text{m}$  in  $r - \phi$  and  $115 \mu\text{m}$  in the  $z$ -direction. The SCT has a design similar to the PD. The lower track density due to the larger area allows to use strips instead of pixels. The spatial resolution is  $16 \mu\text{m}$  in  $r - \phi$  and  $580 \mu\text{m}$  in the  $z$ -direction. Straw tubes filled with a Xenon based gas mixture form the TRT. The TRT covers the pseudorapidity range  $|\eta| < 2$  and provides only  $r - \phi$  information. It is an important tool for the identification of electron candidates.

### 3.2.2 The Calorimeters

An overview of the ATLAS calorimeters is presented in Figure 3.8. The hadronic calorimeter surrounds the electromagnetic calorimeter (EM). Both are sampling calorimeters. The EM uses a liquid-argon (LAR) technique whereas the hadronic

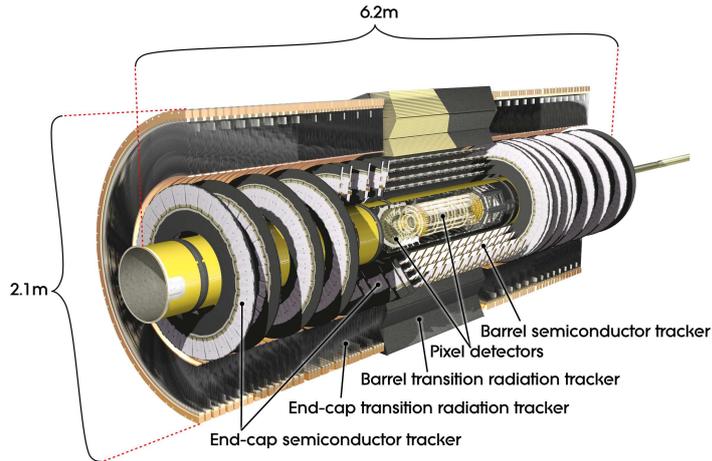


Figure 3.7: Overview of the ATLAS inner detector, from [33].

uses a mixture of LAr and tile calorimeters. The absorber material in the LAr is lead. The EM is composed of the LAr barrel (EMB) and endcap (EMC) calorimeters covering  $|\eta| < 3.2$ . The tile scintillator hadronic barrel calorimeter covers  $|\eta| < 1.7$ . The LAr hadronic endcap calorimeter (HEC) covers  $1.5 < |\eta| < 3.2$ , and the LAr forward calorimeter (FCAL) covers  $3.1 < |\eta| < 4.9$

The highest granularity is reached in the barrel (EMB) region of the LAr EM calorimeter covering  $|\eta| < 1.475$ . It uses the same vacuum vessel as the CS to reduce the present inactive material. The accordion geometry of the calorimeters provides full  $\phi$  symmetry without azimuthal cracks. In the longitudinal direction the EM is divided in three segments (strip, middle, back). The  $\Delta\eta \times \Delta\phi$  granularity is different in all barrel segments with about  $0.003 \times 0.1$  in the strip,  $0.025 \times 0.025$  in the middle and  $0.05 \times 0.025$  in the back. The total thickness of the EM is  $> 22X_0$ . In order to have a better measurement of particles starting to shower before the EM a presampler (PS) in the region  $|\eta| < 1.8$  is installed.

The tile calorimeter uses steel as absorber and plastic scintillator tiles as active material. At the end of the tiles wavelength shifting fibres forward the signals to photomultipliers. It is segmented into three layers. The granularity is coarser than the EM. The azimuthal granularity is  $\Delta\phi = 0.1$ . The third layer has a pseudorapidity granularity of  $\Delta\eta = 0.2$  whereas the first two have  $\Delta\eta = 0.1$ .

The HEC use the same LAr technology as the EM. It consists of two independent wheels built out of copper plates. The HEC uses the same cryostats as the EM end-cap. This end-cap cryostat also contains the copper/tungsten LAr forward calorimeter (FCal). The copper supports the measurement of EM showers whereas the tungsten is used for the hadronic showers.

The  $\eta$  dependence of the overall granularity from the detector geometry is illustrated in Figure 3.9. The impact of the topological clusters will be explained

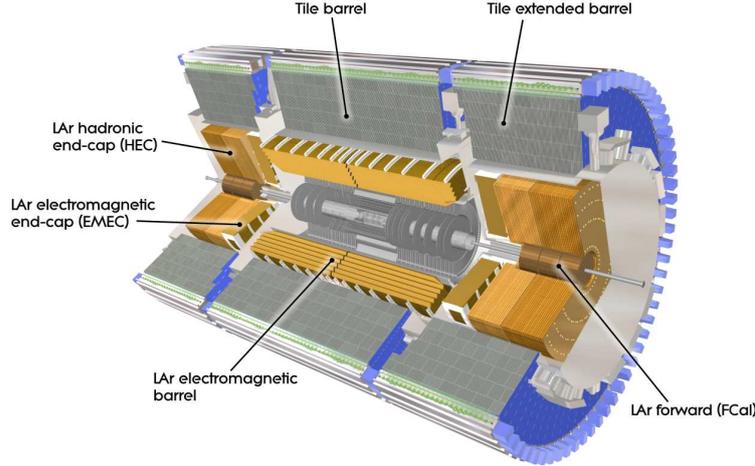


Figure 3.8: Overview of the ATLAS calorimeter system, from [33].

in more detail in Chapter 3.4.4. The highest granularity is reached in the central region  $|\eta| < 2.5$ . For this region the granularity is also roughly constant.

### 3.2.3 The Muon System

The layout of the muon spectrometer is presented in Figure 3.10. It is strongly related to the design of the toroid magnets. Two magnets are at the endcaps with a magnetic field of approximately 1 T and one in the barrel region with approximately 0.5 T. Hence, the magnetic deflection of muons is achieved by a magnetic field which is mostly orthogonal to the particle trajectories. The muon spectrometer provides independent tracking and trigger capabilities. The muon momentum is determined by means of Monitored Drift Tube (MDT) chambers and Cathode Strip Chambers (CSC) in the endcaps.

### 3.2.4 Trigger System

The ATLAS trigger and data-acquisition system is organized in three stages called level-1 (L1), level-2 (L2) and event filter (EF). These three trigger levels are sketched in Figure 3.11. The L1 uses ASICs and FPGAs and is fully hardware based. The L2 and EF form the high level trigger (HLT) and are software based. The main task of the trigger system is to reduce the high interaction rate to a storable amount while keeping enough efficiency for rare new physics processes.

The L1 uses coarse granularity information from all calorimeters for jets, missing transverse energy and total transverse energy. It is directly feed with the up to 40 MHz bunch crossing rate and is designed to achieve a reduction down to 75 kHz. The design of the L1 requires that the decision whether the event passed must be made in less than  $2.5\mu\text{s}$ . The L1 defines Regions-of-interest (RoI) which are used

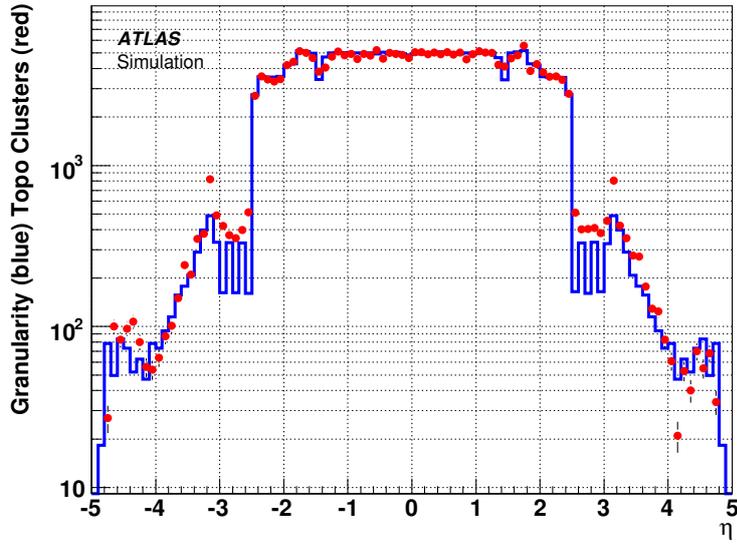


Figure 3.9: The geometric granularity (number of cells per  $\Delta\eta = 0.1$ ) as a function of pseudorapidity (blue) and corresponding number of topological clusters (red) calculated from simulated noise-only events, from [17].

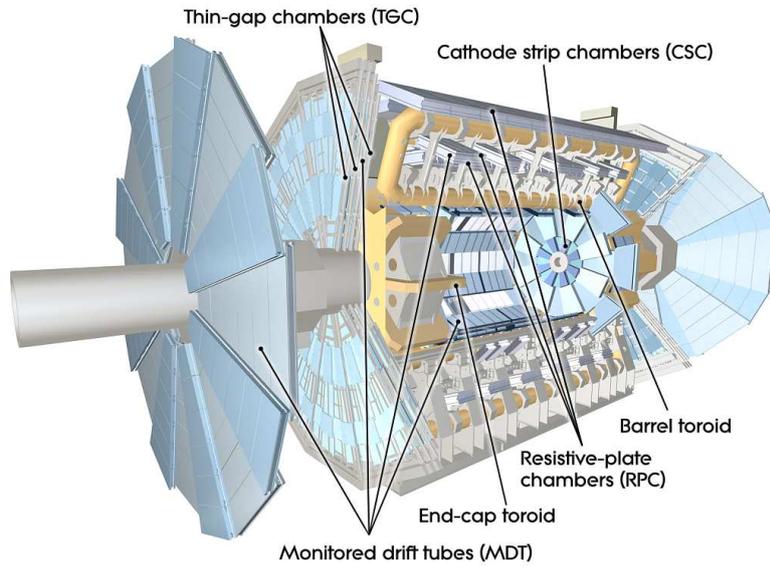


Figure 3.10: Overview of the ATLAS muon system, from [33].

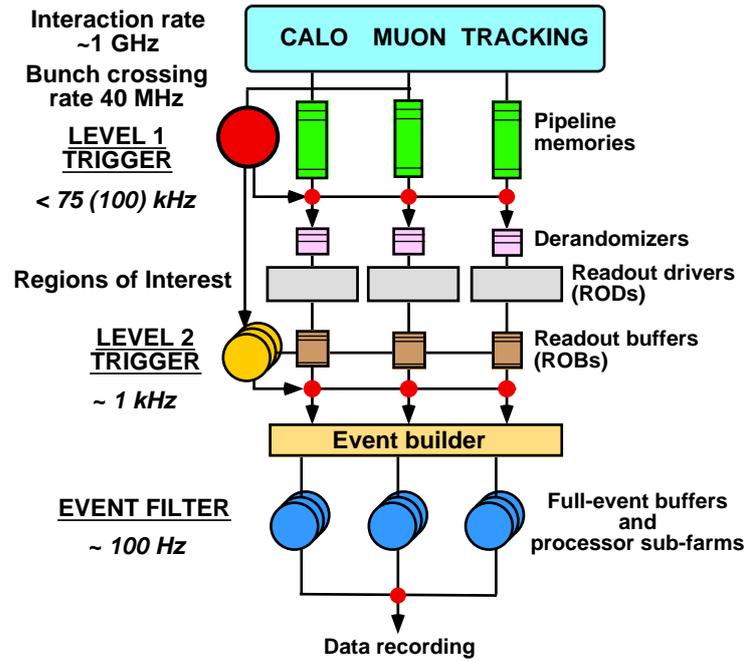


Figure 3.11: Schematic overview of the Trigger/DAQ system and the corresponding design rates, from [35].

to seed the HLT. These RoIs are the possible candidates for trigger objects.

The L2 trigger investigates these candidates in more detail with full granularity. The L2 can use additional information from the ID like tracks. The latency of L2 is approximately 40 ms with a designed reduction of the event rate to approximately 1 kHz.

The EF builds up on the full event information. It reduces the rate down to approximately 100 Hz which is required to store the event on disk. The EF uses offline like reconstruction algorithms but with a looser selection. The average event processing time for the EF is about 4 s.

### 3.3 Data and Monte-Carlo samples

The data used for the comparison to simulated events in Chapter 6 is recorded in the ATLAS detector in pp collisions in the 2012 run at a center of mass energy of 8 TeV. The total recorded good quality data is  $20.3 \text{ fb}^{-1}$ . In Chapter 6 only a subset corresponding to  $2.3 \text{ fb}^{-1}$  is used.

A general introduction into Monte-Carlo generators used in LHC analyses can be found in [36] or [37]. SM  $t\bar{t}$  events are simulated with the MC@NLO generator [38, 39] with CT10 parton density functions (PDF) [40]. Final-state parton showers are simulated and hadronized using the HERWIG [41] program interfaced to the JIMMY underlying event model [42] with the ATLAS AUET2 tune [43, 44].

The  $t\bar{t}$  production cross section for  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 8$  TeV is  $\sigma_{t\bar{t}} = 253_{-15}^{+13}$  pb for a top quark mass of 172.5 GeV. It is calculated at next-to-next-to leading order (NNLO) in QCD including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms with top++2.0 [45, 46, 47, 48, 49, 50]. The PDF and  $\alpha_S$  uncertainties were calculated using the PDF4LHC prescription [51] with the MSTW2008 68% CL NNLO [52, 53], CT10 NNLO [40, 54] and NNPDF2.3 5f FFN [55] PDF sets, added in quadrature to the scale uncertainty. The NNLO+NNLL value is about 3% larger than the exact NNLO prediction, as implemented in Hathor 1.5[56].

The performance of the top quark reconstruction is validated with relatively high- $p_T$  top quarks from  $pp \rightarrow Z' \rightarrow t\bar{t}$  samples which are produced using PYTHIA [57] with the MSTW2008 LO PDF [52] and the ATLAS AU2 tune [58]. The rejection of background by the top tagging algorithms is studied using a dijet sample generated with PYTHIA with the CT10 LO PDF and the AU2 tune.

The pile-up conditions in the 2012 data are considered in all Monte-Carlo samples. The in-time pile-up as well as the out-of-time pile-up is modeled by overlaying one signal collision per bunch crossing with hits from minimum-bias events that are produced with Pythia using the A2M tune and the MSTW2008LO PDF. The number of overlaid events depends on  $\mu$ . It is modeled using a Poisson distribution with mean  $\mu$ . The number of simulated bunch crossings is increased for detector subsystems that are more effected by out-of-time pile-up. The  $\mu$  distribution for data is measured by the ATLAS luminosity detectors [59]. The events of the Monte-Carlo simulation are reweighted to describe the measured  $\mu$  distribution.

The response of the ATLAS detector to the Monte-Carlo particles is modeled by the GEANT4 toolkit [60, 37].

## 3.4 Reconstruction and selection of physics objects and events

This chapter summarizes the definition of physic objects reconstructed from detector information. They are used to select events and as an input to the bucket algorithm.

### 3.4.1 Electrons

Electron candidates are seeded from energy clusters in the EM that are matched to an ID track. Electrons are required to pass the identification corresponding to the “tight” selection as described in Reference [61]. These criteria provide a set of selection cuts on calorimeter and tracking variables, as well as combinations of these. It ensures a separation between isolated electrons and jets. Only central electrons with  $|\eta| < 2.47$  are considered. The range  $1.37 < |\eta| < 1.52$  is excluded

taking into account the crack between barrel and end-cap calorimeters. The  $|\eta|$ -coordinate of the cluster with respect to the primary vertex position is used. The electron candidates should satisfy  $E_T = E_{\text{cl}}/\cosh \eta_{\text{track}} > 25$  GeV. The energy is calculated from the cluster and the direction from the associated track. The criteria for rejection of electrons overlapping with a jet are described in Chapter 3.4.7. The electrons also have to satisfy object quality requirements. Isolation is imposed on the electron by requiring that the  $p_T$  sum of all tracks in a cone of variable size  $\Delta R = 10 \text{ GeV}/p_T^{\text{electron}}$  around the candidate's direction must be less than 5% of its  $E_T$ . The longitudinal impact parameter  $z_0$  is the  $z$  coordinate of the electron track with respect to the primary vertex (see Sec. 3.4.3). The impact parameter  $z_0$  is required to be less than 2 mm.

### 3.4.2 Muons

Muons are identified by an algorithm combining information from the ID and the muon subsystems. The performance of the muon selection criteria and tracking quality cuts is described in Reference [62]. In addition, the muon longitudinal impact parameter  $z_0$  relative to the primary vertex is required to be less than 2 mm. Isolation is imposed on the muon by requiring that the  $p_T$  sum of all tracks in a cone of variable size  $\Delta R = 10 \text{ GeV}/p_T^\mu$  around the candidate's direction must be less than 5% of the muon  $p_T$ . Acceptance requirements,  $p_T > 25$  GeV and  $\eta < 2.5$ , are applied to all muons.

### 3.4.3 Tracks and vertices

In the inner detector (ID) tracks are reconstructed within the full acceptance  $|\eta| < 2.5$ . The reconstruction uses “hits” in the ID sub-detectors [63]. In a first step track candidates are build from 3-point seeds in the silicon detectors (Pixel and SCT). Outlying clusters, ambiguities and fake tracks are removed before the track is extrapolated to the TRT. This is the inside-out algorithm which aims to reconstruct primary charged particles. These are the particles with a mean lifetime greater than  $3 \times 10^{-11}$  s produced directly in a pp interaction or in subsequent decays or interactions of particles with a mean lifetime shorter than  $3 \times 10^{-11}$  s. Tracks from the inside-out algorithm have to fulfill  $p_T \geq 400$  MeV .

The position of the primary vertices is determined by an iterative vertex finding algorithm [64]. The input objects are tracks. The z-position of the reconstructed tracks at the beamline is used as a seed. An iterative  $\chi^2$  fit associates a weight to each track classifying its compatibility with the fitted vertex.

Tracks with a distance of more than  $7\sigma$  from the vertex are used to seed a new vertex and the procedure is repeated until no additional vertices can be found. The hard-scatter vertex is the one with largest  $\sum p_T^2$  of the tracks associated to it.

The in-time pile-up in an event can be quantified by the number of primary vertices with at least two tracks  $N_{\text{PV}}$ . It is directly correlated to the mean number

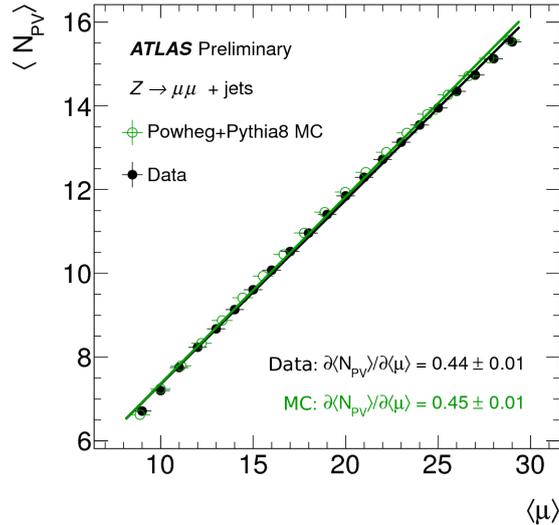


Figure 3.12: Correlation between mean number of primary vertices  $\langle N_{PV} \rangle$  and mean number of interactions per bunch crossing  $\langle \mu \rangle$ , from [17].

of interactions per bunch crossing  $\langle \mu \rangle$  which captures both in-time and out-of-time pile-up. The correlation is shown in Figure 3.12.

### 3.4.4 Topological clusters

Before reconstructing jets their input objects, topological clusters, have to be constructed [65]. In order to have noise suppression the signal inside a cell is required to pass a threshold of  $4\sigma$  to be considered as a seed for a cluster. Here  $\sigma$  is defined as the sum in quadrature of the standard deviations of the distributions of electronic and estimated pile-up noise. All cells with signal greater  $2\sigma$  around the seed in three dimensions are added to the cluster. A single layer of cells directly surrounding the  $2\sigma$  region is also taken into account. The case where topological clusters are merged is considered by a cluster splitting algorithm [65].

The local cluster weighting (LCW) algorithm [18, 66] calibrates the topological cluster energy. The LCW method distinguishes whether a topological cluster has electromagnetic or hadronic origin and applies appropriate energy corrections. It also includes information about non-instrumented regions of the calorimeters and energy lost due to noise suppression.

### 3.4.5 Jets

In data and Monte-Carlo simulations various types of jets are build. These different jets are sketched in Figure 3.13. In Monte-Carlo simulations it is possible to use the Monte-Carlo particles to construct truth jets. For this purpose stable and interacting final state particles from the hard-scatter interaction are used. Interacting particles are the ones depositing most of their energy in the calorimeters.

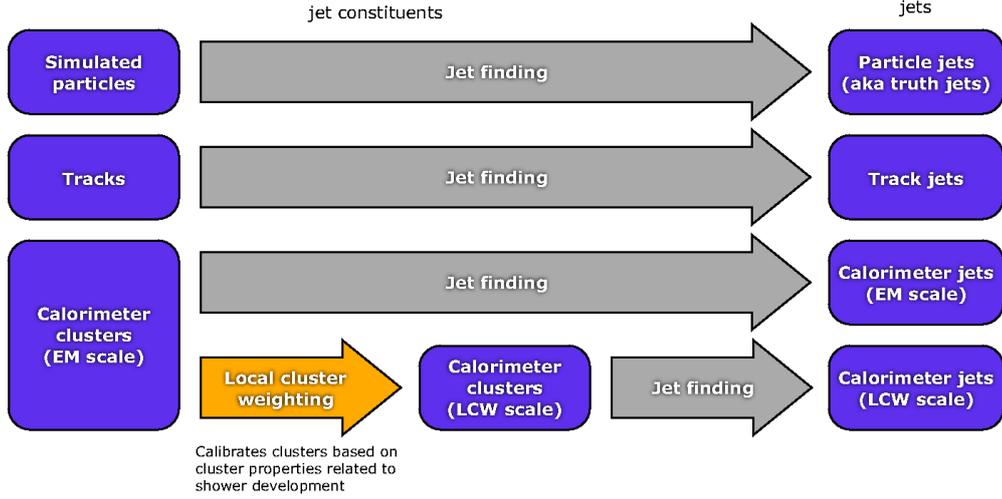


Figure 3.13: The ATLAS jet reconstruction scheme. The jet four momentum is defined by the four momentum sum of its constituents, from [68].

Stable particles are required to have a decay length  $c\tau > 1$  cm. Calorimeter jets are reconstructed from topological clusters using the Anti- $k_T$  jet algorithm [67] with a distance parameter of  $R = 0.4$  and full four-momentum recombination using the FastJet software [15, 16].

The jets are calibrated in several stages. The jet response depends on the pile-up conditions. This pile-up dependence is accounted for using the jet area subtraction method [17]. In a first step the pile-up subtraction is supposed to remove the influence of pile-up on the jet transverse momentum. There still remains a calorimeter response difference between single particles and jets. This difference is considered in the next step by the jet-energy-scale (JES) response correction. The JES correction aims at improving the kinematical agreement between simulated truth jets and the reconstructed calorimeter jets. All jets are required to have  $p_T > 25$  GeV and  $|\eta| < 2.5$ .

The jet vertex fraction (JVF) is a function of the jets  $\text{jet}_i$  and primary vertices  $\text{PV}_j$  in an event [17]. For the calculation of JVF only tracks with  $p_T > 500$  MeV are used.

$$\text{JVF}(\text{jet}_i, \text{PV}_j) = \frac{\sum_k p_T(\text{track}_k^{\text{jet}_i}, \text{PV}_j)}{\sum_n \sum_l p_T(\text{track}_l^{\text{jet}_i}, \text{PV}_n)} \quad (3.4.1)$$

where  $k$  runs over all tracks originating from  $\text{PV}_j$  matched to  $\text{jet}_i$ ,  $l$  runs over all tracks originating from  $\text{PV}_n$  matched to  $\text{jet}_i$  and  $n$  runs over all primary vertices. For the selection of jets the JVF with respect to the hard-scatter vertex is considered. The jet vertex fraction is required to be  $|\text{JVF}| > 0.5$  only for jets with  $p_T \leq 50$  GeV and  $|\eta| < 2.4$ . These selected jets are the ones considered when resolving overlapping objects. Hardware problems in the calorimeter, cosmic-ray induced

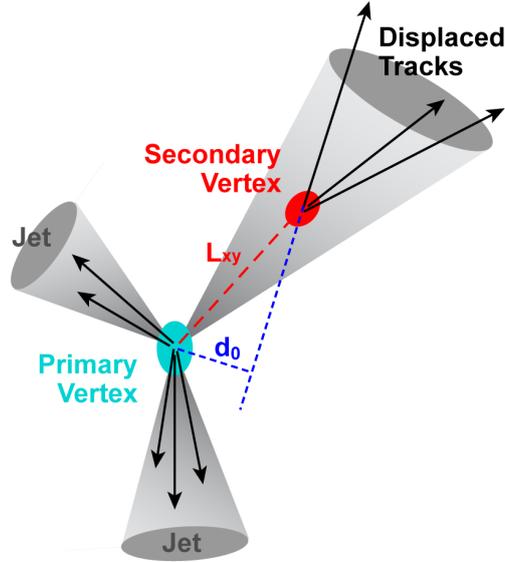


Figure 3.14: Illustration of the decay topology of a B-hadron, from [69]

showers and LHC beam-gas interactions can cause “bad” jets. These events are rejected by requiring the event to fulfill jet quality criteria.

### 3.4.6 Bottom-jets

As the algorithm presented in this thesis relies fundamentally on  $b$ -tagging the ability of the ATLAS Detector to correctly identify jets coming from  $b$ -quarks is a crucial ingredient.

The relatively long lifetime of B-hadrons causes a significant displacement between the B-hadron’s decay vertex and the primary vertex as illustrated in Figure 3.14. Furthermore, the B-hadron decay in the secondary vertex gives rise to displaced tracks with measurable impact parameters defined as the minimum distance of the linearized track from the primary vertex. These inner detector quantities are the objects used in  $b$ -tagging algorithms.

Jets are identified as originating from the decay of a  $b$ -quark if they pass the MV1 jet tagger. The MV1  $b$ -tagging algorithm is based on a neural network using the output weights of the JetFitter+IP3D, IP3D and SV1 algorithms as input. The JetFitter exploits the decay topology of the  $b$ -hadrons inside the jet. The SV1 algorithm relies on secondary vertex information whereas the IP3D algorithm is impact parameter based. The algorithms are described in more detail in Reference [70].

If not stated otherwise the working point with a 70%  $b$ -tagging efficiency is used. The relation between  $b$ -jet efficiency and rejection is shown in Figure 3.15 based on a sample with simulated  $t\bar{t}$  events.

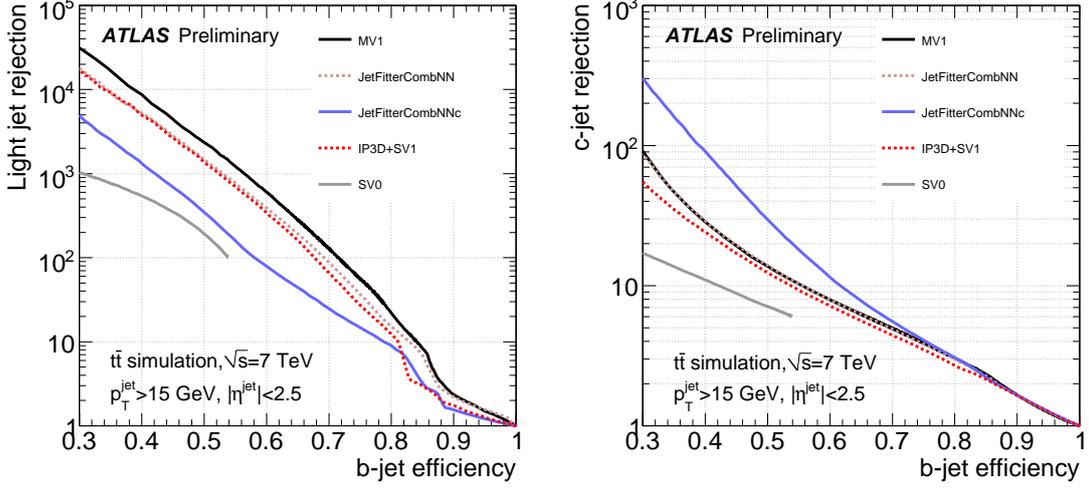


Figure 3.15:  $b$ -tag efficiencies and light-jet and  $c$ -jet rejection of the used  $b$ -tagging algorithms. The efficiencies are calculated using simulated  $t\bar{t}$  events, from [71]

### 3.4.7 Overlap removal

In order to clearly separate leptons from hadronic jets an overlap removal is performed. Overlapping objects are compared using the distance  $\Delta R$  in the  $\eta - \phi$  plane ( $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ ). The order of the overlap removal starts with the electrons. As electrons deposit energy in the calorimeter they are usually additionally reconstructed as jets. To reject such electron-jet duplicates the closest jet to an electron passing all selection criteria is rejected if the distance between both objects in the  $\eta - \phi$  plane is smaller than 0.2. Hence, if an electron and a jet are within  $\Delta R < 0.2$  the jet is removed and the object is considered an electron. Next, muons which are located in  $\Delta R < 0.4$  to a jet are removed. If a jet and an electron are located with  $0.2 \leq \Delta R \leq 0.4$  the electron is removed and the object is treated as a jet.

### 3.4.8 Event selection

Events are required to pass a basic selection cuts to be considered as top quark candidates.

- the event must not contain less than 5 jets and  $p_T^{\text{jet}} \geq 25$  GeV
- if isolated electrons or muons as described in their selection are found the event is rejected
- the event has to contain at least two  $b$ -tagged jets depending on the chosen working point. If not stated otherwise the “medium” working point corresponding to 70%  $b$ -jet efficiency is used.

### 3.4.9 Missing transverse energy

On top of the basic selection additional variables are used in e.g. the SUSY study. Stable particles which are only weakly interacting like neutrino or WIMP candidates in extensions of the standard model like SUSY will be invisible for the detector and cannot be directly reconstructed. But it is possible to get information about them by looking into the transverse plane of the detector. The initial state resulting from the pp collision has generally very small  $p_T$ . Therefore by measuring the momentum conservation in the transverse plane from reconstructed objects the missing transverse energy has to account for the difference. Hence, the measured energy is decomposed in its three-dimensional components. The transverse part has to fulfill:

$$0 = \sum \mathbf{E}_T^{\text{visible}} + \sum \mathbf{E}_T^{\text{invisible}} \quad (3.4.2)$$

As a consequence the missing transverse energy is the negative sum of all reconstructed energies. The calculation of the missing transverse energy in the ATLAS detector is object based. The basic contribution are calorimeter cells. The calorimeter energy deposits associated to hard objects according to their calibration are summed. Each term in the sum defining the global missing transverse energy corresponds to the  $p_T$  sum for each class of objects. The objects contributing in each term are given below

1. Ele: electrons
2. Jet: Jets with  $p_T > 20$  GeV
3. Muon: muons
4. CellOut: cells outside reconstructed objects

The  $x$ - and  $y$ - components are summed and used to calculate the scalar missing transverse energy  $E_T^{\text{miss}}$ :

$$E_{x,y}^{\text{miss}} = E_{x,y}^{\text{Ele}} + E_{x,y}^{\text{Jet}} + E_{x,y}^{\text{Muon}} + E_{x,y}^{\text{CellOut}} \quad (3.4.3)$$

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2} \quad (3.4.4)$$

The muons as minimally-ionizing particles can escape the calorimeter. As a consequence their momentum can not be described just by a calorimeter component. Therefore the muon term combines information from the calorimeters and the muon spectrometer.

### 3.4.10 Trigger

In this thesis triggers which rely on calorimeter information like multijet triggers or missing transverse energy triggers are used. In addition,  $b$ -jet triggers are investigated which are also based on tracking information. A more detailed description

of the L1 calorimeter trigger can be found in Reference [72]. The evolution of the trigger menu during the 2011 and 2012 run is presented in Reference [73].

- multijet trigger:  
A  $8 \times 8$  trigger-tower grid in  $\eta$  and  $\phi$  in the electromagnetic and hadronic layers is used to identify the jet RoIs. The actual jet finding is performed by a sliding window algorithm based on a  $\Delta\eta \times \Delta\phi$  granularity of  $0.2 \times 0.2$  for the search of jet clusters [72]. At L1 four jet candidates with transverse energy  $E_T > 15$  GeV are required. The coarse granularity at L1 results in poor jet position resolution. Therefore the L1 RoIs are not used as seeds for L2 jet finding. Instead, at L2 an Anti- $k_T$  ( $R = 0.4$ ) algorithm is executed with all L1 trigger towers as input. This jet finding procedure is referred to as level 2 full scan (L2FS). At the EF level, topological clusters are calibrated at the hadronic scale. These topological clusters are the input for an Anti- $k_T$  ( $R = 0.4$ ) algorithm. At least five jets with  $E_T > 55$  GeV are required at the EF.
- $b$ -jet trigger:  
Tracking information is needed to perform a  $b$ -jet selection. Therefore,  $b$ -jet triggering starts at the HLT level. Generally, the higher rejection from the  $b$ -jet trigger allows lower transverse energy thresholds at L1. To select jets likely to originate from  $b$ -quarks, events are selected using the IP3D+SV1 algorithm [70]. The IP3D+SV1 algorithm is an impact parameter based and secondary vertex based algorithm. The  $b$ -jet identification is performed on top of the calorimeter jet reconstruction as described for the multijet triggers. For the 2012 run all tracks in the event are used to reconstruct the  $z$ -position of the primary vertex [74]. Two different  $b$ -jet triggers were investigated. The first  $b$ -jet trigger requires one  $b$ -jet with  $E_T > 45$  GeV and a working point corresponding to 50%  $b$ -jet selection efficiency (medium) at EF level. In addition, four Anti- $k_T$  ( $R = 0.4$ ) jets with  $E_T > 45$  GeV calibrated at the hadronic scale must be identified. The second  $b$ -jet trigger requires two  $b$ -jets with  $E_T > 35$  GeV and a working point corresponding to 60%  $b$ -jet efficiency (loose) at EF level. In addition, three Anti- $k_T$  ( $R = 0.4$ ) jets with  $E_T > 35$  GeV calibrated at the hadronic scale must be identified.
- $E_T^{\text{miss}}$  trigger:  
The missing transverse energy trigger considered is based on the full sum of all trigger tower transverse energies at L1. Pile-up noise suppression is used in the FCal to keep the  $E_T^{\text{miss}}$  thresholds low with increasing luminosity [73]. The corresponding “loose” working point for the noise cut is considered here. The  $E_T^{\text{miss}}$  thresholds at L1 required  $E_T^{\text{miss}} > 40$  in 2012. At the EF the trigger uses the vector sum of all topological clusters calibrated at the LCW-scale and requires  $E_T^{\text{miss}} > 80$  GeV.

# 4 Description of the Bucket Algorithm

The construction of the bucket algorithm as first proposed in Reference [9] is summarized in this chapter.

The bucket algorithm targets a moderate transverse momentum regime of top quark pairs between  $p_T^{\text{top}} = 100 - 400$  GeV. As developed in Reference [9] it is designed to identify and reconstruct events containing a pair of top quarks decaying both fully hadronically.

In contrast to fatjet based reconstruction algorithms it also targets low  $p_T$  top quarks where the angular separation between the decay products of the top quarks are widely separated making it impossible to be captured by a reasonable size fatjet. The  $p_T$  dependence of the geometrical separation  $\Delta R_{bjj}$  of the decay products of SM  $t\bar{t}$  is illustrated in Figure 4. In order to tag top quarks at small  $p_T < 200$  GeV it is necessary to combine jets coming from a large area. Not only low  $p_T$  top quarks but also high  $p_T$  top quarks can have separations above the size typically used by fatjet based taggers. These large separations drastically limit the efficiency at low  $p_T$  leading to a threshold to find fatjets around 200 GeV and reduce the efficiency for moderate boosts in the range  $p_T = 200 - 400$  GeV.

Classical resolved top tagging approaches are in contrast to substructure methods making use of small and isolated jets. Therefore they are limited at higher  $p_T$  by merging of the decay product jets depending on the used jet radius parameter according to

$$p_T \simeq \frac{2m}{R} \tag{4.0.1}$$

In order to provide the input jets for the top quark reconstruction it is often necessary depending on the method to consider quite hard jets and to have at least six jets in an event. Also events where due to radiation the number of jets describing the initial top quark is larger than three are not correctly considered.

The bucket algorithm aims to complement these existing techniques by reducing the momentum threshold present in substructure methods and by loosening the selection requirements of the resolved reconstruction.

The starting point of the algorithm is therefore a moderately high multiplicity of Anti- $k_T$  ( $R = 0.4$ ) jets. Out of these, two jets with a bottom tag are required. The need of the  $b$ -tagged jets is a main ingredient as the reconstructed top quarks are seeded around them allowing an accurate reconstruction of the top quark decay.

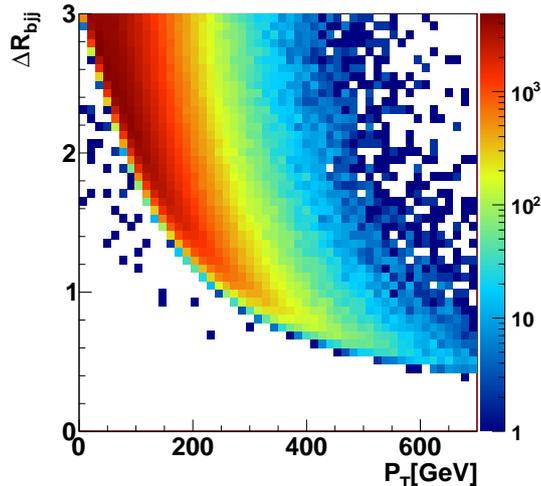


Figure 4.1: Separation in  $\eta - \phi$  plane  $\Delta R_{bjj}$  of three partons from top quark decay for Standard Model  $t\bar{t}$  MC as a function of the  $p_T$  of the top quark, from [75]

The aim is to group all jets in three different container called buckets. The first and second bucket ( $B_1$ ,  $B_2$ ) correspond to the two top quarks whereas the third bucket ( $B_{ISR}$ ) contains the extra hadronic activity of the event.

Targeting the “alljets” channel of the top quark pair decay makes it possible to reconstruct the four momenta of the top quarks by selecting the jets which originated from its decay partons. It also offers the highest branching ratio 46% of all three decay channels of the top quark pairs. Without the presence of isolated leptons, the challenge regarding this channel is the background from QCD multijet production and hadronic activity in the event not related to the top quark pair like initial state radiation. At higher  $p_T^{\text{top}}$  of the top quarks it is expected that the extra hadronic activity is less problematic due to the geometric separation between the jets as illustrated schematically in Figure 4.2. This separation is what fatjet based taggers usually make use of. At lower  $p_T^{\text{top}}$  it becomes more difficult to distinguish the decay jets by kinematical or geometrical means. Independent of the considered reconstruction procedure combinatorial background is most likely the dominant challenge for less boosted events. The idea of a simultaneous reconstruction of both top quarks is to provide as many constraints as possible to handle the combinatorics.

The basic selection with respect to Anti- $k_T$  ( $R = 0.4$ ) jets applied in the bucket algorithm is summarized in Section 3.4.8.

The selection of at least five jets is a unique feature of the bucket algorithm compared to a fully resolved analysis. It is motivated by the fact that the bucket algorithm is designed to reconstruct top quark pair events even when one jet of the top quark decay is not selected due to the momentum threshold. Generally this not selected jet is mostly the softer jet from the W decay. The  $p_T$  distribution of the

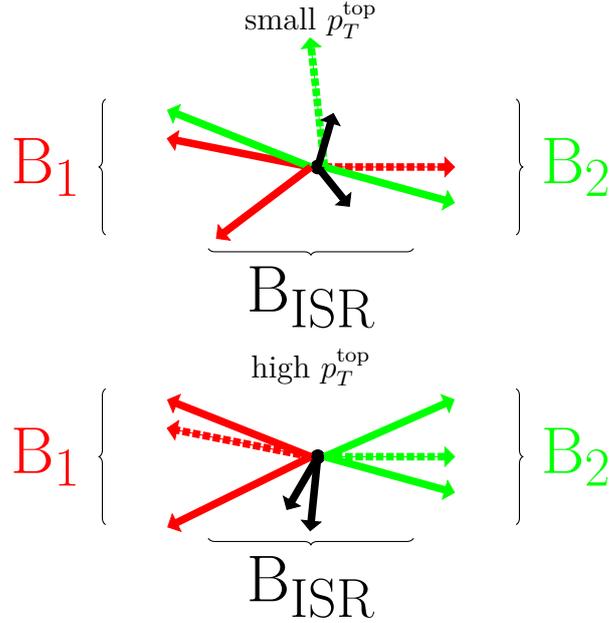


Figure 4.2: Illustration of the  $p_T$  dependence of the topology of a top quark decay. Shown are the decay products of the two tops (red and green) each with a bottom jet (dotted). The extra hadronic activity like initial state radiation is shown in black.

six hardest truth jets in a SM  $t\bar{t}$  sample where the top quarks decay hadronically is shown in Figure 4.3. The softest jet actually peaks at the momentum threshold. Hence, it can be expected that it is not used for the reconstruction in roughly half of the events. As a consequence it is not possible to reconstruct the W subsystem of the top quark decay. To deal with this issue the bucket algorithm is organized in two steps. The W reconstruction is considered in the second reconstruction step of the bucket algorithm.

The only discriminant variable when assigning the jets to the three buckets in the first step is the invariant mass of the buckets itself. This invariant mass is supposed to be sufficient to distinguish the top quark decay jets from the rest of the event. In order to calculate the grouping  $(B_1, B_2, B_{\text{ISR}})$  for each event a metric which is a function of the four momenta of the jets is minimized in the first step. For an individual bucket this metric is

$$\Delta_{B_i} = |m_{B_i} - m_t| \quad \text{with} \quad m_{B_i}^2 = \left( \sum_{j \in B_i} p_j \right)^2 \quad (4.0.2)$$

The metric  $\Delta_{B_i}$  compares the invariant mass of the sum of the four momenta in one bucket to the true top quark mass. Having one  $b$ -tagged jet in each top bucket is the only requirement while calculating this metric. No additional information in terms of e.g. jet multiplicity is imposed on the buckets. From this point of view

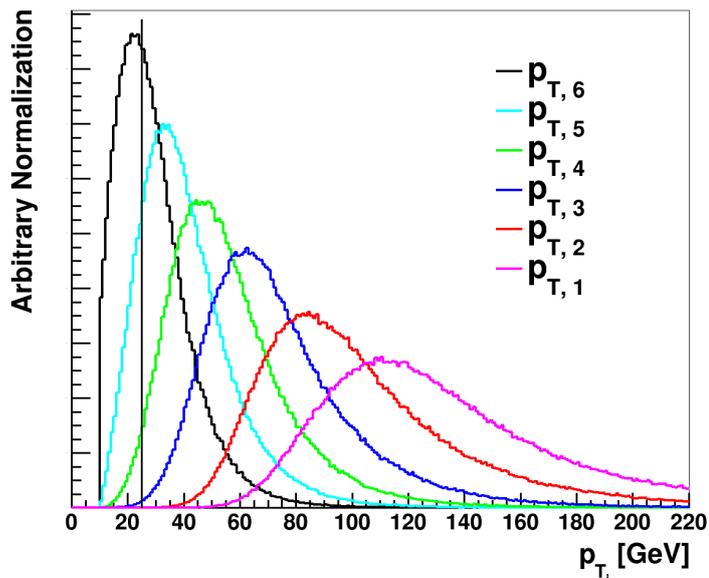


Figure 4.3: Distribution of the six hardest Anti- $k_T$  ( $R = 0.4$ ) truth jets in a  $pp \rightarrow t_h \bar{t}_h$  MC sample.

it is possible to select events with even less than five jets. But considering this smaller jet multiplicities would result in a significant increase of the QCD multijet background.

As mentioned before, one top quark will probably miss a decay jet. In this case it is unlikely that the invariant mass metric is small. For the whole event a global metric  $\Delta$  summing the two  $\Delta B_i$  is minimized. The fact that in many events only one bucket will have its invariant mass close to the top quark mass is taken into account by applying a weighting  $\omega$  between bucket  $B_1$  and bucket  $B_2$ .

$$\Delta^2 = \omega \Delta_{B_1}^2 + \Delta_{B_2}^2 \quad (4.0.3)$$

$$\omega = 100 \longrightarrow \Delta_{B_1} < \Delta_{B_2} \quad (4.0.4)$$

As a result the first bucket ( $B_1$ ) always has an invariant mass closer to the true top quark mass. This behavior can be seen in the distribution of the three bucket masses in Figure 4.4. Due to the minimization of  $\Delta$  the top buckets peak around the true top quark mass. The peak is more pronounced and narrower for the first bucket. The second bucket has a dip at the true top quark mass originating from the weighting factor. It should be mentioned that a well reconstructed top quark mass does not automatically imply that the combination of jets assigned to a bucket is indeed the correct one. The performance of the correct reconstruction of the top quark four momentum will be further investigated in Chapter 5.

The reconstruction is considered successful if the top buckets are within a mass

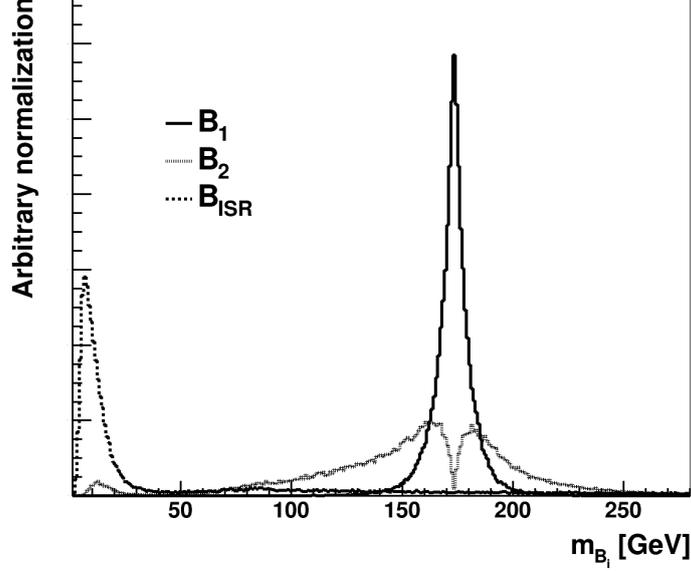


Figure 4.4: Masses of the three buckets for all events passing the basic selection (at least five jets and two of them  $b$ -tagged) in a  $pp \rightarrow t_h \bar{t}_h$  MC sample.

window

$$155 \text{ GeV} < m_{B_{1,2}} < 200 \text{ GeV} \quad (4.0.5)$$

This mass window selection is mostly motivated by its power to suppress QCD multijet background.

Further information can help to reduce the background. Therefore, the two-jet distribution inside a bucket neglecting the  $b$ -tagged jet is investigated for  $W$  candidates. This search for  $W$  candidates is done via a mass ratio cut comparing the ratio of masses of a combination of two jets  $m_{j_k, j_l}$  and the total bucket mass  $m_{B_i}$  to the ratio of the  $W$  boson mass  $m_W$  and the top quark mass  $m_t$ :

$$\left| \frac{m_{j_k, j_l}}{m_{B_i}} - \frac{m_W}{m_t} \right| < 0.15 \quad (4.0.6)$$

If a bucket fulfills equation 4.0.6 it is labeled  $\mathbf{t}_w$  otherwise  $\mathbf{t}_-$ . This classification allows to group the events into four different categories where the first entry corresponds to the bucket with the better top quark reconstruction ( $B_1, B_2$ ):

1. ( $\mathbf{t}_w, \mathbf{t}_w$ )
2. ( $\mathbf{t}_w, \mathbf{t}_-$ )
3. ( $\mathbf{t}_-, \mathbf{t}_w$ )
4. ( $\mathbf{t}_-, \mathbf{t}_-$ )

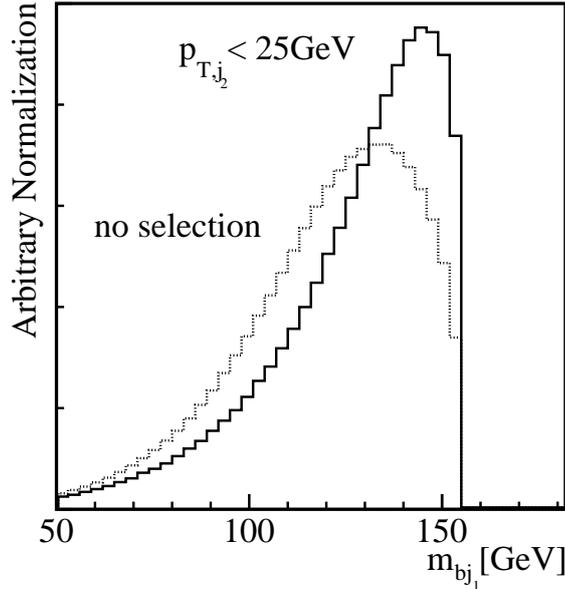


Figure 4.5: Parton level distribution of the invariant mass of the  $b$  quark and the harder W decay jet (dotted). The case where the softer jet from the W decay is below  $p_{T,j_2} < 25$  GeV is also shown (solid), from [9]

A concentration on category  $(\mathbf{t}_w, \mathbf{t}_w)$  where all six decay jets are supposedly contained in the top buckets would result in a low signal efficiency. The invariant mass of the  $\mathbf{t}_-$  buckets is generally not as close to the true top quark mass as the invariant mass of the  $\mathbf{t}_w$  buckets. The initially applied weighting between  $B_1$  and  $B_2$  in Equation 4.0.3 already accounted for this difference. Nevertheless the information that the two remaining jets after the  $p_T$  threshold cut are the  $b$ -jet which is generally the hardest jet in a hadronic top quark decay and the harder jet from the W decay can be used in the second step of the bucket algorithm. This step relies on the fact that the three body decay nature of the top quark decay induces an additional peak in the invariant mass distribution of these two decay products [76]. The invariant mass distribution  $m_{bj}$  of the  $b$ -quark and the harder quark from the W decay is shown in Figure 4.5 at parton level. The distribution has an endpoint at  $m_{bj} < \sqrt{m_t^2 - m_W^2} \simeq 155$  GeV and a quite pronounced peak around 145 GeV if the softer jet from the W decay is below  $p_T = 25$  GeV. It was shown that the two partons carry enough information about the initial top quark to accurately reconstruct the top quark four momentum [9].

The structure of the  $m_{bj}$  distribution is reflected in the new metric  $\Delta_B^{bj}$  used to reconstruct two-jet buckets.

$$\Delta_B^{bj} = \begin{cases} |m_B - 145 \text{ GeV}| & \text{if } m_B \leq 155 \text{ GeV} \\ \infty & \text{else} \end{cases} \quad (4.0.7)$$

The  $\mathbf{t}_w$  buckets are unchanged. The metric in Equation 4.0.7 takes into account

all jets in the event except for the jets contained in  $\mathbf{t}_w$ . For  $(\mathbf{t}_w, \mathbf{t}_-)$  and  $(\mathbf{t}_-, \mathbf{t}_w)$  these are the jets previously in the  $\mathbf{t}_-$  bucket and in  $B_{\text{ISR}}$ . Even though a two-jet combination will contain less information than a three-jet combination it is expected that the reduced number of jets used for the minimization of  $\Delta_B^{bj}$  will also lead to less combinatorics. In the case where both top buckets do not contain a W candidate i.e. for the event category  $(\mathbf{t}_-, \mathbf{t}_-)$  the global metric defined as

$$\Delta^{bj} = \Delta_{B_1}^{bj} + \Delta_{B_2}^{bj} \quad (4.0.8)$$

is minimized using all selected jets in the event.

Again as in the first step, in order to reject background, a mass window selection is applied:

$$75 \text{ GeV} < m_{bj} < 155 \text{ GeV} \quad (4.0.9)$$

The ranges of the mass windows in Equation 4.0.5 and in Equation 4.0.9 produce a separation of the two classes of buckets in the invariant mass distribution.

Until now the algorithm is applicable to any boost of the top quarks. The selection of top quark pair candidates is based on the mass window cuts in Equation 4.0.5 and Equation 4.0.9. Even after the mass selection the buckets can contain wrong assignments of jets. In Reference [9] it was shown that from a combinatoric perspective the grouping is only supposed to work successfully if significant boost is present for the top quarks. The events with a wrong grouping of jets into buckets tend to populate small transverse momenta of the top buckets. This  $p_T$  dependence of the wrong grouping is respected by imposing a  $p_T$  cut on the reconstructed top quarks:

$$p_T^{\text{reco}} > 100 \text{ GeV} \quad (4.0.10)$$

The need for this cut is further investigated in Chapter 5.

# 5 Performance of the Bucket Algorithm

A top tagging algorithm in the fully hadronic decay mode of the top quarks has to be able to handle hadronic activity and combinatorics. The influence of pile-up on the reconstruction performance is expected to be rather weak as the bucket algorithm uses ATLAS Anti- $k_T$  ( $R = 0.4$ ) jets which are relatively small and already corrected for the presence of pile-up. It is known [9] that a higher transverse momentum of the reconstructed top quarks can help to reduce combinatorics. The main experimental challenge for the performance of the bucket algorithm are supposed to come from worse position resolution and different  $p_T$  distributions of the reconstructed jets in the ATLAS detector compared to Reference [9]. The aim of this performance study with simulated events is to show that even in a realistic experimental environment it is still possible to reconstruct top quark pairs with low transverse momentum  $p_T < 200$  GeV.

The performance is tested with two different Monte-Carlo samples, SM  $t\bar{t}$  and  $pp \rightarrow Z' \rightarrow t\bar{t}$ . Both samples allow to test different  $p_T$  regions of the algorithm. The background is simulated by QCD dijet samples. The ability to suppress QCD multijet background as well as the accurate reconstruction of the four momenta of the top quark pairs is analyzed.

To distinguish inefficiencies coming from acceptance to those induced by the algorithm, events are preselected if they have a hadronically decaying top quark pair with  $p_T > 100$  GeV and  $|\eta| < 2.5$  at truth level.

As a measure of the accuracy of the reconstruction the top buckets are matched to truth level top quarks in the  $(\eta - \phi)$  plane. For each top bucket  $B_i$  the matching distance  $R_i$  is calculated. In order to consider the reconstruction successful, it is required that the matching distance  $R_i$  fulfills

$$R_i = \min[\Delta R(\vec{B}_i, \vec{t}_1), \Delta R(\vec{B}_i, \vec{t}_2)] < 0.5 \quad (5.0.1)$$

$$\text{with } \Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (5.0.2)$$

The  $t_-$  (two-jet) buckets are supposed to only miss the softest jet from the top quark decay. Therefore the geometric matching should be sufficient to describe well-reconstructed top quarks.

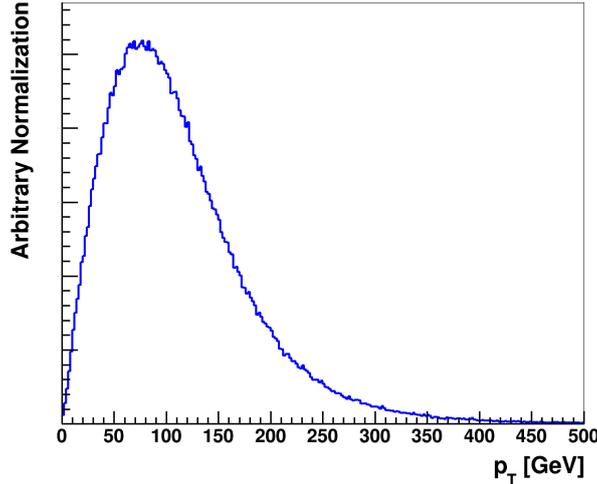


Figure 5.1: Truth level  $p_T$  distribution of the top quarks in the allhadronic decay mode of SM  $t\bar{t}$ .

## 5.1 Performance of simulated Standard Model events

The  $p_T$  distribution of SM top quark pairs in the all hadronic mode are shown in Figure 5.1. The  $p_T$  distribution peaks shortly below the  $p_T$  cut ( $p_T^{\text{reco}} > 100$  GeV) imposed on the reconstructed top quarks in the bucket algorithm. It is clear that it is less likely for a low  $p_T$  top quark to fake a substantial higher transverse momentum for the top buckets. This shift to higher values of the transverse momentum would require a significant fraction of additional  $p_T$  in an event. But as there are so many top quarks close to the threshold and the distribution is strongly falling the selection at truth level for the top quarks is necessary to see the performance of the bucket algorithm in its targeted momentum regime. The impact of low  $p_T$  top quarks will be discussed separately while comparing the Monte-Carlo with data in Chapter 6.

The invariant mass of the reconstructed  $t_{-}$  buckets using the second metric  $\Delta_B^{bj}$  is shown in Figure 5.2. The distribution is without further selection on the reconstructed top quarks. The peak around 145 GeV is present but broader compared to the parton level distribution in Figure 4.5.

It is necessary to find a way to distinguish the well matched ( $R_i < 0.5$ ) from the imperfect ( $R_i > 0.5$ ) reconstructed top quarks. A simple and effective way is a cut on the  $p_T^{\text{reco}}$ . The distribution of the transverse momentum of buckets is presented in Figure 5.3 separately for both classes of reconstructed top quarks. The imperfectly reconstructed top quarks populate predominantly low  $p_T^{\text{bucket}}$ . Of course a higher cut on  $p_T^{\text{bucket}}$  would increase the fraction of these well matched buckets. But in the light of signals with relatively small boost a higher cut on  $p_T^{\text{bucket}}$  would equally drastically reduce the signal efficiency. As one task of this thesis is to investigate SM  $t\bar{t}$  in ATLAS data the value proposed in [9] is unchanged.

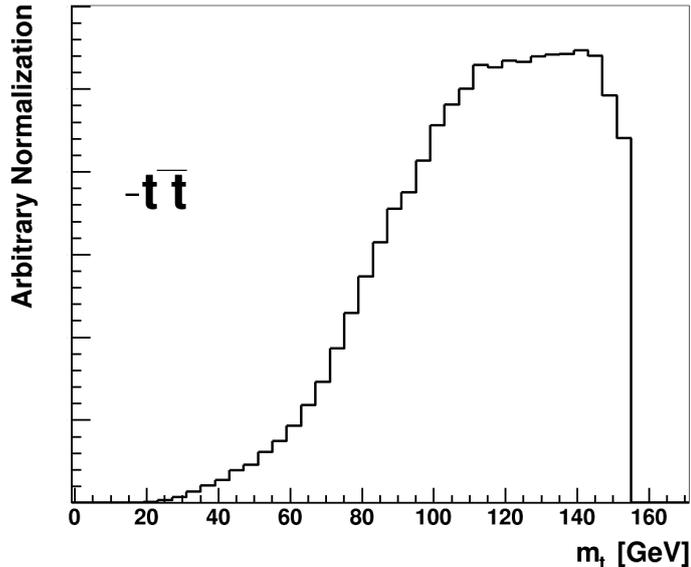


Figure 5.2: Distribution of the invariant mass of reconstructed  $\mathbf{t}_-$  buckets without further cuts on the reconstructed top quarks for SM  $t\bar{t}$  MC.

Additional discrimination power to reject background and to select well reconstructed buckets could be achieved by combining the  $p_T^{\text{bucket}}$  information with the overall geometric size of the top bucket. For  $\mathbf{t}_w$  buckets this size is the maximal geometric separation  $\Delta R_{bjj}$  between any two jets in the bucket and for  $\mathbf{t}_-$  buckets this is the separation  $\Delta R_{bj}$  between the  $b$ -tagged jet and the second jet inside the  $\mathbf{t}_-$  bucket. The correlations are shown separately for matched  $R_i < 0.5$  and unmatched  $R_i > 0.5$  buckets in Figure 5.4. Unmatched buckets not only have a smaller boost but also a larger separation. The distribution of  $\Delta R_{bjj}$  additionally indicates how the bucket algorithm is able to successfully reconstruct top quarks with a separation of the decay jets above  $\Delta R_{bjj} > 1.5$  which is a typical fatjet size making it possible to consider a larger signal fraction compared to substructure based methods. This advantage is also valid for higher boosts as will become more evident when discussing the  $Z'$  samples in Chapter 5.2.

### Resolution

The cut  $p_T^{\text{reco}} > 100$  GeV still guarantees a large signal efficiency for SM  $t\bar{t}$ . With this cut on the reconstructed top quarks it is now possible to investigate systematically the quality of the top quark reconstruction. The position resolution as well as the  $p_T$  reconstruction performance is illustrated in Figure 5.5. The matching distance  $R_i$  between a bucket  $B_i$  and the nearest top quark is shown in Figure 5.5(a). Using the value  $R = 0.5$  to define successful matching it can be seen that both kinds of buckets have a similar fraction of well matched buckets with 79% for  $\mathbf{t}_w$  and 78% for  $\mathbf{t}_-$ . The  $R_i$  distribution of  $\mathbf{t}_w$  buckets peaks close to zero which is the result of

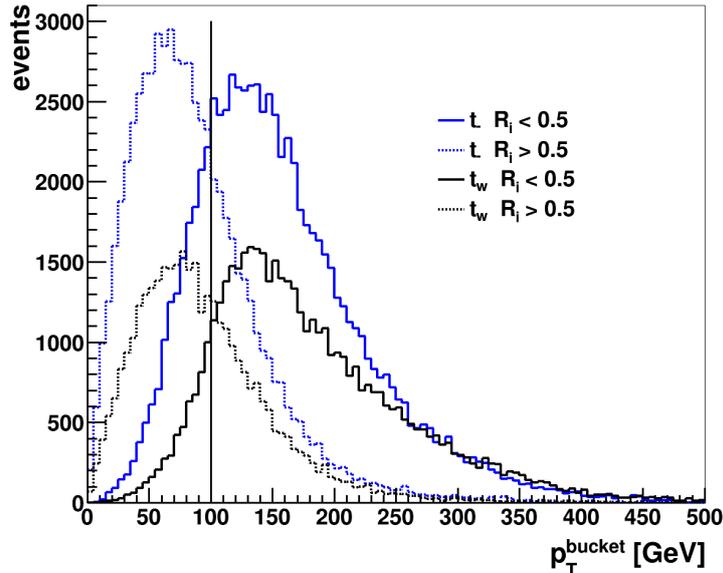


Figure 5.3:  $p_T^{\text{bucket}}$  dependence of well matched (solid line) and imperfect reconstructed (dotted line) buckets of the SM  $t\bar{t}$  MC. The buckets are shown after the mass window selections. For a doubletag both top buckets in an event would additionally have to pass the complete bucket selection ( $p_T^{\text{reco}} > 100$  GeV).

the correct assignment of all three decay jets. In contrast, the  $R_i$  distribution of  $t_-$  buckets is broader with a peak further away from zero. The shift of the peak compared to  $t_w$  is caused by the jet missing from the top quark decay. However, the partial reconstruction of the top quark decay jets is able to adequately reconstruct the direction of the top quark in the  $\eta - \phi$  plane. It should be mentioned that this plot only reveals the fraction of well matched buckets per a single bucket. The efficiency to find two buckets in one event satisfying  $R_{1,2} < 0.5$  will be slightly lower and investigated in more detail as a function of  $p_T$  in the following.

For the two bucket classes the absolute  $p_T$  and relative  $p_T$  differences to the nearest top quark are shown in Figure 5.5(b) and 5.5(c). Both plots indicate that the resolution for  $t_w$  buckets is better with a narrower distribution peaking at zero. The  $p_T$  distribution of  $t_-$  buckets is shifted towards negative values and has a broader distribution with respect to  $t_w$  buckets. This behavior of the  $p_T$  distribution of  $t_-$  buckets is expected as one jet from the top quark decay with  $p_T < 25$  GeV is missing. The position of the peak in the absolute  $p_T$  difference is also compatible with a missing soft decay jet, see Figure 5.5(b). The fraction of buckets with good momentum resolution satisfying  $|\Delta p_T / p_T^{\text{top}}| < 20\%$  is 61% for  $t_-$  buckets compared to 71% for  $t_w$  buckets.

The correlation between the position resolution quantified by  $R_i$  and the  $p_T$  resolution was further investigated. The buckets which are geometrically matched indeed are the ones with good momentum reconstruction.

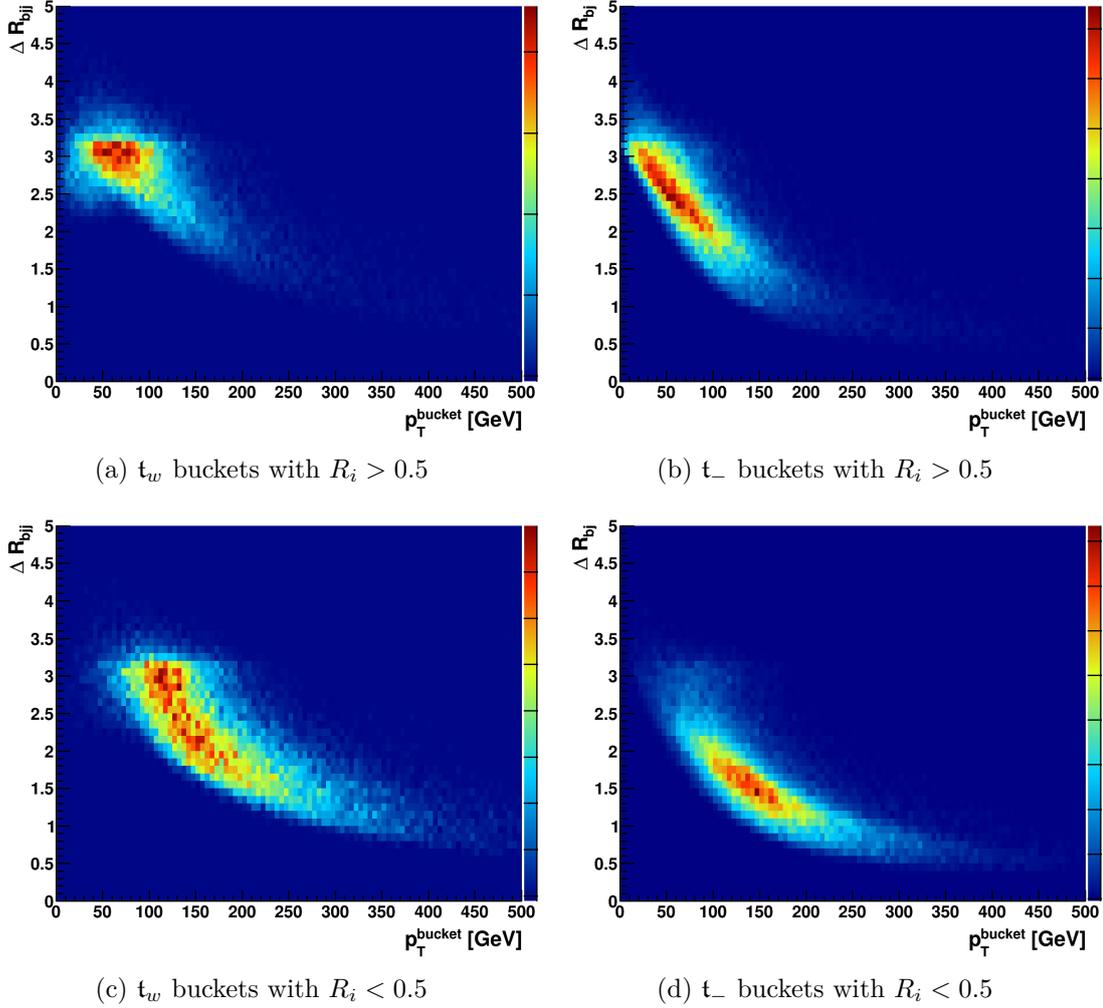


Figure 5.4: Correlation between the geometric size and the  $p_T$  of the unmatched (upper row) and matched (lower row) reconstructed top quarks for a SM  $t\bar{t}$  MC. The  $t_w$  buckets consisting of three jets are shown on the left,  $t_-$  buckets with two jets on the right.

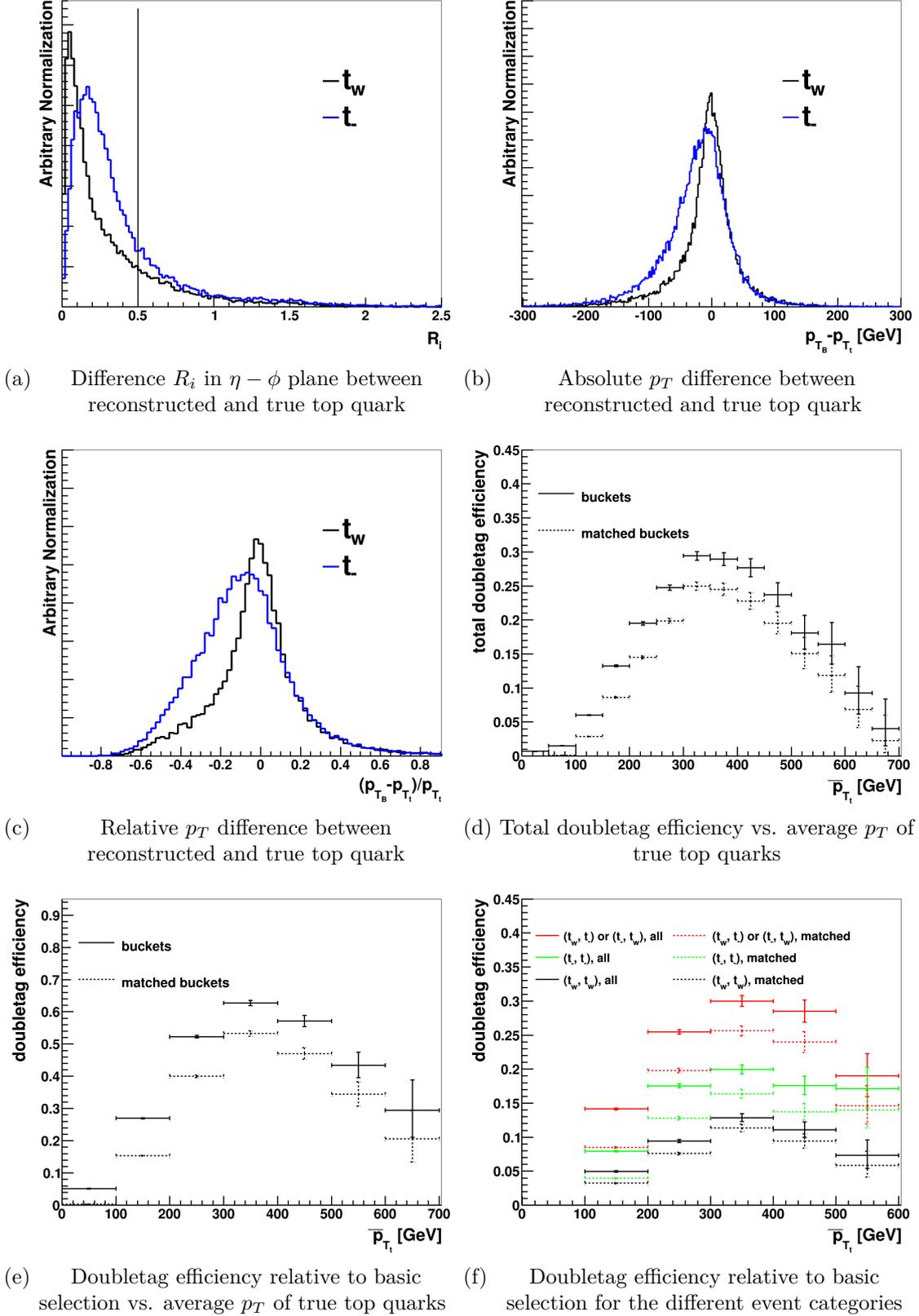


Figure 5.5: Performance plots of bucket algorithm for SM  $t\bar{t}$  MC. For a doubletag both top buckets in the event have to pass the bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV). The buckets  $\mathfrak{t}_w$  containing a W candidate are shown in black the two-jet buckets  $\mathfrak{t}_-$  in blue. 53

## Efficiencies

The construction procedure of the bucket algorithm with its low transverse momentum thresholds for the individual jets used for the assignment to buckets guarantees a high selection efficiency.

Because the baseline selection efficiency has no impact on the performance of the algorithm itself the efficiency is separated in the efficiency of the basic selection and the bucket selection efficiency. The basic selection efficiency is defined in the following way

$$\epsilon(\text{basic selection}) = \frac{(n_{\text{jets}} \geq 5) \text{ AND } (n_{b\text{-jets}} \geq 2)}{\text{generated events}} \quad (5.1.1)$$

This basic selection efficiency already has a specific  $\eta$  and  $p_T$  dependence originating mainly from the corresponding  $b$ -tagging efficiencies. The  $b$ -tagging algorithms used are discussed in Chapter 3.4.6.

The doubletag efficiency  $\epsilon(\text{doubletag})$  is the efficiency to find two top buckets passing the bucket selection criteria

$$\epsilon(\text{doubletag}) = \frac{(\text{mass windows}) \text{ AND } (p_T^{\text{reco}} > 100 \text{ GeV})}{\text{basic selection}} \quad (5.1.2)$$

This efficiency is used to study the dependencies of the bucket algorithm on various quantities. The total doubletag efficiency  $\epsilon(\text{total doubletag})$  which is the product of the basic selection efficiency and the doubletag efficiency allows to compare the efficiency of the bucket algorithm for a given sample to other algorithms.

$$\epsilon(\text{total doubletag}) = \epsilon(\text{basic selection}) \times \epsilon(\text{doubletag}) \quad (5.1.3)$$

Additionally the efficiency  $\epsilon(\text{matched})$  of matching the selected two buckets to the nearest true top quark within  $R_i < 0.5$  is investigated.

$$\epsilon(\text{matched}) = \frac{(R_1 < 0.5) \text{ AND } (R_2 < 0.5)}{\text{basic selection}} \quad (5.1.4)$$

The overall efficiencies for SM  $t\bar{t}$  are summarized in Table 5.1. They are also compared to a QCD dijet sample to highlight the intrinsic background suppression of the bucket algorithm. The doubletag efficiency for fully hadronic SM  $t_h\bar{t}_h$  is 34.2%. Out of these selected events the fraction of matched events is 66.6%. The unmatched events sit at lower  $p_T^{\text{bucket}}$  (see Figure 5.3). The category  $(\mathbf{t}_w, \mathbf{t}_w)$  constitutes 6.3%. Adding the  $\mathbf{t}_-$  class of buckets provides roughly four times the signal compared to only considering  $(\mathbf{t}_w, \mathbf{t}_w)$ . The matched fraction for  $(\mathbf{t}_w, \mathbf{t}_w)$  and  $(\mathbf{t}_w, \mathbf{t}_-)$  are comparable ( $\approx 70\%$ ). This similar matching performance reveals that  $\mathbf{t}_-$  buckets can provide the same reconstruction quality as  $\mathbf{t}_w$  buckets. Going to  $(\mathbf{t}_-,$

category	SM $t\bar{t}$ MC		QCD dijet MC
	$\epsilon(\text{doubletag})$	$\epsilon(\text{matched})$	$\epsilon(\text{doubletag})$
total	34.2%	22.8% (66.6%)	3.3%
$(\mathbf{t}_w, \mathbf{t}_w)$	6.3%	4.6% (72.3%)	0.1%
$(\mathbf{t}_w, \mathbf{t}_-)$	15.2%	10.6% (69.6%)	1.0%
$(\mathbf{t}_-, \mathbf{t}_w)$	2.1%	1.2% (54.7%)	0.1%
$(\mathbf{t}_-, \mathbf{t}_-)$	10.6%	6.5% (61.2%)	2.0%

Table 5.1: Doubletag efficiency of finding two buckets in event passing the bucket selection cuts with respect to the basic selection. Doubletag efficiency for buckets matched to true top quark within  $R_i < 0.5$  with respect to basic selection. In parenthesis the relative matched fraction with respect to the doubletag is quoted.

$\mathbf{t}_-$ ) the fraction of matched events decreases slightly to 61.2%. The reason for this decrease is that the efficiency of finding a correct second bucket is correlated with the efficiency to find the first one. In  $(\mathbf{t}_w, \mathbf{t}_-)$  the  $\mathbf{t}_w$  bucket is already found when the two-jet bucket is constructed. Hence, for the reconstruction of the  $\mathbf{t}_-$  bucket only a reduced number of jets in the event is considered. As a consequence the combinatorics are also reduced. In contrast in  $(\mathbf{t}_-, \mathbf{t}_-)$  both buckets are reconstructed at once taking into account all jets in the event. The higher considered multiplicity of jets makes it easier to select a wrong combination.

The third category  $(\mathbf{t}_-, \mathbf{t}_w)$  has the smallest doubletag efficiency (2.1%) and the smallest fraction of matched events (54.7%). Having a  $\mathbf{t}_-$  bucket in the first entry means that the bucket which had the better top quark mass in the first reconstruction step did not contain a W candidate. An explanation could be that two of the top quark decay jets merged resulting in a bucket with good top quark mass reconstruction in the first reconstruction step but without a W candidate in the two-jet distribution. In this case the  $\Delta_B^{bj}$  minimization would hardly find the correct jet assignment as the two aimed jets are not distinct anymore.

The doubletag efficiency for the dominant QCD background process of multijet production is 3.3%. The background suppression is achieved by requiring the buckets to be in appropriate mass windows and by the  $p_T^{\text{reco}}$  cut which not only reduces wrong combinations in the signal sample but also rejects background events. Event categories containing a  $\mathbf{t}_-$  bucket like the  $(\mathbf{t}_w, \mathbf{t}_-)$  category with a doubletag efficiency of 1% and the  $(\mathbf{t}_-, \mathbf{t}_-)$  category with 2% give the dominant contribution to the overall QCD background doubletag efficiency. The additional information contained in  $\mathbf{t}_w$  buckets helps to suppress background events resulting in smaller doubletag efficiencies for event categories containing  $\mathbf{t}_w$  buckets. Demanding one  $b$ -tagged jet in each top bucket from the beginning raises the background combinatorics. The  $p_T$  dependence of the  $b$ -tagger supports that the background looks like top quark pair events.

To study the pile-up dependence of the bucket algorithm the efficiencies of different steps of the selection against the average number of interactions per bunch crossing  $\langle\mu\rangle$  are compared in Figure 5.6. The basic selection efficiency is split in the efficiency of finding at least five jets and the efficiency of finding two  $b$ -tagged jets in these jets. The doubletag efficiency for buckets and matched buckets is also shown. The doubletag efficiency of the buckets and the  $b$ -jet efficiency have only a weak  $\langle\mu\rangle$  dependence. The only weak pile-up dependence of the doubletag efficiency is expected because Anti- $k_T$  ( $R = 0.4$ ) jets which are already pile-up corrected are used for the reconstruction. The size of the Anti- $k_T$  ( $R = 0.4$ ) jets would be too small to improve their composition by further looking into their substructure. The pile-up mostly affects the  $p_T$  of the reconstructed Anti- $k_T$  ( $R = 0.4$ ) jets [17]. It is less likely that pile-up produces additional hard jets in the event which could increase the multiplicity after the basic selection cuts. Such an effect could be problematic as additional jets eventually lead to a reduction of correct jet assignments to the three buckets. The efficiency of finding at least five jets is also contained in Figure 5.6 indicating that no further jets arise from pile-up. The successful pile-up subtraction [17] guarantees that the  $p_T$  resolution of the reconstructed jets is only weakly pile-up depended as well. Hence, also the matched doubletag efficiency has only a weak pile-up dependence.

The efficiency of finding at least five jets even decreases for higher  $\langle\mu\rangle$ . This decrease can be explained by an overcorrection from the pile-up subtraction [17]. It is a known issue that the JVF cut as presented in Chapter 3.4.5 leads to a reduced hard-scatter jet efficiency [77] with increasing  $\langle\mu\rangle$  as the calculation of JVF depends on the number of primary vertices in an event.

As described in the introduction the bucket algorithm targets a moderate transverse momentum regime of top quarks. The lower bound comes from the transverse momentum requirements on the reconstructed top quarks necessary to reject unmatched events. At higher boosts of the top quark it can be expected that the jets from the top quark decay start to merge. This merging would reduce the efficiency at higher top quark  $p_T$ . In order to study their  $p_T$  dependence, the efficiencies are shown against the average  $\bar{p}_T$  of the true top quarks.

$$\bar{p}_T = \frac{p_T^{\text{top1}} + p_T^{\text{top2}}}{2} \tag{5.1.5}$$

Therefore the preselection on the  $p_T$  of the true top quarks is removed in Figure 5.5(d), 5.5(e) and 5.5(f). The maximum of the doubletag efficiency is reached for  $\bar{p}_T = 300 - 400$  GeV with  $\epsilon(\text{doubletag}) \approx 60\%$ , see Figure 5.5(e). The doubletag efficiency decreases starting at 400 GeV. This decrease is consistent with the turnon

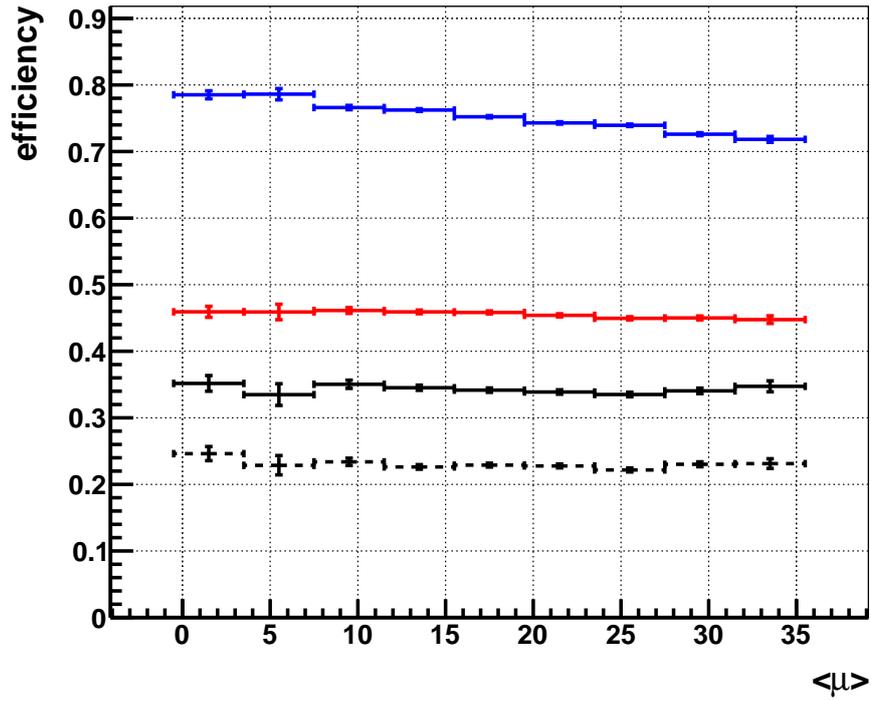


Figure 5.6: Pile-up dependence of efficiencies for SM  $t\bar{t}$  MC. The efficiency to find at least five jets is shown in blue. The efficiency that two of these jets are  $b$ -tagged jets is given in red. The doubletag efficiency for buckets is shown in solid black. The doubletag efficiency for matched buckets is shown in dotted black. For a doubletag both top buckets in an event have to pass the bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV)

for the merging which can be approximately described by

$$p_T \approx \frac{m}{R} \approx 430 \text{ GeV} \quad (5.1.6)$$

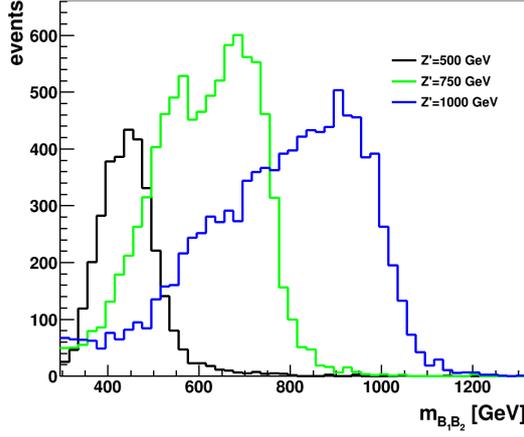
In the case of merged decay jets the whole event fails to pass the bucket selection criteria. Therefore the matched fraction stays roughly constant for high  $\bar{p}_T > 400$  GeV, see Figure 5.5(e). The  $\bar{p}_T$  dependence is different for the decrease towards smaller transverse momentum  $\bar{p}_T < 300$  GeV. The distribution falls sharper and the fraction of matched buckets decreases as well. There still remains a significant doubletag efficiency for low  $\bar{p}_T$  with  $\approx 28\%$  in  $\bar{p}_T = 100 - 200$  GeV. The relative grouping into the four categories is similar over the whole  $\bar{p}_T$  range as shown in Figure 5.5(f).

## 5.2 Validation with $Z'$ boson samples

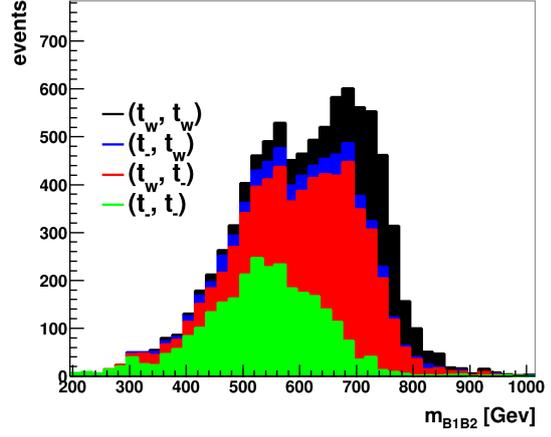
The performance of the bucket algorithm in its high efficiency regime is further validated using various  $Z' \rightarrow t\bar{t}$  samples. The heavy  $Z'$  bosons decay into  $t\bar{t}$  pairs which allows to directly apply the bucket algorithm as presented in Chapter 4. The process  $pp \rightarrow Z' \rightarrow t\bar{t}$  with a fully hadronic final state was previously studied in ATLAS excluding the mass range  $m_{Z'} < 1$  TeV [78].

For this purpose  $Z' \rightarrow t\bar{t}$  samples with  $Z'$  boson masses  $m_{Z'} = 500, 750, 1000$  GeV are chosen. The interesting observable is the invariant mass of the reconstructed top quark pair. From the discussion in the previous chapter it is known that the bucket algorithm generally provides a good directional reconstruction of the top quark. The reconstruction of the  $p_T$  should also be good as only the softest decay jet is missing. The invariant mass of the  $\mathbf{t}_-$  buckets differs from the top quark mass by construction. This imperfect reconstruction of the top quark mass for  $\mathbf{t}_-$  buckets causes a broadening and shift of the invariant mass  $m_{B_1B_2}$  of the top bucket pair as shown in Figure 5.7. The  $(\mathbf{t}_w, \mathbf{t}_w)$  reconstructs  $m_{Z'}$  reasonably well whereas  $(\mathbf{t}_w, \mathbf{t}_-)$  and  $(\mathbf{t}_-, \mathbf{t}_-)$  are shifted towards smaller  $m_{B_1B_2}$ . To account for these issues the  $\mathbf{t}_-$  buckets are ad hoc corrected to the true top quark momentum. The correction is done by setting the reconstructed mass to the true top quark mass and adding  $p_T$  according to the peak position in the absolute difference between reconstructed and true  $p_T^{\text{top}}$ . The “corrected”  $m_{B_1B_2}$  distributions are shown in the middle row of Figure 5.7.

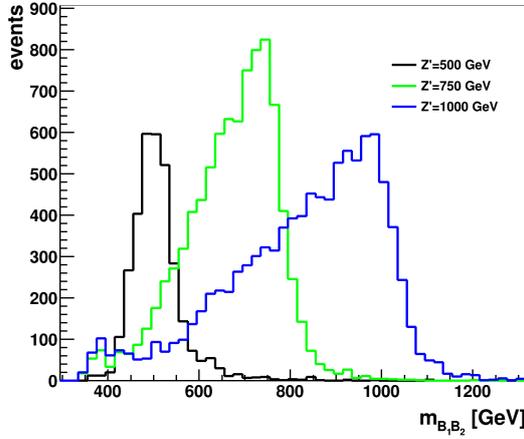
Even after this “correction” the distribution are relatively broad with tails towards smaller  $m_{B_1B_2}$ . These tails are produced by missing larger amounts of  $p_T$  in the reconstructed top quarks. If a high  $p_T$  jet which originated from the top quark decay is missing in a top bucket it is not lost but appears in the other buckets preferably in the ISR-bucket if the second top bucket is well reconstructed. After the application of the bucket algorithm this wrong jet assignment can be studied by looking for quantities describing the ISR-bucket in the event like e.g. its  $p_T$  fraction



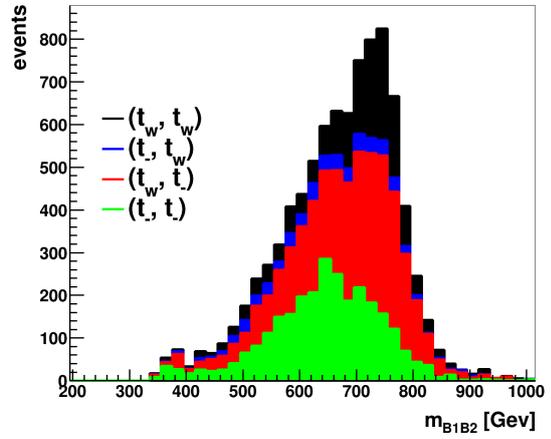
(a)  $m_{B_1 B_2}$  uncorrected



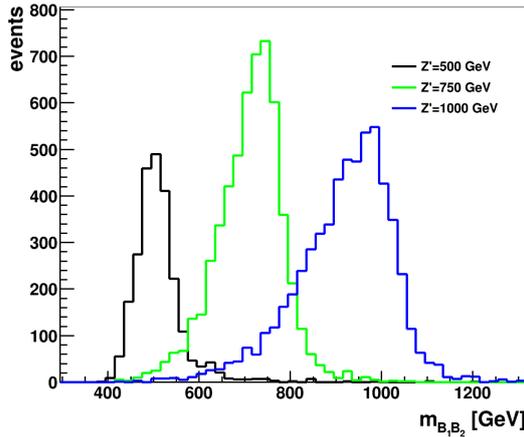
(b)  $m_{B_1 B_2}$  uncorrected of event categories for  $m_{Z'} = 750$  GeV



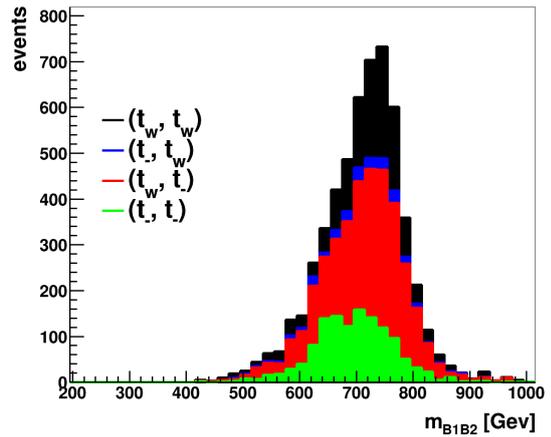
(c)  $m_{B_1 B_2}$  corrected



(d)  $m_{B_1 B_2}$  corrected of event categories for  $m_{Z'} = 750$  GeV



(e)  $m_{B_1 B_2}$  corrected and  $R_{pT}^{B_1 B_2} > 0.8$



(f)  $m_{B_1 B_2}$  corrected and  $R_{pT}^{B_1 B_2} > 0.8$  of event categories for  $m_{Z'} = 750$  GeV

Figure 5.7: Invariant mass distribution  $m_{B_1 B_2}$  of bucket algorithm for different  $Z'$  boson masses  $m_{Z'}$  (left) and corresponding  $m_{B_1 B_2}$  distribution of the bucket event categories for  $m_{Z'} = 750$  GeV (right).

with respect to the total event.

$$R_{p_T}^{\text{ISR}} = \frac{\sum_{\text{jets} \in B_{\text{ISR}}} p_T}{\sum_{\text{jets}} p_T} \quad (5.2.1)$$

$$R_{p_T}^{\text{B1B2}} = \frac{\sum_{\text{jets} \in B_{1,2}} p_T}{\sum_{\text{jets}} p_T} = 1 - R_{p_T}^{\text{ISR}} \quad (5.2.2)$$

To quantify the accurate reconstruction of the  $p_T$  the top buckets are considered to be matched if additionally to the geometric match the requirement  $|\Delta p_T/p_T^{\text{top}}| < 20\%$  is satisfied. The resulting  $R_{p_T}^{\text{B1B2}}$  distribution is shown in Figure 5.8 for matched and unmatched events. Obviously matched events favor large  $R_{p_T}^{\text{B1B2}}$  which reflects the fact that the initial state radiation is generally more forward compared to the more transverse top quark pairs. It is possible to match events with smaller  $R_{p_T}^{\text{B1B2}}$  values but the matched  $R_{p_T}^{\text{B1B2}}$  distribution clearly peaks at high values above 0.9. The “unmatched” events are not only containing buckets where the reconstruction failed. The cut on the  $p_T$  resolution is quite strict and selects only the top buckets with very good top quark momentum reconstruction. The effect of an additional cut on  $R_{p_T}^{\text{ISR}} > 0.8$  on the  $m_{B1B2}$  distribution is shown in Figure 5.7(e) and Figure 5.7(f). It mainly affects the categories containing a  $\mathbf{t}_-$  bucket as the  $\mathbf{t}_w$  reconstruction is already good before the cut. The still quite broad distribution of the  $(\mathbf{t}_-, \mathbf{t}_-)$  category is caused by the sometimes inadequate ad hoc correction of the  $\mathbf{t}_-$  buckets.

The impact of all improvements combined is quantified by the mean and root-mean-square of the  $m_{B1B2}$  distributions and is summarized in Table 5.2. With the additional requirement on  $R_{p_T}^{\text{B1B2}}$  the total doubletag efficiency of the bucket algorithm is reduced e.g. from 22% to 12.8% for  $m_{Z'} = 1000$  GeV. For a  $Z'$  boson with smaller mass  $m_{Z'} = 500$  GeV and resulting top quarks with smaller transverse momenta the total doubletag efficiency is only slightly reduced from 7.8% to 5.2% but also without providing a significant improvement of the  $Z'$  boson mass reconstruction accuracy.

The requirement of two  $b$ -tagged jets and the merging of the jets from the top quark decay results in a reduction of efficiency for higher  $Z'$  boson masses  $m_{Z'} > 1000$  GeV. As the  $Z'$  boson mass region where the bucket algorithm would potentially complement fatjet based analysis is already excluded, the bucket algorithm is not a prime candidate for searches for heavy resonances. However, this study proves that the bucket algorithm is well capable of reconstructing invariant masses of particles decaying to top quark pairs.

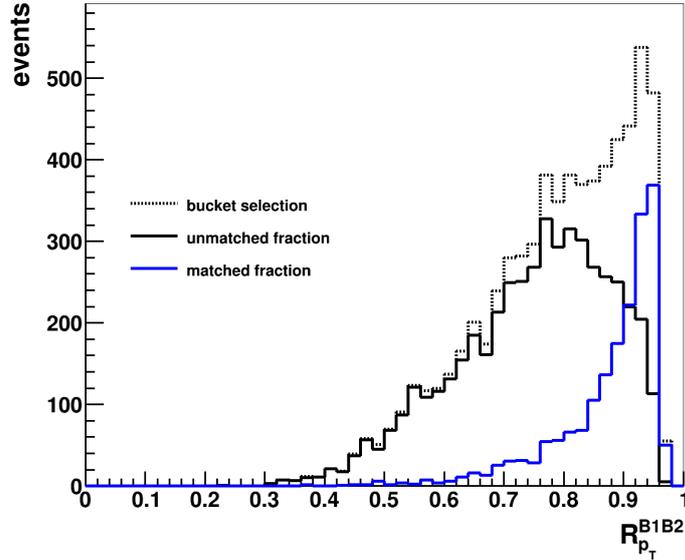


Figure 5.8: The distribution of  $R_{p_T}^{B1B2}$  for  $Z' \rightarrow t\bar{t}$  MC with  $m_{Z'} = 750$  GeV after bucket selection cuts. The matched fraction (blue) comprises top buckets that are matched geometrically within  $R_i < 0.5$  and additionally matched by requiring  $\Delta p_T/p_T < 20\%$  with respect to the true top quarks.

$m_{Z'}$ [GeV]	Buckets default			Buckets $R_{p_T}^{B1B2} > 0.8$		
	efficiency	mean [GeV]	rms [GeV]	efficiency	mean [GeV]	rms [GeV]
500	7.8%	511 (450)	66 (75)	5.2%	514	65
750	19.5%	670 (609)	105 (120)	12.1%	712	77
1000	22.0%	831 (772)	170 (184)	12.8%	918	107

Table 5.2: Total doubletag efficiency for several  $Z' \rightarrow t\bar{t}$  samples with respect to only fully hadronically decaying top quark pairs. The mean and root-mean-square of the dibucket mass  $m_{B1B2}$  distribution is given for the default bucket selection and with an additional cut on  $R_{p_T}^{B1B2} > 0.8$ . The numbers in parenthesis correspond to the uncorrected values.

# 6 Model to Data comparison for the Bucket Algorithm

In this chapter various kinematic distributions of the reconstructed top quarks and the rest of the event e.g. properties of the ISR-Bucket are presented. The procedure of data selection and background modeling aims at illustrating the performance of the bucket algorithm which is applied to data for the first time. Under the assumption that the cross section for SM  $t\bar{t}$  production is well known the data to Monte-Carlo comparison can also show how well the bucket algorithm suppresses backgrounds. As background for Standard Model  $t\bar{t}$  production only QCD multijets Monte-Carlo samples are considered. Because the bucket algorithm relies on a moderately high jet multiplicity, which are poorly modeled in the available dijet samples, it is necessary to use a data driven approach for the background modeling. In addition to the basic and bucket selection described in Chapter 5, a requirement on the geometric separation between the reconstructed top quark pairs is imposed. They have to satisfy  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ . This requirement aims at rejecting events with an inaccurate reconstruction of the top quark four momentum.

## 6.1 Trigger study

By construction the bucket algorithm requires always two  $b$ -tagged jets. Due to its two-step procedure targeting hadronically decaying top quark pairs the bucket algorithm also needs a high multiplicity of jets and a relatively low  $p_T$  threshold for these jets.

In the data to Monte-Carlo comparison, as presented in the following chapters, a trigger with the requirement of five jets each with transverse momentum of at least  $p_T \geq 55$  GeV at the EF is used. It is the unrescaled multijet trigger with the lowest  $p_T$  requirement for at least five jets. The trigger efficiency corresponding to the multijet trigger for events that have been selected with the bucket algorithm is 40% for the fully hadronic SM  $t\bar{t}$  Monte-Carlo and 18% for the QCD dijet Monte-Carlo, see Table 6.1. This multijet trigger was fully efficient during 2012 data taking. Besides the multijet trigger, two  $b$ -jet triggers are investigated. A detailed description of the considered triggers is given in Chapter 3.4.10.

An alternative study of buckets with leptonically decaying top quarks would need a significant modification of the reconstruction algorithm compared to hadronically top quark decays making the channels difficult to compare. It is also unclear

trigger	SM $t\bar{t}_{\text{allhad}}$	QCD dijet
five jets with $p_T^{\text{jet}} \geq 55$ GeV	40%	18%
one $b$ -jet with $p_T \geq 45$ and four jets with $p_T^{\text{jet}} \geq 45$ GeV	47%	31%
two $b$ -jets with $p_T \geq 35$ and three jets with $p_T^{\text{jet}} \geq 35$ GeV	48%	32%

Table 6.1: Trigger efficiency after the bucket selection for various unrescaled triggers for a fully hadronic SM  $t\bar{t}$  MC and a QCD dijet MC. The trigger requirements correspond to the selection at EF level. The trigger used in the data to Monte-Carlo comparison is highlighted in blue. The trigger requiring two  $b$ -jets imposes looser requirements on the  $b$ -tagging than the one  $b$ -jet trigger. The triggers are described in more detail in Chapter 3.4.10.

whether in the semileptonic channel which can give only one mass constraint for the top quarks, it is possible to reduce the combinatorial background. This channel is evidently more troublesome because it also needs to select the correct  $b$ -tagged jet coming from the hadronic decay just by the minimization of the mass metric. Finally, in the lepton channel with high multiplicity of jets and rather low  $p_T$  thresholds the transverse momentum distribution of all decay jets from the top quark pair will look other than in the allhadronic decay mode. It could be more likely to have all jets from the  $W$  decay above the jet  $p_T$  threshold as only one hadronic top quark is present. Having more often a full reconstruction of the  $W$  decay would make it questionable if the reconstruction procedure of 2jet-buckets gives a serious improvement for the signal efficiency.

Without the possibility to look into the semileptonic decays of the top quarks and therefore no easy way to trigger, one needs a trigger in accordance with the main topology for the buckets: a moderately high jet multiplicity and two bottom-tags. However, all existing unrescaled multijet triggers require high transverse momentum for the jets. These high momentum thresholds are not especially favorable for testing the bucket algorithm as they enhance the probability to miss the softest jet from the  $W$ -decay increasing the fraction of  $t_{\perp}$  buckets. Additionally, the higher  $p_T$  thresholds can cause that not only the softest but even harder jets from the top quark decay are rejected. The lack of further jets from the top quark decay would lead to a less accurate reconstruction of the top quark four momentum as this scenario is not considered in the construction procedure of the bucket algorithm. Nevertheless a trigger requiring one hard jet or multiple harder jets compared to the standard bucket selection will also shift the  $p_T$  spectrum of the top quarks under consideration towards higher values where the bucket algorithm is more efficient. The influence of the multijet-trigger on the performance of the bucket algorithm can be seen from the baseline doubletag efficiency of SM  $t_h\bar{t}_h$  as a function of the average true top quark  $\bar{p}_T$  in Figure 6.1 where the applied multijet-trigger is com-

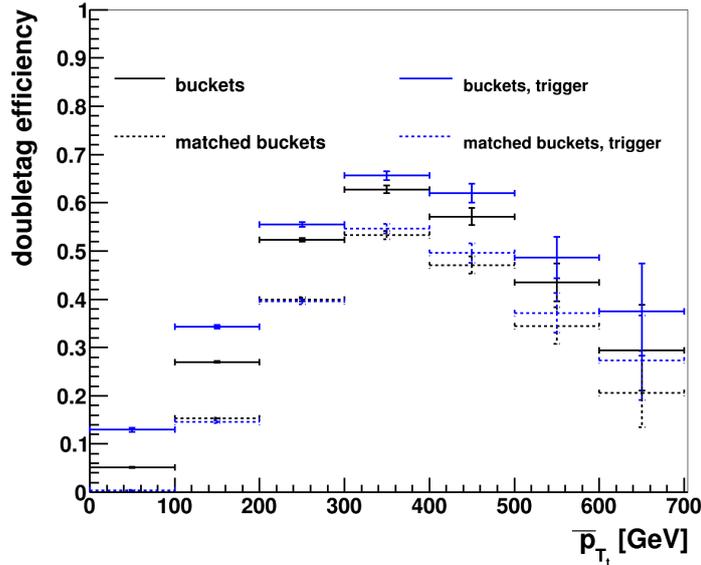


Figure 6.1: Efficiency with respect to the basic selection of finding two buckets which pass the bucket selection cuts (solid) or additionally the geometrical truth matching (dotted) as a function of the average  $\bar{p}_T$  of the true top quarks. The black points are the efficiency for a SM  $t_h\bar{t}_h$  MC without a trigger and the blue points are the efficiency with the multijet trigger requiring five jets with  $p_T > 55$  GeV at the EF level.

pared to the doubletag efficiency without a trigger. The doubletag efficiency for the bucket selection including the trigger is larger over the whole  $\bar{p}_T$  range whereas the doubletag efficiency for the matched buckets is unchanged. The shift of the top quark  $p_T$  spectrum to higher transverse momenta and the reduced matching efficiency compensate each other. As a consequence the overall fraction of matched buckets  $\approx 60\%$  after the bucket selection does not change for  $t_h\bar{t}_h$  with or without the application of the multijet trigger.

As mentioned above the purpose of a data to Monte-Carlo comparison of the bucket algorithm is to investigate its background suppression power and whether the two step structure of the algorithm with its two different classes of buckets is correctly described in Monte-Carlo. Therefore it is not necessary to have a high trigger efficiency for SM  $t\bar{t}$  especially considering that the algorithm has a significant doubletag efficiency in the  $p_T$  range populated by SM  $t\bar{t}$  events which are produced with a high rate at the LHC. Hence, even a relatively inefficient trigger selects enough top quarks to test the algorithm. As it is desired to test the bucket algorithm for only moderately boosted top quarks ( $p_T^{\text{top}} < 200$  GeV), where fatjet based taggers are normally inefficient, a trigger with a very high requirement on the  $p_T$  of a single jet or a very large  $\sum E_T$  is not used in the data to Monte-Carlo comparison presented here as these triggers select mostly top quark pairs

with higher  $p_T$ . One way to come closer to the proposed basic selection of the bucket algorithm is to consider a  $b$ -jet trigger in combination with a multijet trigger. Because in all cases a data driven background modeling is necessary the available  $b$ -jet triggers reduce the possibilities to define appropriate background control regions as the  $b$ -jet reconstruction in the trigger operates at a smaller  $b$ -jet efficiency than the one corresponding to the considered offline working point.

## 6.2 Data to Monte-Carlo comparison

In a first step, the expected cross section for SM  $t\bar{t}$  and QCD dijet samples are used to estimate the yields in the signal region. In Figure 6.2 the distributions are shown with the predicted normalization for  $t\bar{t}$  and QCD. The jet multiplicity, the grouping into the four event categories and the transverse momentum distributions of the buckets are presented. The QCD dijet Monte-Carlo is composed of several samples generated according to the leading transverse momentum of the jets. Samples containing high  $p_T$  jets are more likely to have high jet multiplicities and to survive the basic selection cuts. Therefore, buckets with smaller transverse momentum have a worse statistical precision as can be seen in e.g. Figure 6.2(d). The Monte-Carlo yields do not describe the number of events in the data adequately. Under the assumption that the SM  $t\bar{t}$  rate is correct the QCD dijet sample would require a correction factor of  $\approx 0.5$  to satisfy  $n_{\text{data}} \simeq n_{\text{QCD}}^{\text{MC}} + n_{t\bar{t}}^{\text{MC}}$ . Obviously this approach is not suited for testing the buckets.

In addition, the QCD dijet Monte-Carlo distributions are normalized to the difference between data and  $t\bar{t}$  Monte-Carlo. The jet multiplicity, the grouping into the four event categories and the transverse momentum distributions of the buckets are shown in Figure 6.3. Within the errors the distributions seem to agree quite well. But a mismodeling of a global event quantity like the jet multiplicity already indicates that the grouping into different bucket categories can not be correctly described as this quantity is strongly depending on the number of jets, e.g. the  $(\mathbf{t}_w, \mathbf{t}_w)$  category already requires at least six jets. Nevertheless, the kinematics of the reconstructed top quarks in terms of  $p_T$  distributions are accurately described.

## 6.3 Background estimate from data

In order to test the bucket algorithm it is necessary to model the background well, especially for global event variables like the multiplicity of jets as even in the signal region the QCD background dominates the shape of the distributions which makes it hard to check the description of SM  $t\bar{t}$  distributions. With such a high background the kinematic distributions of the SM  $t\bar{t}$  Monte-Carlo can not be validated in great detail but at least the probability of missing a jet from the  $W$ -decay due to the momentum threshold and the corresponding grouping in  $\mathbf{t}_w$

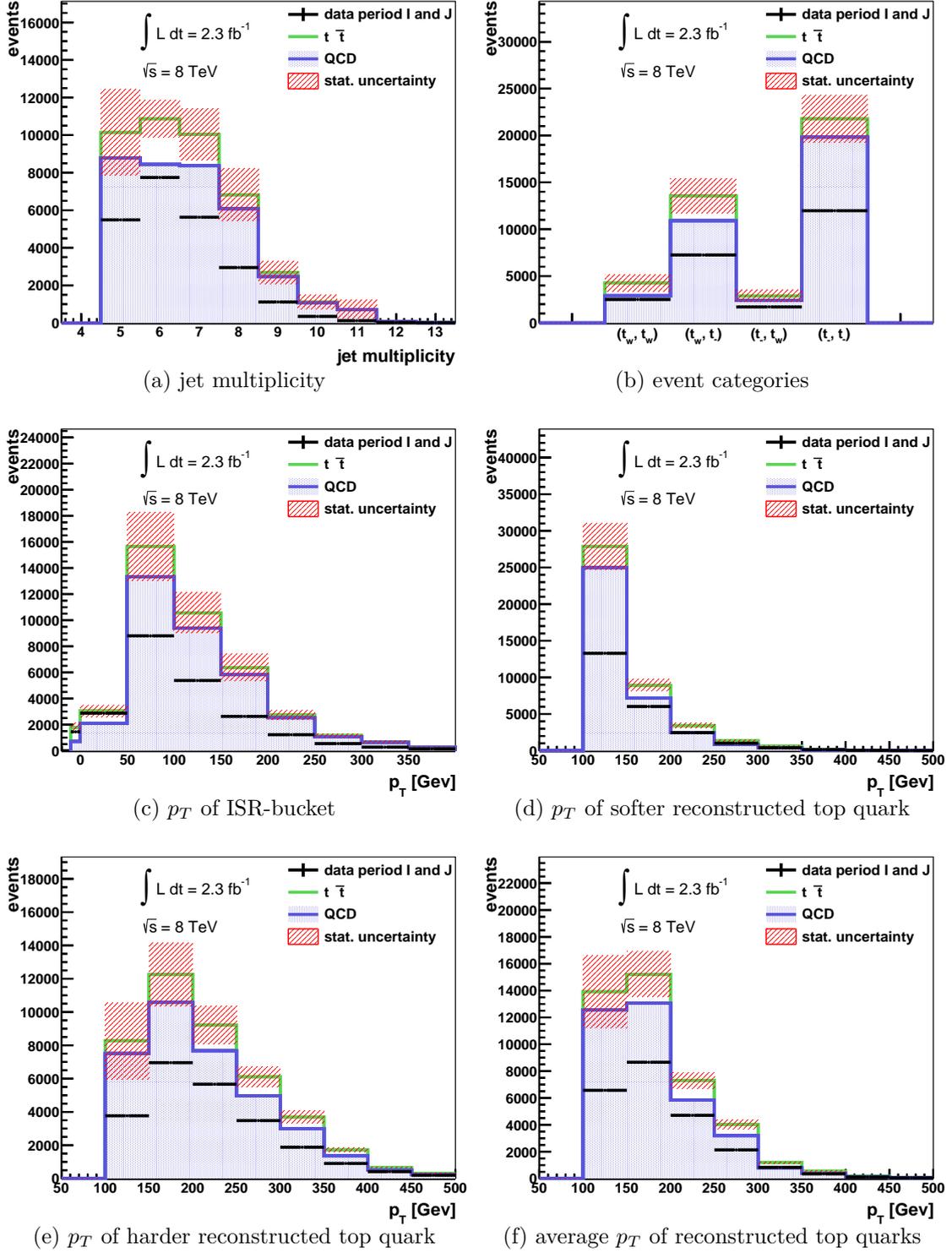


Figure 6.2: Data to Monte-Carlo comparison plots after bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV,  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ ) in the signal region. The QCD MC is modeled by a QCD dijet sample. The SM  $t\bar{t}$  is modeled by a MC sample of the fully hadronic decay mode and a MC sample of the leptonic decay modes.

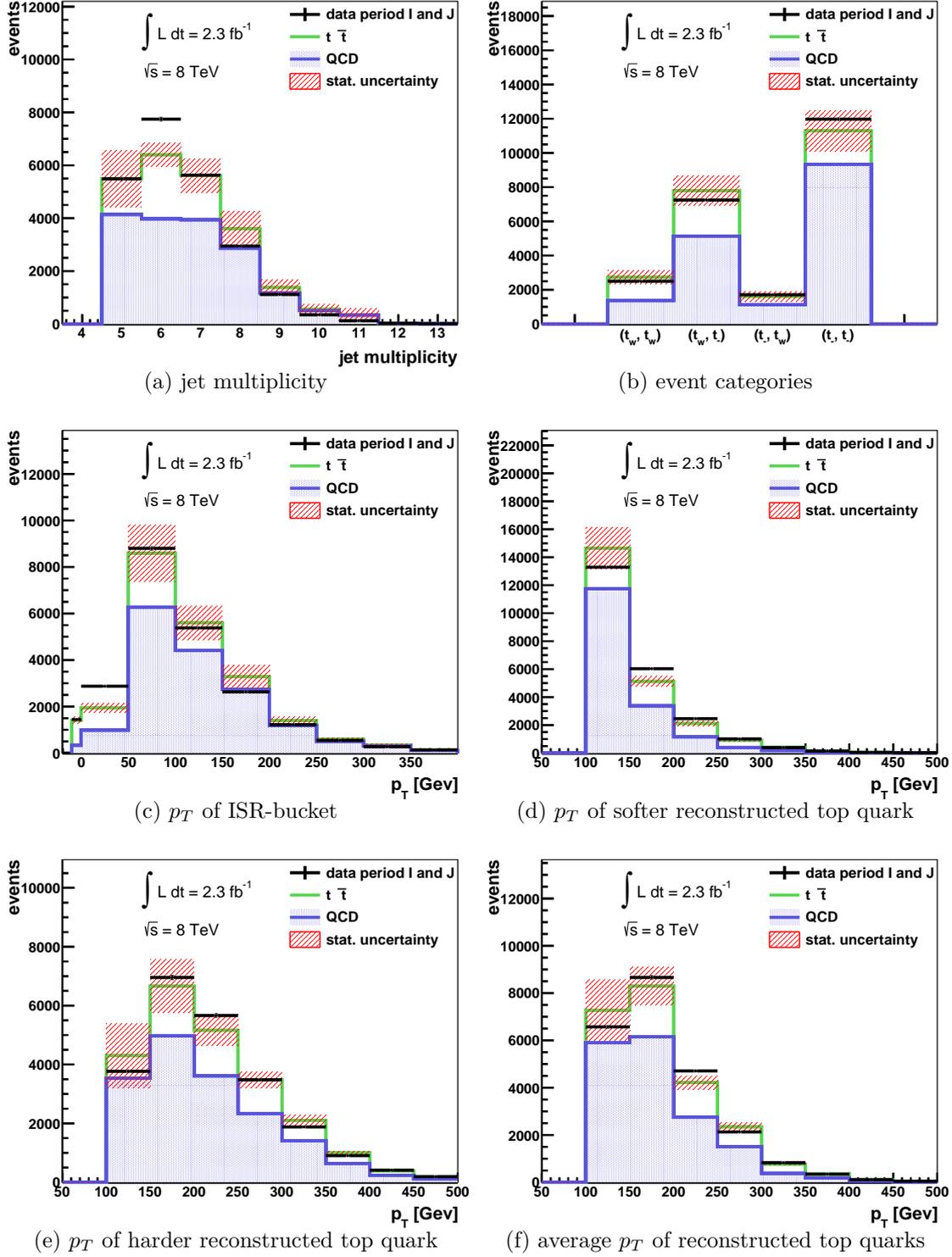


Figure 6.3: Data to Monte-Carlo comparison plots after bucket selection (mass windows,  $p_T^{\text{reco}} > 100 \text{ GeV}$ ,  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ ) in the signal region. The QCD MC is normalized according to  $n_{\text{QCD}}^{\text{MC}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ . The QCD MC is modeled by a QCD dijet sample. The SM  $t\bar{t}$  is modeled by a MC sample of the fully hadronic decay mode and a MC sample of the leptonic decay modes.

and  $\mathbf{t}_-$  buckets should be testable. The study of this grouping is of particular interest because the assumption that sometimes only the  $b$ -jet and the harder jet from the  $W$  decay are selected is actually the main motivation behind the improved part of the bucket algorithm where  $\mathbf{t}_-$  buckets are constructed by the minimization of  $\Delta_B^{bj}$  as defined in Equation 4.0.7. Furthermore, a more detailed model of the background could help defining additional cuts to separate signal from background which is difficult by solely looking at the QCD Monte-Carlo with its bad numerical precision. Therefore, a QCD template is constructed from data in a background dominated region. The SM  $t\bar{t}$  Monte-Carlo plus the QCD template are compared to data in the signal region.

### 6.3.1 Construction and validation of background template

As already mentioned, the background region needs to contain two bottom-jets due to the simultaneous reconstruction of both top quarks in the event. As the buckets are seeded around the  $b$ -tagged jets a different description of the kinematics of the  $b$ -jets in the background region compared to the  $b$ -jets in the signal region will potentially affect the distribution of the buckets as well.

To reduce the correlation between the kinematics of the  $b$ -tagged jet in the region used to construct the background template and the kinematics of the buckets, a region orthogonal to the signal region is constructed by demanding different  $b$ -tagging working points (WP) as shown in Table 6.2. The two constructed regions are labeled *background control region* and *signal region* in the following. The *signal region* corresponds to the  $b$ -jet selection as used in the previous chapters. The *background control region* is chosen in a way of a looser  $b$ -tag WP which has also a worse rejection for light flavor jets guaranteeing that the *background control region* stays highly dominated by QCD multijet events even after the bucket selection cuts, see Table 6.3. However, the *background control region* contains some  $t\bar{t}$  events which are subtracted for the template used in the *signal region*. The normalization of the template in the *signal region* is given by the difference between SM  $t\bar{t}$  and data  $n_{\text{bkg}}^{\text{template}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$  where  $n_{\text{data}} = 23426$  and  $n_{t\bar{t}}^{\text{MC}} = 6478.6$  for the *signal region*, see Table 6.3.

In Figure 6.4 event observables and variables in the *background control region* are shown. The QCD multijet and SM  $t\bar{t}$  are described by Monte-Carlo. The QCD Monte-Carlo is normalized according to  $n_{\text{QCD}}^{\text{MC}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ . The Monte-Carlo prediction is highly dominated by QCD multijet events. The SM  $t\bar{t}$  purity is only  $\approx 4\%$ , see Table 6.3. The good agreement justifies the usage of the QCD template for the *signal region*. The QCD template consists of the data in the *background control region* whereof the only small ( $\approx 4\%$ ) SM  $t\bar{t}$  Monte-Carlo distributions are subtracted.

That the QCD template appropriately describes the background in the *signal region* is further investigated by looking at distributions in the *signal region* before

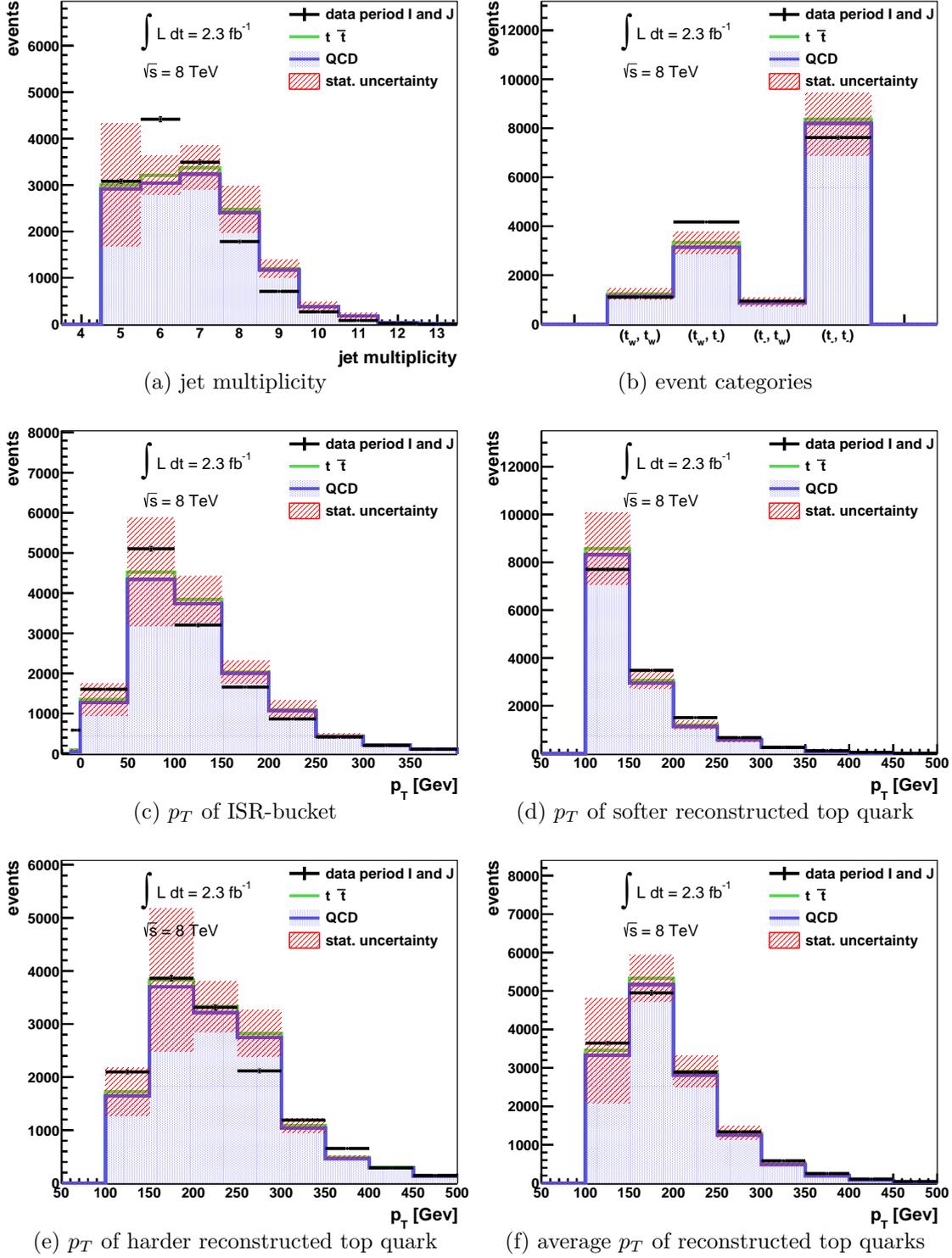


Figure 6.4: Data to Monte-Carlo comparison plots after bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV,  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ ) in the *background control region*. The QCD MC is normalized according to  $n_{\text{QCD}}^{\text{MC}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ . The QCD MC is modeled by a QCD dijet sample. The SM  $t\bar{t}$  is modeled by a MC sample of the fully hadronic decay mode and a MC sample of the leptonic decay modes.

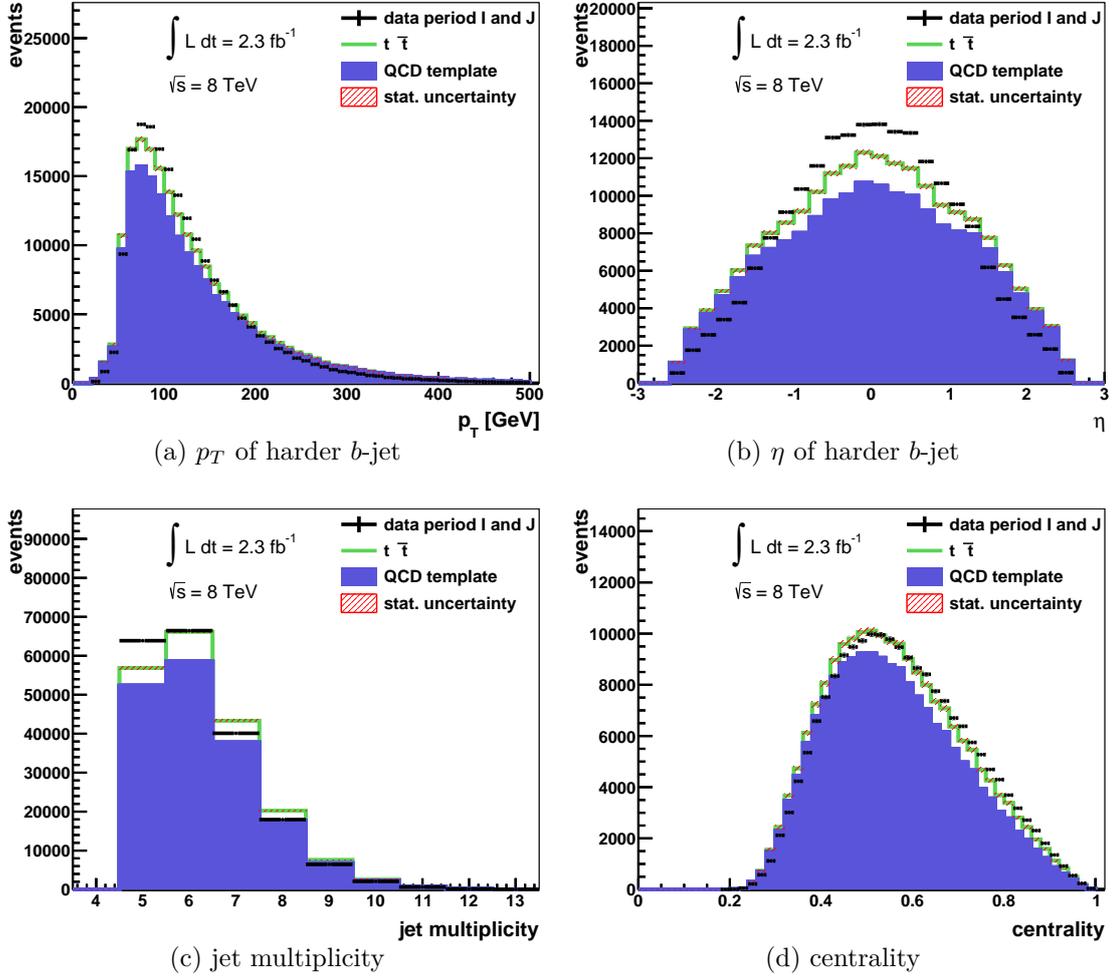


Figure 6.5: Data to Monte-Carlo plus background model comparison plots after baseline selection (at least five jets, two of them  $b$ -tagged), i.e. before running the bucket algorithm, in *signal region*. The QCD template is normalized according to  $n_{\text{QCD}}^{\text{template}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ .

region	MV1 $b$ -tag working points	
	<i>loose</i> 80%	<i>medium</i> 70%
<i>background control region</i>	✓	✗
<i>signal region</i>	✗	✓

Table 6.2: Illustration of the definition of the background region used to construct the background template and the signal region. The different  $b$ -tagging WPs correspond to the present  $b$ -jet efficiency as described in Chapter 3.4.6. The *medium* WP is the default one used in the performance study of the bucket algorithm. The *loose* WP has a higher  $b$ -jet efficiency and reduced rejection against light flavor jets.

region	data yield	$t\bar{t}$ MC yield	$S/B$	$t\bar{t}$ purity
<i>background control region</i>	13844	487.3	0.04	0.04
<i>signal region</i>	23426	6478.6	0.38	0.28

Table 6.3: Data yields and SM  $t\bar{t}$  MC yields after bucket selection corresponding to  $\int Ldt = 2.3 \text{ fb}^{-1}$ . The ratio of Standard Model  $t\bar{t}$  over the background and SM  $t\bar{t}$  purity in data after bucket selection. The values are calculated assuming correct normalization of SM  $t\bar{t}$ . Therefore, the background is normalized to the difference to data.

applying the bucket algorithm. In Figure 6.5 some event observables and variables describing the  $b$ -jets are shown after the basic selection in the *signal region* region. The centrality is defined as

$$\text{centrality} = \frac{\sum_{\text{jets}} p_T}{\sum_{\text{jets}} |p|} \quad (6.3.1)$$

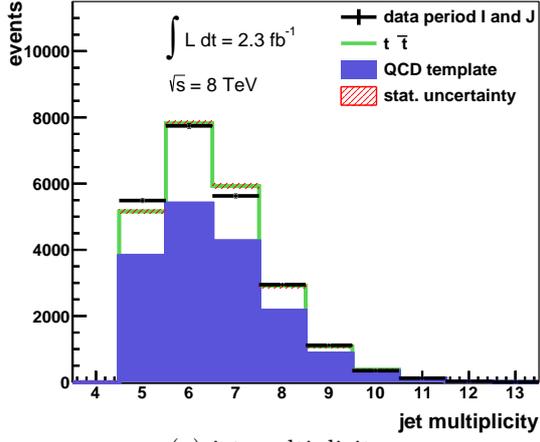
where  $|p|$  is the three-dimensional magnitude. It is a function between zero and one and describes the fraction of the total momentum in the event carried by the  $p_T$  of the jets. At this stage of the selection the SM  $t\bar{t}$  purity is only 8% compared to the SM  $t\bar{t}$  purity of 28% after the bucket selection given in Table 6.3. Therefore, every deviation is likely caused by the background template. Central  $b$ -jets (see Figure 6.5(b)) and  $b$ -jets with a  $p_T$  around 100 GeV (see Figure 6.5(a)) are underestimated. The differences arise because the bottom tagging is also  $p_T$  and  $\eta$  dependent. Despite the slight deviations of the jet multiplicity (see Figure 6.5(c)) and centrality (see Figure 6.5(d)) the background template should still allow to test the grouping of the buckets.

### 6.3.2 Data to Monte-Carlo plus background model comparison in the signal region

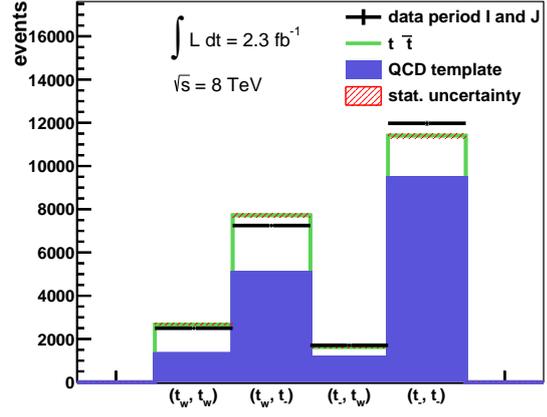
Several event observables and  $p_T$  distributions of the reconstructed top quarks after the bucket selection are shown in Figure 6.6 for data,  $t\bar{t}$  Monte-Carlo and the QCD template. The model slightly underestimates data in the *signal region* for events with less than six jets, see Figure 6.6(a). It should be mentioned that the jet multiplicity distributions differ for the four different bucket event categories. The five-jet bin for example can not contain the  $(\mathbf{t}_w, \mathbf{t}_w)$  category which requires at least six jets. Equally, higher jet multiplicities will more likely contain  $\mathbf{t}_w$  buckets as can be seen by the distribution of the four categories in Figure 6.6(b). The categories containing a  $\mathbf{t}_w$  like  $(\mathbf{t}_w, \mathbf{t}_-)$  are slightly overestimated whereas the  $(\mathbf{t}_-, \mathbf{t}_-)$  category is underestimated. The  $(\mathbf{t}_-, \mathbf{t}_-)$  category is more likely composed of five-jet events compared to categories containing a  $\mathbf{t}_w$  bucket. The distribution of the event categories serves as a good test of the bucket algorithm because each category contains a different  $S/B$  ratio indicating that the agreement is not just due to a wrong normalization of the SM  $t\bar{t}$  MC. The correct description of the grouping emphasizes an appropriate simulation of non-reconstruction of all three top quark decay jets and a correct modeling of the efficiency of building  $\mathbf{t}_-$  buckets.

An additional way to check the correct  $\mathbf{t}_- - \mathbf{t}_w$  grouping is to look into the mass distribution of the harder reconstructed top quark or the softer reconstructed top quark as shown in Figure 6.7(a) and Figure 6.7(b). The events in the high mass range  $m_B > 155$  GeV contain  $\mathbf{t}_w$  buckets while the events  $m_B < 155$  GeV contain  $\mathbf{t}_-$  buckets. Again, the grouping and the mass distribution is correctly described. The high mass range has the higher signal to background ratio resulting from the additional background suppression from the W boson constraint and a well modeled peak at the top quark mass. As expected for the low mass range the peak at 145 GeV is broader than the mass peak around the top quark mass. The low mass peak is also more pronounced for the signal than for the background. Therefore, the mass peak at 145 GeV is expected to be less distinctive in this more background dominated mass range. The same applies for the mass distribution of the softer bucket. Overall the uncertainties for the low mass range do not allow a strong statement about the quality of the shape of the distribution.

The kinematics of the reconstructed top quarks are represented by the  $p_T$  distributions of the harder and softer reconstructed top quark as well as the average  $p_T$  of both reconstructed top quarks in Figure 6.6(d), Figure 6.6(e) and Figure 6.6(f). Generally, in all three cases the events with high  $p_T$  are overestimated. This overestimation starts at  $p_T > 250$  GeV for the harder bucket,  $p_T > 200$  GeV for the softer bucket and around  $p_T > 250$  GeV for the average  $p_T$  of the buckets. This disagreement could be due to some mismodeling in SM  $t\bar{t}$  or the different selection induced by the  $b$ -tagging WP of the *background control region*. The lack of  $b$ -jets within  $p_T \approx 100 - 175$  GeV already at the basic selection, see Figure 6.5(a), could



(a) jet multiplicity



(b) event categories

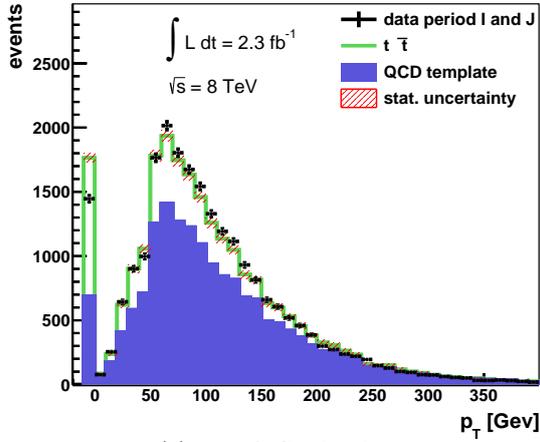
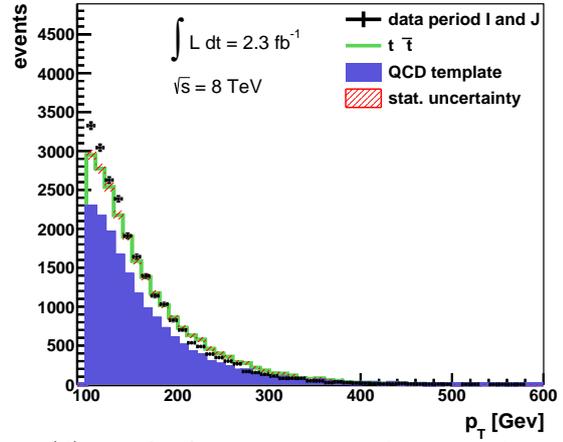
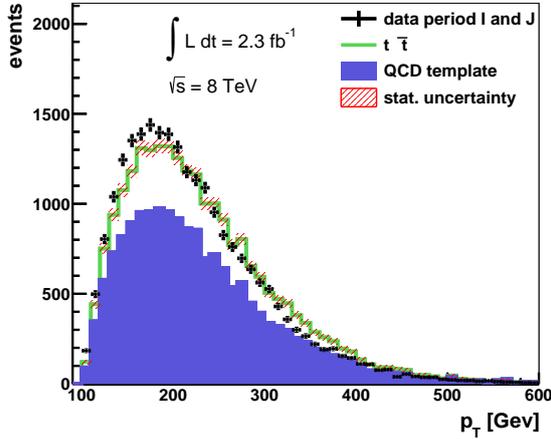
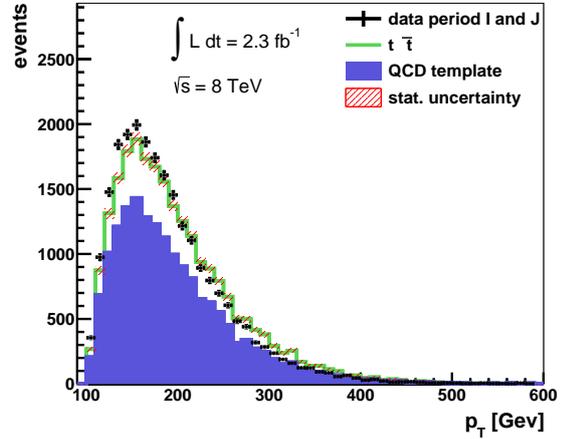
(c)  $p_T$  of ISR-bucket(d)  $p_T$  of softer reconstructed top quark(e)  $p_T$  of harder reconstructed top quark(f) average  $p_T$  of reconstructed top quarks

Figure 6.6: Data to Monte-Carlo plus background model comparison plots after bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV,  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ ) in the *signal region*. The QCD template is normalized according to  $n_{\text{QCD}}^{\text{template}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ .

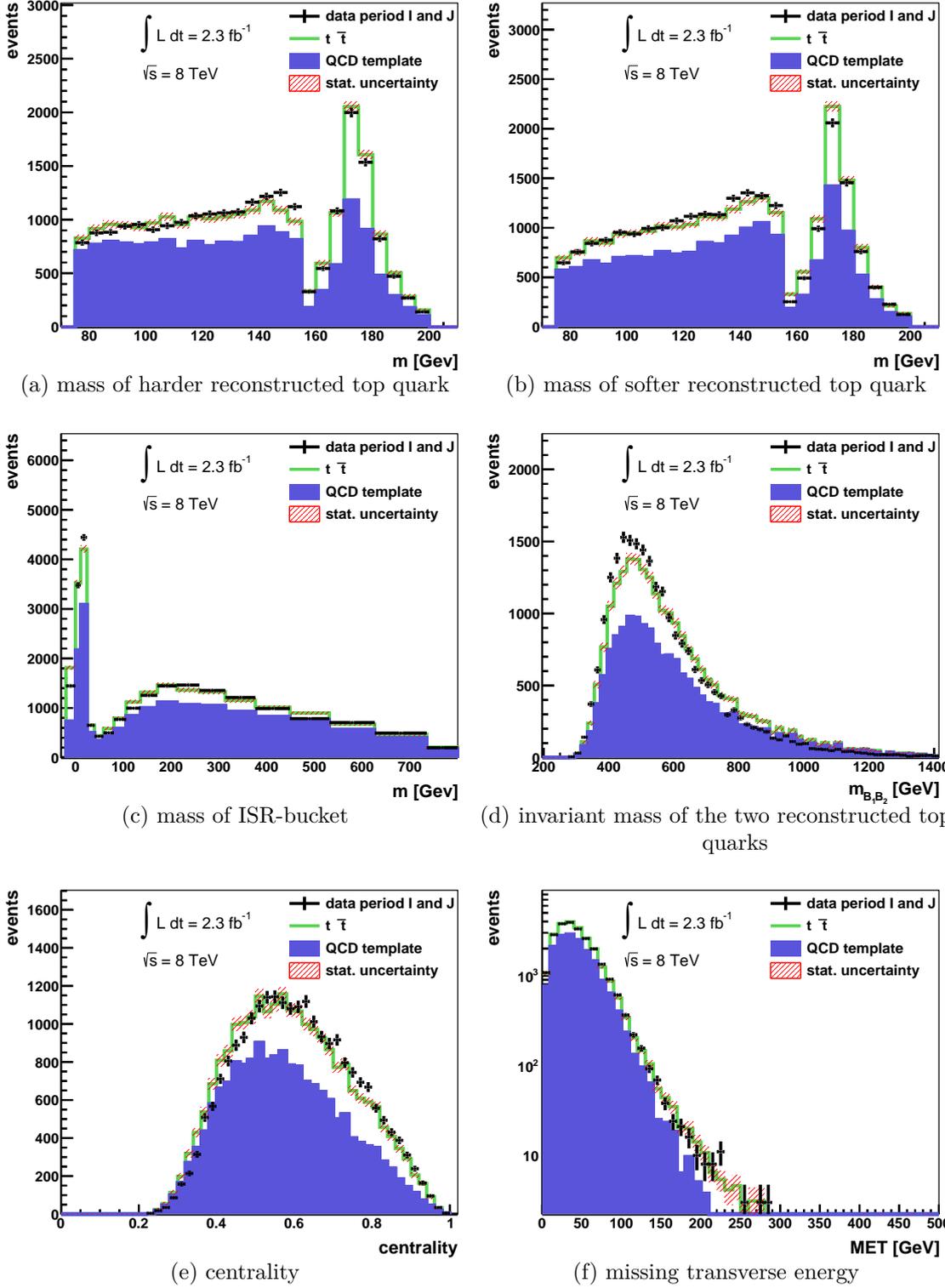


Figure 6.7: Data to Monte-Carlo plus background model comparison plots after bucket selection (mass windows,  $p_T^{\text{reco}} > 100$  GeV,  $\Delta R(\vec{B}_1, \vec{B}_2) > 2.4$ ) in the *signal region*. The QCD template is normalized according to  $n_{\text{QCD}}^{\text{template}} = n_{\text{data}} - n_{t\bar{t}}^{\text{MC}}$ .

as well explain the discrepancy after the bucket selection. Buckets with relative low  $p_T$  consisting for example of two jets whereof one is a  $b$ -jet with a  $p_T$  around 100 GeV will follow more or less the  $b$ -jet distribution. This scenario is quite likely as the  $b$ -jet distribution peaks for the harder  $b$ -jet around  $p_T \approx 100$  GeV.

That the background estimate from the used template approach is not perfect can be seen by looking at the centrality distribution in Figure 6.7(e). Here the data in the *signal region* has more central events compared to the signal simulation plus background model. For low centralities ( $< 0.4$ ) the background alone starts to exceed the data. This excess indicates that either the background normalization is wrong or the *signal region* contains more central  $b$ -jets as the *background control region* resulting in more central total events.

A well modeled event observable is the missing transverse energy  $E_T^{\text{miss}}$  as shown in Figure 6.7(f). This quantity is particularly important for searches for supersymmetry as discussed in more detail in the next chapter.

In addition to these global event variables the ISR-bucket allows to test the description of the rest of the event besides the two top quarks as illustrated in Figure 6.6(c) and Figure 6.7(c). Here a negative entry corresponds to an empty ISR-bucket. The slight overestimation of the amount of empty buckets reflects again the overestimation of the categories containing a  $t_w$  bucket. In these categories the top buckets are composed of more jets leaving less jets for the ISR-bucket. The description of the  $p_T$  distribution of the ISR-bucket in Figure 6.6(c) is good and shows no strong underestimation of soft buckets. Again this agreement is probably due to the fact that the  $b$ -jets with their deviating kinematics sit in the top buckets and not in the ISR-bucket. A similar adequate description is achieved for the mass of the ISR-bucket in Figure 6.7(c). The peak at low masses is produced by ISR-buckets containing just one jet.

The agreement between data and the signal Monte Carlo simulation plus the background model is tested for various kinematic and event variables. Furthermore, reconstructed quantities obtained by the grouping of the jets into buckets are examined. Within the uncertainties and under the consideration of the rather simple model of the  $t\bar{t}$  and background normalization, almost all distributions are adequately described. Of course it is difficult to judge whether a deviation of a bucket quantity is caused by the insufficient background model or the  $t\bar{t}$  simulation. It should be emphasized again that an approach in a physics analysis would be different and needed a more detailed study of the correlation between different  $b$ -tagging WP regions, the associated influence on the basic selection of jets used to feed the bucket algorithm or even required another ansatz making use of e.g. sidebands depending on the analysis.

## 7 Study of scalar top partner decays

The applicability of the bucket algorithm in new physics scenarios is illustrated in this chapter. A direct production of a scalar top pair and a subsequent decay to top quarks and neutralinos with 100% branching ratio, i.e.

$$pp \rightarrow \tilde{t}\tilde{t}^* + X \rightarrow t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0, \quad (7.0.1)$$

is suitable to show the applicability. A direct application of the bucket algorithm is possible as the signature of the final state is mainly two top quark pairs which decay with a BR of around 46% fully hadronically and additionally characterized by high missing transverse energy due to the neutralinos.

Furthermore, searches for scalar partners of the top quark in supersymmetry models at the TeV scale are motivated by the potential of these models to solve the hierarchy problem. The observation of a small Higgs mass makes the hierarchy problem an even stronger question. For solving the hierarchy problem stops with masses below 1 TeV are favored. The allowed range of stop masses is already constrained to  $m_{\text{stop}} > 660$  GeV assuming negligible neutralino masses [79], see Figure 7.1. Higher stop masses will likely result in higher boosts for the top quarks. Therefore, it is possible to use the previous selection of the bucket algorithm and additionally requiring  $E_T^{\text{miss}} > 200$  GeV. This  $E_T^{\text{miss}}$  selection would make it also easier to trigger on the events compared to those described in Chapter 6. After applying the  $E_T^{\text{miss}}$  selection cut the dominant background will be SM  $t\bar{t}$  in the leptonic decay channel of the top quark as the neutrinos present in this decay channel produce missing transverse energy. For simplicity the proof of concept analysis presented here only considers this background. As the bucket algorithm uses looser basic selection cuts compared to existing searches in the zero-lepton channel [79] it is expected that the bucket algorithm will find more signal events with a good significance. Even though an easy way to get rid of leptonic  $t\bar{t}$  and to increase the significance would be a higher  $E_T^{\text{miss}}$  cut, the value  $E_T^{\text{miss}} > 200$  GeV is fixed in this analysis partially due to limited sample sizes and in order to compare the results to existing analyses. Furthermore, it highlights the possibilities to reduce backgrounds making use of the reconstructed four-momentum of the top quarks.

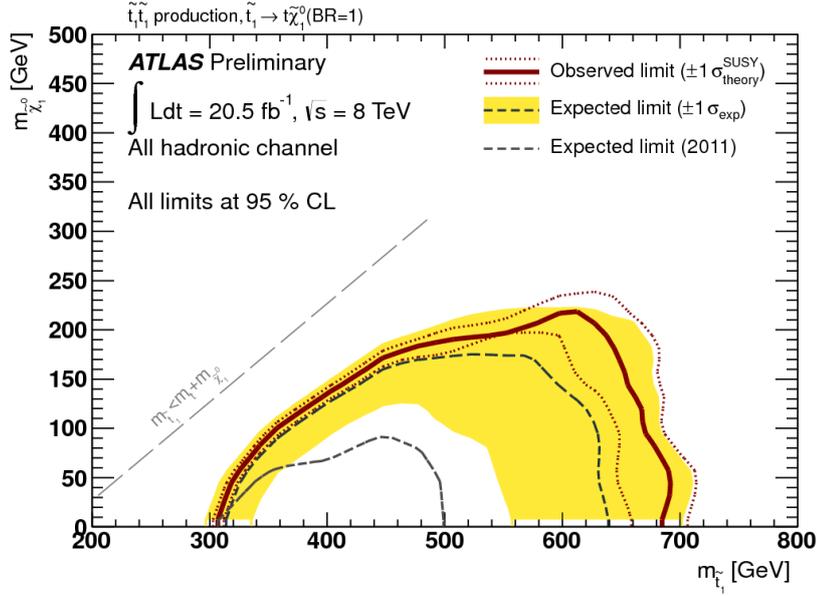


Figure 7.1: Expected and observed exclusion limits at 95% CL for the model of  $pp \rightarrow t\bar{t} \rightarrow t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$  with 100% branching ratio of  $\tilde{t} \rightarrow t\tilde{\chi}_1^0$  [79].

## 7.1 MT2

As mentioned before the fully hadronic decay channel for the top quark pairs from stop decays suffers from a large background contribution of leptonically decaying SM  $t\bar{t}$  events. If it is not possible to suppress this background efficiently by requiring high  $E_T^{\text{miss}}$ , an additional observable helping to discriminate the signal from this background is needed. The need of further discriminant variables becomes especially important if the considered signal sample contains a neutralino with relative high mass which generally results in a reduction of  $E_T^{\text{miss}}$  compared to a scenario with the same stop mass but smaller neutralino mass.

One way to relate  $E_T^{\text{miss}}$  to a physics process involving one invisible particle is the transverse mass  $m_T$ . It can be used for example to describe a leptonic decay of the  $W$  boson  $W \rightarrow l\nu$ .

$$m_T^2 = m_l^2 + m_\nu^2 + 2(E_T^l E_T^\nu - \mathbf{p}_T^l \mathbf{p}_T^\nu) \quad (7.1.1)$$

$$\text{with } E_T^i = \sqrt{m_i^2 + |\mathbf{p}_T^i|^2} \quad (7.1.2)$$

where  $m_i$  is the mass of the particle and  $\mathbf{p}_T^i$  the momentum vector in the transverse plane. The  $m_T$  distribution has an upper endpoint given by  $m_W$  because

$$m_W^2 \geq m_T^2 \quad (7.1.3)$$

and equality only if the lepton and neutrino have the same rapidity. For the case

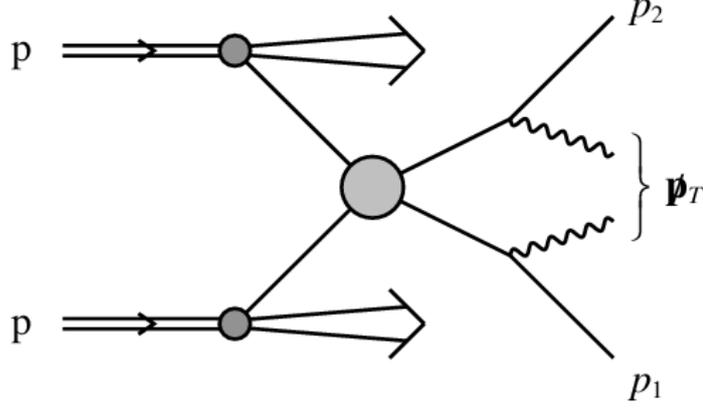


Figure 7.2: Illustration of the considered process. The measurable particle momenta of the top quark pair would correspond to  $p_1$  and  $p_2$ . The not directly observed momenta of the neutralinos (wavy lines) can only be inferred from the missing transverse momentum vector  $\cancel{p}_T$ , from [80].

of direct stop pair production and decay to  $t\tilde{\chi}_1^0$  with 100% BR it is not possible to relate the  $E_T^{\text{miss}}$  to one neutralino. The resulting partition of the missing transverse momentum is illustrated in Figure 7.2 where the momentum  $p_1, p_2$  would correspond to the visible four-momenta of the two top quarks and  $\cancel{p}_T \sim E_T^{\text{miss}}$  to the combined missing transverse momentum vector of the neutralinos.

The two challenges in this scenario are that the mass of the invisible particle and the partition of  $E_T^{\text{miss}}$  into  $E_{T_1}^{\text{miss}}$  and  $E_{T_2}^{\text{miss}}$  is not known. The solution proposed in Reference [80] to overcome this problem is to calculate the function  $M_{T2}(m_\chi)$  for every event

$$M_{T2} = \min_{\mathbf{p}_T^1 + \mathbf{p}_T^2 = \cancel{\mathbf{p}}_T} \left[ \max \left\{ m_T^2(\mathbf{p}_T^1, \mathbf{p}_T^1, m_\chi), m_T^2(\mathbf{p}_T^2, \mathbf{p}_T^2, m_\chi) \right\} \right] \quad (7.1.4)$$

If the invisible particle mass is known, this quantity has an endpoint at the mass of the scalar top partner. Additionally,  $M_{T2}$  shows a threshold at the invisible particle mass. To simplify the analysis for different stop configurations it is assumed that  $m_\chi = 0$  GeV for the minimization of  $M_{T2}$ . For the calculation of  $M_{T2}$  the bisection method as presented in Reference [81] is used. As already indicated in Reference [80],  $M_{T2}$  is only useful for physics processes where the high mass tail is strongly populated which turns out to be the case for the considered process if the four-momenta of the top quarks are well reconstructed. Even though  $M_{T2}$  is not necessarily the best quantity to separate signal from SM  $t\bar{t}$  background, it has the advantage that it shows how the reconstruction of the top quark four-momentum can improve an analysis.

From the construction of  $M_{T2}$ , designed to measure particle masses, it is not

evident that it can be used as a “selection variable” at all. The reasons of the background suppression power of  $M_{T2}$  are known [82] and involve that  $M_{T2}$  is shifted to smaller values for small  $E_T^{\text{miss}}$  or back-to-back QCD-like events.  $M_{T2}$  is also smaller when  $E_T^{\text{miss}}$  is parallel to one of the visible particles. As this parallelism is more likely to happen for fake  $E_T^{\text{miss}}$  produced through detector inefficiencies or resolution it also tends to suppress experimentally bad events where the  $E_T^{\text{miss}}$  is mismeasured.

## 7.2 Monte-Carlo samples

Monte-Carlo simulated samples are used for the description of the background and the SUSY signal. The main backgrounds to this search arise from SM  $t\bar{t}$  production where one top quark decays leptonically. Leptons especially taus, can be misidentified as jets. In turn this misidentification leads to a higher jet multiplicity and therefore an enhancement of events passing the basic selection cuts. Other possible background processes which are not considered here could be QCD multijets,  $Z$  Boson production in association with light-flavor jets, Boson production in association with light- and heavy-flavor jets, single top quark,  $t\bar{t} + W$ ,  $t\bar{t} + WW$  and diboson production. As the basic selection and the requirement for  $E_T^{\text{miss}}$  should highly suppress these processes it is sufficient for the purpose of this study to merely investigate the dominant contribution from SM  $t\bar{t}$ . The leptonic decay modes of the top quark pair are simulated like the fully hadronic mode described in Chapter 3.3.

For the generation of the signal samples HERWIG++ [83] is used. The mixing matrices for the stop and for the neutralinos are chosen such that the  $\tilde{t}$  is mostly the partner of the right-handed top quark. Signal cross sections are calculated to next-to-leading order in the strong coupling constant, adding the resummation of soft gluon emission at next-to-leading-logarithmic accuracy (NLO+NLL) [84, 85, 86]. An envelope of cross section predictions using different PDF sets and factorization and renormalization scales is used to calculate the nominal cross section and the uncertainty as described in Reference [87].

## 7.3 Signal Selection and Comparison

The primary event selection is the same as described in section 3.4. The  $E_T^{\text{miss}}$  is constructed out of electrons, jets, muons and cells outside reconstructed objects as described in Chapter 3.4.9. As in this chapter a pure simulation study is presented no data quality requirements or triggers are used. Nevertheless it would be relatively easy to trigger on the signal events by requiring a missing transverse energy trigger. The efficiency of the lowest unrescaled  $E_T^{\text{miss}}$ -trigger as a function of the offline  $E_T^{\text{miss}}$  for a signal sample with  $m_{\tilde{t}} = 600$  GeV and  $m_{\tilde{\chi}} = 1$  GeV is shown in Figure 7.3. The trigger corresponds to  $E_T^{\text{miss}} > 80$  GeV at the event filter level and

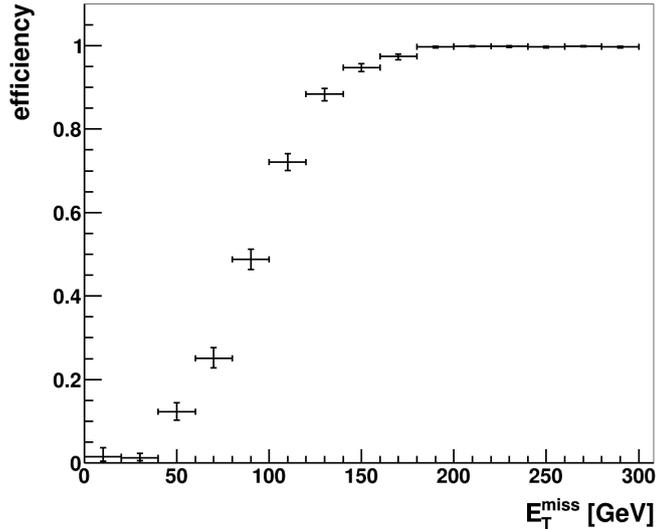


Figure 7.3: Trigger efficiency after bucket selection for the signal sample  $m_{\tilde{t}} = 600$  GeV and  $m_{\tilde{\chi}} = 1$  GeV as a function of offline  $E_T^{\text{miss}}$ . The trigger used requires  $E_T^{\text{miss}} > 80$  GeV at the event filter level.

is described in more detail in Chapter 3.4.10.

To investigate the performance of the bucket algorithm in a supersymmetry search a stop sample with a stop mass  $m_{\tilde{t}} = 600$  GeV and a nearly massless neutralino  $m_{\tilde{\chi}} = 1$  GeV is chosen. It allows to directly compare the bucket algorithm to existing searches [79]. The bucket selection performance achieved with samples of higher stop mass and a nearly massless neutralino is comparable to the  $m_{\tilde{t}} = 600$  GeV scenario. As the production rate for direct stop pairs reduces with increasing stop mass, the advantage of the bucket algorithm with a high signal efficiency should become apparent for these higher stop masses.

This study investigates samples with small neutralino masses. A heavy neutralino generally leads to a reduced boost of the top quarks making the application of the bucket algorithm less feasible. Of course the situation changes with higher center of mass energies.

The main background, SM  $t\bar{t}$  in the semileptonic channel, has only four hadronic decay products at parton level resulting in a falling distribution of the jet multiplicity for higher values. Further jets can arise from e.g. initial state radiation or leptons predominantly taus misidentified as jets. The signal with its hadronically decaying top quark pair has a distribution with higher multiplicity of jets. The different jet multiplicity distributions of signal and background illustrate directly the advantages and disadvantages of an application of buckets in this stop scenario. Compared to a resolved analysis which usually requires at least six relatively hard jets, the bucket algorithm allows to loosen the selection requirement. As a consequence better signal efficiencies for the overall sample even in  $p_T^{\text{top}}$  regimes of the top quark where the bucket algorithm is not highly efficient can be achieved. But

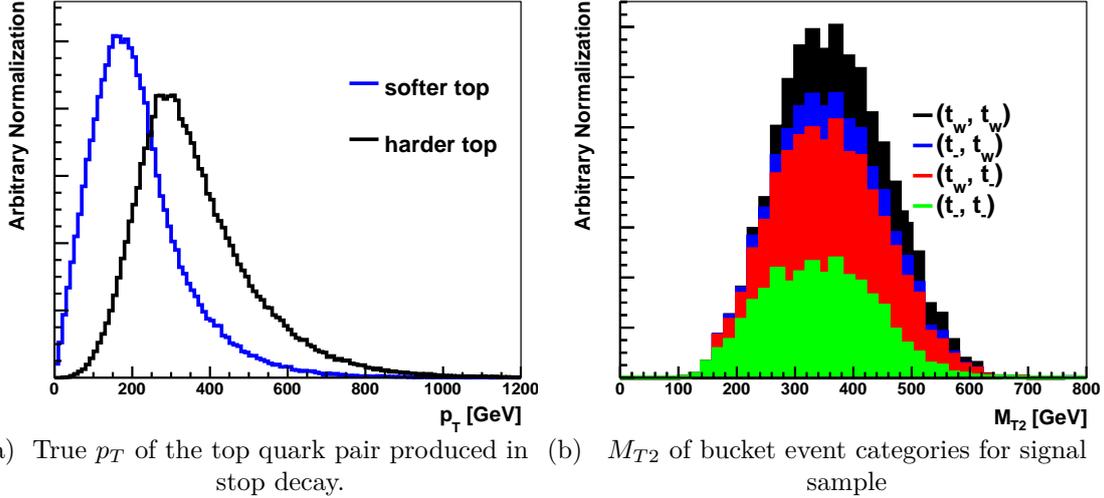


Figure 7.4: Kinematic distributions for a stop MC sample with  $m_{\tilde{t}} = 600$  GeV and  $m_{\tilde{\chi}} = 1$  GeV.

the higher signal efficiency comes with the cost of more selected background events as well.

A major indication for the expected performance of the bucket algorithm is the transverse momentum distribution of the two top quarks from the stop decays as presented in Figure 7.4(a) for the signal sample. A large fraction ( $\approx 50\%$ ) of events where the softest top quark is below 200 GeV indicates that a purely substructure based approach is not appropriate. The  $p_T$  of the harder top quark is in the optimal regime of the bucket algorithm making it a promising test sample.

In order to make the study comparable it is tried to mimic some discrimination variables also used in [79]. For simplicity a signal region corresponding to  $E_T^{\text{miss}} > 200$  GeV is analyzed even though a higher  $E_T^{\text{miss}}$  selection could achieve better significance. The transverse mass  $m_T(b\text{-jet}, E_T^{\text{miss}})$  between the  $E_T^{\text{miss}}$  vector and the  $b$ -tagged jet closest to the  $E_T^{\text{miss}}$  is only slightly correlated with  $E_T^{\text{miss}}$ . Therefore, the transverse mass  $m_T(b\text{-jet}, E_T^{\text{miss}})$  is an additional handle to suppress  $t\bar{t}$  background. The transverse mass is defined by

$$m_T(b, E_T^{\text{miss}}) = \sqrt{2p_T^b E_T^{\text{miss}} (1 - \cos(\Delta\phi(\vec{p}_T^b, \vec{p}_T^{\text{miss}})))} \quad (7.3.1)$$

For the semileptonic SM  $t\bar{t}$  background it is expected that  $m_T(b, E_T^{\text{miss}})$  does not strongly exceed the top quark mass. The corresponding  $m_T(b, E_T^{\text{miss}})$  distribution is shown in Figure 7.5(c). In contrast,  $M_{T2}$  is relatively strong correlated with  $E_T^{\text{miss}}$  for the signal. The correlation is responsible for the sharp turnon in the  $M_{T2}$  distributions as shown in Figure 7.5(b). This strong correlation is the reason why the requirement on  $M_{T2}$  with  $M_{T2} > 300$  GeV is chosen quite conservatively in terms of rejection. As already mentioned the upper edge of the  $M_{T2}$  distribution

can allow for a mass measurement of the scalar top partner if the neutralino mass is known. This mass measurement needs a highly populated upper tail of the  $M_{T2}$  distribution. The  $M_{T2}$  distribution for the four different bucket categories is shown in Figure 7.4(b). As investigated in Chapter 5 the  $\mathbf{t}_w$  buckets generally achieve a better resolution for the reconstruction of the four momentum of the top quark. This different performance of the top quark four momentum reconstruction of the four event categories is also reflected in the  $M_{T2}$  distribution. The imperfect reconstruction for the  $p_T$  shifts the  $(\mathbf{t}_-, \mathbf{t}_-)$  category towards smaller  $M_{T2}$  whereas the  $(\mathbf{t}_w, \mathbf{t}_w)$  category populates higher values. A correction of the four momentum of the  $\mathbf{t}_-$  buckets as done in Chapter 5.2 could cure this behavior of the  $M_{T2}$  distribution. But as such a correction would also affect the background distribution in a similar way it was not applied here. To reduce the influence from events where one tau lepton is reconstructed as a jet, the geometric distance in the  $\phi$ -direction between all non  $b$ -tagged jets and the  $E_T^{\text{miss}}$  vector is required to be greater than  $0.2\pi$ . The  $\Delta\phi(\text{jet}, E_T^{\text{miss}})$  distribution is shown in Figure 7.5(d). The additional selection applied after the bucket selection is summarized in the following

- $E_T^{\text{miss}} > 200 \text{ GeV}$
- $M_{T2} > 300 \text{ GeV}$
- $m_T(b\text{-jet}, E_T^{\text{miss}}) > 175 \text{ GeV}$
- $\Delta\phi(\text{jet}, E_T^{\text{miss}}) > 0.2\pi$

After applying the above signal selection the bucket algorithm would still leave enough room for further optimizations, e.g. the boost of the hard top quark present in the signal would allow to impose even stronger  $p_T$  requirements on the reconstructed top quark. Bad statistical precision in the background sample limits the detailed investigation of those.

The expected Monte-Carlo yields corresponding to  $\int Ldt = 20.5 \text{ fb}^{-1}$  for the signal and background are given in Table 7.1. The doubletag efficiency relative to the baseline selection is around 41% and 14% for signal and leptonic SM  $t\bar{t}$  respectively. This quantifies the rejection power of the bucket algorithm against background events. Nevertheless the much higher production rate of the background requires to further exploit additional variables which contribute the largest to the background suppression. The impact of the  $E_T^{\text{miss}}$  and transverse mass variables also illustrates that in terms of background suppression the precise top quark reconstruction only plays a minor role. All additionally considered discriminant variables are shown in Figure 7.5. As already mentioned, the boost of the top quarks is not significantly increased by higher stop masses. Hence, the signal selection efficiency is roughly the same for other stop samples. The application of the additionally considered discriminant variables would result in a  $4.6\sigma$  significance for  $m_{\tilde{t}} = 600 \text{ GeV}$  and  $m_{\tilde{\chi}} = 1 \text{ GeV}$ . Furthermore, the used discriminant variables

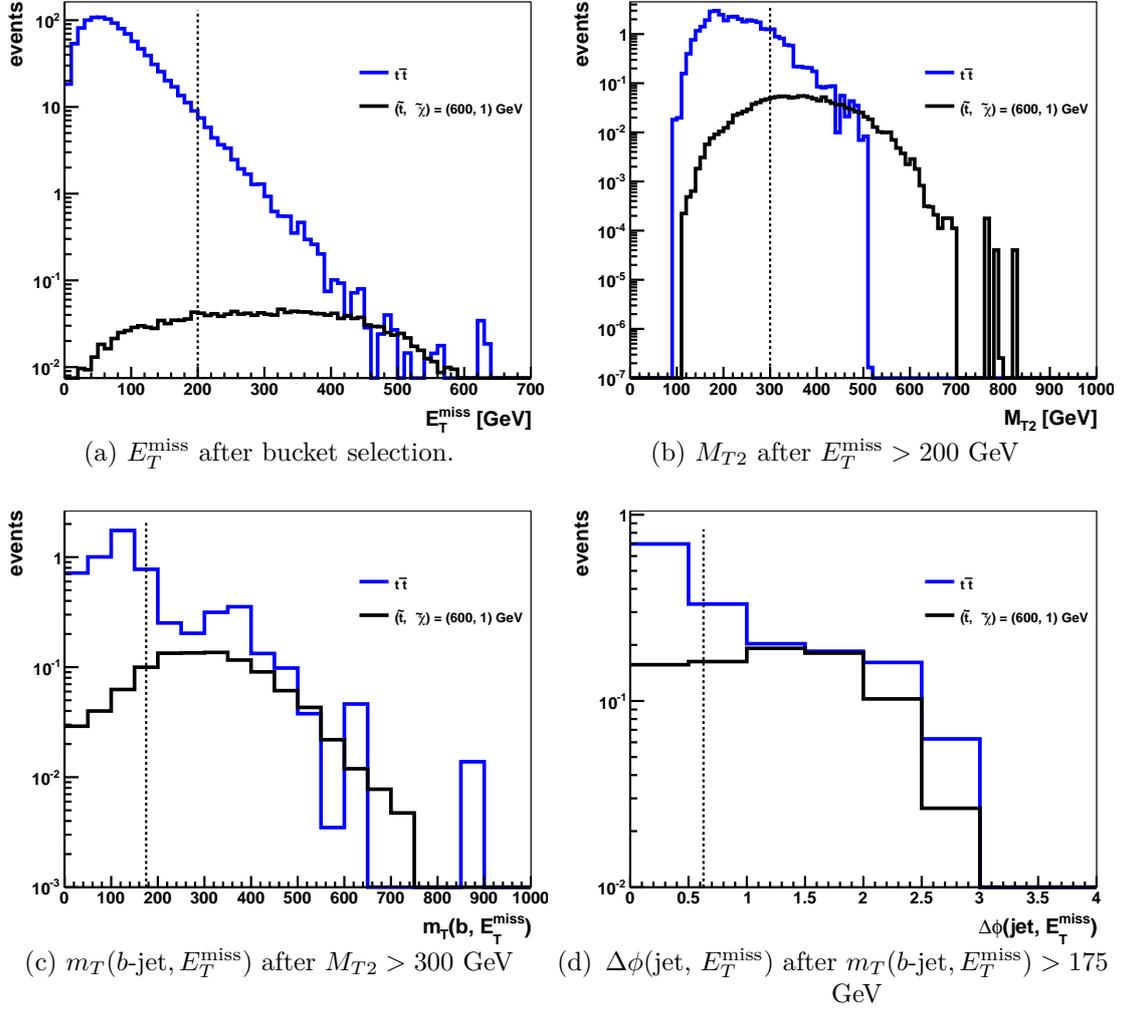


Figure 7.5: Kinematic distributions for SM  $t\bar{t}$  (blue) and signal (black). The SM  $t\bar{t}$  is filtered to contain the leptonic decay modes of the top quark pairs. The signal is shown for a sample with  $m_{\tilde{t}} = 600$  GeV and  $m_{\tilde{\chi}} = 1$  GeV. The yields are normalized to  $\int L dt = 1 \text{ fb}^{-1}$ . All additional discriminant variables used after bucket selection are shown.

	SM $t\bar{t}$ LeptonFilter yield	$\epsilon_r$ [%]	$\epsilon_a$ [%]	$\tilde{t}\tilde{t}^*$ yield	$\epsilon_r$ [%]	$\epsilon_a$ [%]
No selection	2814905.3	–	100.0	508.4	–	100.0
Lepton veto	1554869.1	55.2	55.2	358.1	70.4	70.4
Jet multiplicity and $p_T$	489817.2	31.5	17.4	271.7	75.9	53.4
$\geq 2$ $b$ -tagged jets	220139.2	44.9	7.8	115.0	42.3	22.6
Bucket selection	29885.9	13.6	1.1	47.2	41.1	9.3
$E_T^{\text{miss}} > 200$ GeV	1017.0	3.4	$36.1 \times 10^{-3}$	34.8	73.7	6.8
$M_{T2} > 300$ GeV	151.6	14.9	$5.4 \times 10^{-3}$	25.9	74.5	5.1
$m_T(b\text{-jet}, E_T^{\text{miss}}) > 175$ GeV	43.6	28.8	$1.6 \times 10^{-3}$	21.4	82.4	4.2
Tau veto	12.2	27.9	$0.4 \times 10^{-3}$	16.2	75.9	3.2

Table 7.1: Monte-Carlo yields for  $\int L dt = 20.5 \text{ fb}^{-1}$ , relative ( $\epsilon_r$ ) and absolute ( $\epsilon_a$ ) selection efficiency for each cut step. The signal is represented by a  $pp \rightarrow \tilde{t}\tilde{t}^*$  sample with  $m_{\tilde{t}} = 600$  GeV and  $m_{\tilde{\chi}} = 1$  GeV. The SM  $t\bar{t}$  sample comprises the leptonic decay channels of the top quark pair.

reveal that the bucket algorithm is able to achieve a significance similar to previous studies in the phase space  $E_T^{\text{miss}} > 200$  GeV. Given the same  $E_T^{\text{miss}}$  requirement, the background rate for SM  $t\bar{t}$  in Reference [79] with about 9.8 expected events and the corresponding expected signal events with about 10.3 result in a significance of approximately  $3.3\sigma$ . Of course this comparison only holds for the quoted  $E_T^{\text{miss}}$  selection. Knowing that a future higher center of mass energy of the LHC will lead to more boost of the top quarks an even better performance of the bucket algorithm is expected.

## 8 Discussion and outlook

In this thesis a new method to reconstruct pairs of hadronically decaying top quarks as presented originally in Reference [9] is validated with ATLAS specific Monte-Carlo simulations and LHC data from the 2012 run at a center of mass energy of 8 TeV.

The method targets a moderate transverse momentum regime  $p_T^{\text{top}} = 100 - 400$  GeV of the top quarks with the intend to complement existing reconstruction techniques like substructure based methods or resolved reconstruction. For this purpose conventional ATLAS Anti- $k_T$  ( $R = 0.4$ ) jets are used. They are build out of topological clusters. Jet area based pile-up subtraction and the consideration of the jet vertex fraction aim to reduce the influence of pile-up.

First, tests of the method using Monte-Carlo simulation to study the efficiency of signal selection and background rejection are performed. The signal efficiency for selecting two buckets agrees with the previous “proof-of-concept” study [9] taking into account the reduced resolution present in the detector. The bucket algorithm is most efficient around  $p_T^{\text{top}} \approx 400$  GeV and the efficiency reduces towards higher transverse momentum of the top quarks similar to resolved analysis. At smaller transverse momenta  $p_T^{\text{top}} = 100 - 200$  GeV the bucket algorithm still provides a significant selection efficiency. But the reconstruction of the top quark four momenta is more accurate for higher boosts of the top quarks. An application of the bucket algorithm to top quarks with transverse momentum above the targeted momentum regime as present in e.g. heavy  $Z' \rightarrow t\bar{t}$  scenarios is possible. It is shown that the bucket algorithm is well capable of reconstructing invariant masses of particles decaying to top quark pairs. The accuracy of the reconstruction of the top quark four momenta can be improved by additional cuts on reconstructed quantities.

Due to the usage of relatively small and pile-up corrected jets the reconstruction performance is only weakly pile-up dependent. This weak pile-up dependence makes the bucket algorithm a promising new tool even at higher center of mass energies as will be reached in the upcoming LHC runs.

It was demonstrated that background from QCD multijet production is suppressed by cuts on the mass and transverse momentum of the reconstructed top quarks.

The bucket algorithm requires generally looser preselection cuts than a resolved analysis and is computationally relatively fast because it just makes use of the invariant mass of the assigned jets. Because of its simplicity it is in principal possible to combine the bucket algorithm with a likelihood based analysis.

A data to Monte-Carlo comparison was performed. The Monte-Carlo samples of SM  $t\bar{t}$  and QCD dijet events correctly describe kinematic variables of the reconstructed top quarks, global event observables as well as the grouping into the four possible bucket event categories. In order to overcome the shortcomings of the Monte-Carlo simulation of QCD multijets, a data driven background template is constructed to estimate the background. Under the assumption that the normalization of the SM  $t\bar{t}$  is correctly described in data, good agreement between data and the SM  $t\bar{t}$  Monte-Carlo plus background model is achieved for distributions describing kinematic variables of the buckets, global event observables and the grouping of the buckets into the event categories. This agreement first of all confirms that the basic idea of the bucket algorithm with the reconstruction using only reduced information of the two harder jets from the top quark decay is well justified and adequately modeled in Monte-Carlo simulation.

It is shown that with the bucket algorithm a better signal efficiency compared to previous searches for supersymmetry involving the decay of scalar top partner pairs could be achieved with a similar significance in the phase space  $E_T^{\text{miss}} > 200$  GeV. The accurate reconstruction of the top quark four momenta allows to calculate the  $M_{T2}$  variable resulting in an improved background rejection and giving the possibility to measure the stop mass.

In this thesis no systematic uncertainties are considered for the bucket algorithm. They could arise from e.g. jet energy scale (JES) or  $b$ -tagging uncertainties. The important observable in the reconstruction of the top quark pair is the invariant mass of the jets assigned to each bucket. This mass observable will certainly show some dependence on the JES. Uncertainties on the  $b$ -tagging efficiency especially occurring at higher transverse momenta of the  $b$ -jets could affect the performance of the reconstruction and influence the  $t\bar{t}$  yield as presented in the data to Monte-Carlo comparison.

Furthermore, bottom quarks decay quite often semileptonically to muons [88]. They can decay directly to muons or through cascade decays via charm and tau. The energy of the muon or the corresponding neutrino can be lost in the calorimeters leading to an underestimation of the  $b$ -jet energy. A mis-measurement of the four momentum of the  $b$ -tagged jet will affect the performance of the bucket algorithm as in each top bucket the  $b$ -jet four momentum contributes to the calculation of the metric used to assign the jets to the three buckets. Available corrections of semileptonically decaying  $b$ -jets [89] should enable to further improve the reconstruction accuracy of the top quark four momentum.

As suggested in Reference [90] the bucket algorithm could also be applied in searches for Higgs boson production in association with top quarks. The process  $pp \rightarrow t\bar{t}H(bb)$  is characterized by a high multiplicity of jets and four  $b$  quarks. It also requires the reconstruction of top quarks with relatively small transverse momentum to achieve a good efficiency. The application of the bucket algorithm

for the top quark reconstruction could help selecting the  $b$  quarks that correspond to the Higgs boson.

# Bibliography

- [1] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012. [arXiv:1207.7214](#), [doi:10.1016/j.physletb.2012.08.020](#).
- [2] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012. [arXiv:1207.7235](#), [doi:10.1016/j.physletb.2012.08.021](#).
- [3] Johannes Erdmann, Stefan Guindon, Kevin Kroeninger, Boris Lemmer, Olaf Nackenhorst, et al. A likelihood-based reconstruction algorithm for top-quark pairs and the KLFilter framework. *Nucl.Instrum.Meth.*, A748:18–25, 2014. [arXiv:1312.5595](#), [doi:10.1016/j.nima.2014.02.029](#).
- [4] Michael H. Seymour. Searches for new particles using cone and cluster jet algorithms: A Comparative study. *Z.Phys.*, C62:127–138, 1994. [doi:10.1007/BF01559532](#).
- [5] Tilman Plehn and Michael Spannowsky. Top Tagging. *J.Phys.*, G39:083001, 2012. [arXiv:1112.4441](#), [doi:10.1088/0954-3899/39/8/083001](#).
- [6] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. *Eur.Phys.J.*, C71:1661, 2011. [arXiv:1012.5412](#), [doi:10.1140/epjc/s10052-011-1661-y](#).
- [7] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, et al. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. *J.Phys.*, G39:063001, 2012. [arXiv:1201.0008](#), [doi:10.1088/0954-3899/39/6/063001](#).
- [8] Sebastian Schaetzel. Boosted Top Quarks and Jet Structure. 2014. [arXiv:1403.5176](#).
- [9] Matthew R. Buckley, Tilman Plehn, and Michihisa Takeuchi. Buckets of Tops. *JHEP*, 1308:086, 2013. [arXiv:1302.6238](#), [doi:10.1007/JHEP08\(2013\)086](#).
- [10] Tilman Plehn. Lectures on LHC Physics. *Lect.Notes Phys.*, 844:1–193, 2012. [arXiv:0910.4182](#), [doi:10.1007/978-3-642-24040-9](#).

- [11] J. Beringer et al. Review of Particle Physics (RPP). *Phys.Rev.*, D86:010001, 2012. doi:10.1103/PhysRevD.86.010001.
- [12] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2–3):31 – 145, 1983. URL: <http://www.sciencedirect.com/science/article/pii/0370157383900807>, doi:[http://dx.doi.org/10.1016/0370-1573\(83\)90080-7](http://dx.doi.org/10.1016/0370-1573(83)90080-7).
- [13] John M. Campbell, J.W. Huston, and W.J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept.Prog.Phys.*, 70:89, 2007. arXiv:hep-ph/0611148, doi:10.1088/0034-4885/70/1/R02.
- [14] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 0804:005, 2008. arXiv:0802.1188, doi:10.1088/1126-6708/2008/04/005.
- [15] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012. arXiv:1111.6097, doi:10.1140/epjc/s10052-012-1896-2.
- [16] Matteo Cacciari and Gavin P. Salam. Dispelling the  $N^3$  myth for the  $k_t$  jet-finder. *Phys.Lett.*, B641:57–61, 2006. arXiv:hep-ph/0512210, doi:10.1016/j.physletb.2006.08.037.
- [17] Pile-up subtraction and suppression for jets in ATLAS. Technical Report ATLAS-CONF-2013-083, CERN, Geneva, Aug 2013.
- [18] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009. arXiv:0901.0512.
- [19] Particle Data Group. 2013 partial update of [11] for the 2014 edition.
- [20] Frank-Peter Schilling. Top Quark Physics at the LHC: A Review of the First Two Years. *Int.J.Mod.Phys.*, A27:1230016, 2012. arXiv:1206.4484, doi:10.1142/S0217751X12300165.
- [21] Fabian Kohn. Measurement of the charge asymmetry in top quark pair production in pp collision data at  $\sqrt{s} = 7$  TeV using the ATLAS detector. 2012. arXiv:1204.0952.
- [22] Stephen P. Martin. A Supersymmetry primer. 1997. arXiv:hep-ph/9709356.
- [23] Search for supersymmetry using razor variables in events with b-jets in pp collisions at 8 TeV. Technical Report CMS-PAS-SUS-13-004, CERN, Geneva, 2013.

- [24] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004.
- [25] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*. LHC Tech. Proposal. CERN, Geneva, 1994.
- [26] *Technical proposal*. LHC Tech. Proposal. CERN, Geneva, 1994. Cover title : CMS, the Compact Muon Solenoid : technical proposal.
- [27] *LHCb : Technical Proposal*. Tech. Proposal. CERN, Geneva, 1998.
- [28] *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC Tech. Proposal. CERN, Geneva, 1995.
- [29] Peter Braun-Munzinger and Johanna Stachel. The quest for the quark-gluon plasma. *Nature*, 448:302–309, 2007. doi:10.1038/nature06080.
- [30] Philippe Mouche. Overall view of the LHC. Vue d’ensemble du LHC. Sep 2006.
- [31] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008.
- [32] ATLAS Collaboration. Public luminosity results, April 2014. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [33] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi:10.1088/1748-0221/3/08/S08003.
- [34] ATLAS Experiment.
- [35] *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999. Electronic version not available.
- [36] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Hoche, et al. General-purpose event generators for LHC physics. *Phys.Rept.*, 504:145–233, 2011. arXiv:1101.2599, doi:10.1016/j.physrep.2011.03.005.
- [37] G. Aad et al. The ATLAS Simulation Infrastructure. *Eur.Phys.J.*, C70:823–874, 2010. arXiv:1005.4568, doi:10.1140/epjc/s10052-010-1429-9.
- [38] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002. arXiv:hep-ph/0204244, doi:10.1088/1126-6708/2002/06/029.

- [39] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 0308:007, 2003. arXiv: hep-ph/0305252, doi:10.1088/1126-6708/2003/08/007.
- [40] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010. arXiv:1007.2241, doi:10.1103/PhysRevD.82.074024.
- [41] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6.5 release note. 2002. arXiv:hep-ph/0210213.
- [42] J.M. Butterworth, Jeffrey R. Forshaw, and M.H. Seymour. Multiparton interactions in photoproduction at HERA. *Z.Phys.*, C72:637–646, 1996. arXiv: hep-ph/9601371, doi:10.1007/s002880050286.
- [43] New ATLAS event generator tunes to 2010 data. Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011.
- [44] ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.
- [45] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys.Lett.*, B710:612–622, 2012. arXiv:1111.5869, doi:10.1016/j.physletb.2012.03.013.
- [46] Peter Baernreuther, Michal Czakon, and Alexander Mitov. Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to  $q\bar{q} \rightarrow t\bar{t} + X$ . *Phys.Rev.Lett.*, 109:132001, 2012. arXiv:1204.5201, doi:10.1103/PhysRevLett.109.132001.
- [47] Michal Czakon and Alexander Mitov. NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels. *JHEP*, 1212:054, 2012. arXiv:1207.0236, doi:10.1007/JHEP12(2012)054.
- [48] Michal Czakon and Alexander Mitov. NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction. *JHEP*, 1301:080, 2013. arXiv: 1210.6832, doi:10.1007/JHEP01(2013)080.
- [49] Michal Czakon, Paul Fiedler, and Alexander Mitov. The total top quark pair production cross-section at hadron colliders through  $\mathcal{O}(\alpha_S^4)$ . 2013. arXiv: 1303.6254.
- [50] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. 2011. arXiv:1112.5675.

- [51] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, et al. The PDF4LHC Working Group Interim Recommendations. 2011. [arXiv:1101.0538](#).
- [52] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009. [arXiv:0901.0002](#), [doi:10.1140/epjc/s10052-009-1072-5](#).
- [53] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Uncertainties on  $\alpha_S$  in global PDF analyses and implications for predicted hadronic cross sections. *Eur.Phys.J.*, C64:653–680, 2009. [arXiv:0905.3531](#), [doi:10.1140/epjc/s10052-009-1164-2](#).
- [54] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, et al. The CT10 NNLO Global Analysis of QCD. 2013. [arXiv:1302.6246](#).
- [55] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. *Nucl.Phys.*, B867:244–289, 2013. [arXiv:1207.1303](#), [doi:10.1016/j.nuclphysb.2012.10.003](#).
- [56] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput.Phys.Commun.*, 182:1034–1046, 2011. [arXiv:1007.1327](#), [doi:10.1016/j.cpc.2010.12.040](#).
- [57] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008. [arXiv:0710.3820](#), [doi:10.1016/j.cpc.2008.01.036](#).
- [58] Summary of ATLAS Pythia 8 tunes. Technical Report ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug 2012.
- [59] Luminosity Determination in pp Collisions at  $\sqrt{s} = 7$  TeV using the ATLAS Detector in 2011. Technical Report ATLAS-CONF-2011-116, CERN, Geneva, Aug 2011.
- [60] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003. [doi:10.1016/S0168-9002\(03\)01368-8](#).
- [61] Expected electron performance in the ATLAS experiment. Technical Report ATL-PHYS-PUB-2011-006, CERN, Geneva, Apr 2011.
- [62] Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data. Technical Report ATLAS-CONF-2013-088, CERN, Geneva, Aug 2013.

- [63] Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. Technical Report ATLAS-CONF-2012-042, CERN, Geneva, Mar 2012.
- [64] Performance of primary vertex reconstruction in proton-proton collisions at  $\sqrt{s} = 7$  TeV in the ATLAS experiment. Technical Report ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010.
- [65] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.
- [66] T Barillari, E Bergeaas Kuutmann, T Carli, J Erdmann, P Giovannini, K J Grahn, C Issever, A Jantsch, A Kiryunin, K Lohwasser, A Maslennikov, S Menke, H Oberlack, G Pospelov, E Rauter, P Schacht, F Spanó, P Speckmayer, P Stavina, and P Strízenec. Local Hadronic Calibration. Technical Report ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun 2008. Due to a report-number conflict with another document, the report-number ATL-LARG-PUB-2009-001-2 has been assigned.
- [67] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.
- [68] Jet energy scale and its systematic uncertainty in proton-proton collisions at  $\sqrt{s}=7$  TeV with ATLAS 2011 data. Technical Report ATLAS-CONF-2013-004, CERN, Geneva, Jan 2013.
- [69] D0 Collaboration. Observation of single top quark production, March 2009. URL: [http://www-d0.fnal.gov/Run2Physics/top/singletop\\_observation/](http://www-d0.fnal.gov/Run2Physics/top/singletop_observation/).
- [70] Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.
- [71] Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb<sup>-1</sup> of Data from the ATLAS Detector. Technical Report ATLAS-CONF-2012-043, CERN, Geneva, Mar 2012.
- [72] R. Achenbach, P. Adragna, V. Andrei, P. Apostologlou, B. Asman, et al. The ATLAS level-1 calorimeter trigger. *JINST*, 3:P03001, 2008. doi:10.1088/1748-0221/3/03/P03001.

- [73] I Hristova. The evolution and performance of the ATLAS calorimeter-based triggers in 2011 and 2012. Technical Report ATL-DAQ-PROC-2012-051, CERN, Geneva, Oct 2012.
- [74] V Cavaliere. b-jet triggering in ATLAS. Technical Report ATL-DAQ-PROC-2012-038, CERN, Geneva, Jun 2012.
- [75] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. Stop Reconstruction with Tagged Tops. *JHEP*, 1010:078, 2010. arXiv:1006.2833, doi:10.1007/JHEP10(2010)078.
- [76] Tilman Plehn and Michihisa Takeuchi. W+Jets at CDF: Evidence for Top Quarks. *J.Phys.*, G38:095006, 2011. arXiv:1104.4087, doi:10.1088/0954-3899/38/9/095006.
- [77] Tagging and suppression of pileup jets with the ATLAS detector. Technical Report ATLAS-CONF-2014-018, CERN, Geneva, May 2014.
- [78] Georges Aad et al. Search for resonances decaying into top-quark pairs using fully hadronic decays in  $pp$  collisions with ATLAS at  $\sqrt{s} = 7$  TeV. *JHEP*, 1301:116, 2013. arXiv:1211.2202, doi:10.1007/JHEP01(2013)116.
- [79] Search for direct production of the top squark in the all-hadronic  $t\bar{t}$  +  $\text{etmiss}$  final state in 21 fb<sup>-1</sup> of p-p collisions at  $\sqrt{s}=8$  TeV with the ATLAS detector. Technical Report ATLAS-CONF-2013-024, CERN, Geneva, Mar 2013.
- [80] C.G. Lester and D.J. Summers. Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders. *Phys.Lett.*, B463:99–103, 1999. arXiv:hep-ph/9906349, doi:10.1016/S0370-2693(99)00945-4.
- [81] Hsin-Chia Cheng and Zhenyu Han. Minimal Kinematic Constraints and  $m(T_2)$ . *JHEP*, 0812:063, 2008. arXiv:0810.5178, doi:10.1088/1126-6708/2008/12/063.
- [82] Christopher Lester and Alan Barr. MTGEN: Mass scale measurements in pair-production at colliders. *JHEP*, 0712:102, 2007. arXiv:0708.1028, doi:10.1088/1126-6708/2007/12/102.
- [83] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, et al. Herwig++ Physics and Manual. *Eur.Phys.J.*, C58:639–707, 2008. arXiv:0803.0883, doi:10.1140/epjc/s10052-008-0798-9.
- [84] W. Beenakker, M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas. Stop production at hadron colliders. *Nucl. Phys.*, B515:3–14, 1998. arXiv:hep-ph/9710451.

- [85] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, and I. Niessen. Supersymmetric top and bottom squark production at hadron colliders. *JHEP.*, 1008:098, 2010. [arXiv:arXiv:1006.4771 \[hep-ph\]](#).
- [86] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, et al. Squark and gluino hadroproduction. *Int.J.Mod.Phys.*, A26:2637–2664, 2011. [arXiv:1105.1110](#), [doi:10.1142/S0217751X11053560](#).
- [87] Michael Kramer, Anna Kulesza, Robin van der Leeuw, Michelangelo Mangano, Sanjay Padhi, et al. Supersymmetry production cross sections in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. 2012. [arXiv:1206.2892](#).
- [88] P. Urquijo et al. Moments of the electron energy spectrum and partial branching fraction of  $B \rightarrow X_c e \nu$  decays at Belle. *Phys.Rev.*, D75:032001, 2007. [arXiv:hep-ex/0610012](#), [doi:10.1103/PhysRevD.75.032001](#).
- [89] Jet energy measurement and systematic uncertainties using tracks for jets and for b-quark jets produced in proton-proton collisions at  $\sqrt{s} = 7$  TeV in the ATLAS detector. Technical Report ATLAS-CONF-2013-002, CERN, Geneva, Jan 2013.
- [90] Matthew R. Buckley, Tilman Plehn, Torben Schell, and Michihisa Takeuchi. Buckets of Higgs and Tops. *JHEP*, 1402:130, 2014. [arXiv:1310.6034](#), [doi:10.1007/JHEP02\(2014\)130](#).

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den .....