

Fakultät für Physik und Astronomie

Ruprecht-Karls-Universität Heidelberg

Bachelorarbeit

Im Studiengang Physik

vorgelegt von

Dominik Mitzel

geboren in Baden-Baden

2012

Untergrundbestimmung

in der

$W + \gamma + \gamma$ Analyse

Die Bachelorarbeit wurde von Dominik Mitzel

ausgeführt am 13.08.2012

Kirchhoff Institut für Physik

unter der Betreuung von

Herrn Prof. Dr. Hans-Christian Schultz-Coulon

Untergrundbestimmung in der $W + \gamma + \gamma$ Analyse:

Orientiert an der in ATLAS durchgeführten und publizierten $WZ + \gamma$ Analyse [1] wird die Ereignisselektion des $p + p \rightarrow W + \gamma + \gamma + X$ Prozesses im Myonzerfallskanal des W-Bosons vorgestellt. Es wird eine auf Daten basierende Untergrundbestimmung der dominanten $W + jets$ Prozesse und eine auf Monte Carlo Daten basierende Abschätzung weiterer elektroschwacher Untergrundprozesse präsentiert. Dabei wird der Einfluss von alternativen Signalregiondefinitionen der $W + \gamma + \gamma$ Analyse auf die statistische und systematische Unsicherheit von Signal- und Untergrundereignisanzahl mit Hilfe von Monte Carlo Studien untersucht.

Aufgrund des kleinen Wirkungsquerschnitts des $W + \gamma + \gamma$ Prozesses treten bei der Untergrundbestimmung statistische Effekte auf, die erst bei niedrigen Ereigniszahlen relevant werden. Diese Effekte werden in die Bestimmung der statistischen und systematischen Unsicherheit einbezogen.

Die Selektion und Untergrundbestimmung des $pp \rightarrow W\gamma\gamma \rightarrow \mu\nu\gamma\gamma$ Prozesses wird in Proton-Proton Kollisionsdaten durchgeführt, die im Jahr 2011 vom ATLAS Detektor bei einer Schwerpunktsenergie von $\sqrt{s}=7$ TeV aufgezeichnet wurden.

Außerdem wird eine Monte Carlo basierende Prognose für die erwartete Anzahl an Signal- und Untergrundereignissen für das Jahr 2012 gegeben.

Background determination in the $W + \gamma + \gamma$ analysis:

Based on the existing and published $WZ + \gamma$ ATLAS analysis [1], the selection for $p + p \rightarrow W + \gamma + \gamma + X$ events with the W decaying in the muon channel is introduced. In order to estimate the amount of dominant $W + jets$ and other electroweak background events, a data-driven method and a monte-carlo driven method are presented, respectively; taking into account the possibility of different definitions of the signal region and its impact on the statistical and systematic uncertainty.

Due to the small cross section of the $W + \gamma + \gamma$ process, the determination of background events reveals statistical effects only relevant at very low numbers of events. The accomplished estimation of the statistical and systematic error takes these effects into consideration.

The event selection and determination of background in the $pp \rightarrow W\gamma\gamma \rightarrow \mu\nu\gamma\gamma$ process is done for proton-proton collision data recorded by the ATLAS detector in 2011 at a center of mass energy $\sqrt{s}=7$ TeV. Furthermore, a monte-carlo driven prediction for the expected amount of signal and background events in 2012 is given.

Inhaltsverzeichnis

1	Einleitung	8
2	Der $W + \gamma + \gamma$ Prozess	9
2.1	Signaldefinition $W + \gamma + \gamma$	9
2.2	Untergrunddefinition	10
2.3	Verwendeter Datensatz	11
2.4	Signal und Untergrund Monte Carlo Datensätze	12
3	Photonrekonstruktion und -Identifikation im ATLAS Detektor	14
3.1	Der ATLAS Detektor	14
3.2	Photonrekonstruktion	16
3.3	Photonidentifikation	17
4	Definition der Selektion	22
4.1	Vorselektion	22
4.2	Finale Selektion	24
5	Untergrundbestimmung	26
5.1	Die $2 \times 2D$ Sideband Methode	26
5.2	Bestimmung der Monte Carlo Input Parameter	36
5.3	Monte Carlo Closure Test	37
5.4	Signal- und Untergrund- Wahrscheinlichkeitsdichteverteilungen	41
5.5	Einfluss der Monte Carlo Unsicherheit auf die Vorhersage des relativen Fehlers	48
5.6	Studie des systematischen Fehlers	49
5.7	Zusammenfassung der Monte Carlo Studie und Entdeckungswahrscheinlichkeit	51
5.8	Anwendung der $2 \times 2D$ Sideband Methode auf Daten	54
6	Zusammenfassung	59
A	Monte Carlo Input Parameter	61
B	Monte Carlo Studie für alternative Signalregiondefinition	65
C	Alternative Studie auf Daten	67
D	Monte Carlo Vorhersage für 2012	69

E	Zusätzliche Rechnung	71
F	Verzeichnisse	72
F.1	Abbildungsverzeichnis	72
F.2	Tabellenverzeichnis	73
G	Literaturverzeichnis	75
	Danksagung	78

1 Einleitung

Seit dem Jahre 2009 werden auf dem Gelände des CERN¹ in der Nähe von Genf wahlweise Proton Pakete oder Blei Ionen im Large Hadron Collider (LHC) [2] zur Kollision gebracht. Der LHC ist ein 27 km langer Ringbeschleuniger, welcher sich in einem Tunnel in etwa 100 m Tiefe unter der Erdoberfläche befindet. Für die Zukunft sind Proton-Proton Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 14$ TeV geplant, wobei die im Jahre 2011 aufgenommenen Daten bei Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV aufgezeichnet wurden.

Die vier Detektoren ATLAS (**A** **T**oroidal **L**HC **A**pperatu**S**), CMS (**C**ompact **M**uon **S**oleniod), ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) und LHCb (**L**arge **H**adron **C**ollider **b**eauty) wurden um die Kollisionspunkte der gegenläufig beschleunigten Protonen gebaut, um die bei der Wechselwirkung entstehende Teilchen und Teilchen aus Sekundärzerfällen zu detektieren, zu vermessen und somit die Vorgänge im Streuprozess zu verstehen.

Als Vielzweckdetektor konzipiert bietet der ATLAS Detektor [3] die Möglichkeit, eine Vielzahl von verschiedenen Teilchen im Endzustand der Proton-Proton Wechselwirkung zu rekonstruieren und zu unterscheiden. Die Vielseitigkeit des ATLAS Detektors führt zusammen mit der hohen Luminosität des LHC dazu, dass Analysen von Prozesse mit immer kleineren Wirkungsquerschnitten möglich werden, die aufgrund ihrer Seltenheit mit vorherigen Beschleunigerexperimenten nicht möglich waren.

Die $W + \gamma + \gamma$ Analyse versucht den Wirkungsquerschnitt des Prozesses zu bestimmen, in dem ein W-Boson und zwei Photonen aus der Proton-Proton Wechselwirkung hervorgehen. Eine besondere Herausforderung dieses Prozesses liegt in dem sehr geringen Wirkungsquerschnitt. Bei einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV wird ein Wirkungsquerschnitt im Bereich von ca. 12 fb erwartet [4]. Außerdem handelt es sich um den ersten messbaren Prozess, dessen Endzustand sich aus drei Bosonen zusammensetzt.

Der kleine Wirkungsquerschnitt dieses Prozesse führt dazu, dass nur sehr wenig Ereignisse in den bisher aufgezeichneten Daten erwartet werden. Im Rahmen dieser Bachelorarbeit wird die $W + \gamma + \gamma$ Ereignisselektion vorgestellt und implementiert. Da es eine Vielzahl von Untergrundereignissen gibt, welche trotz einer sorgfältigen Selektion fälschlicherweise als $W + \gamma + \gamma$ Ereignisse selektiert werden, wird zudem eine auf Daten basierende Untergrundabschätzung vorgestellt, in Monte Carlo Studien getestet und anschließend auf Daten aus dem Jahr 2011 angewandt.

¹Abkürzung für Europäische Organisation für Kernforschung

2 Der $W + \gamma + \gamma$ Prozess

Eine zentrale Rolle in dieser Bachelorarbeit spielen der $p + p \rightarrow W + \gamma + \gamma + X$ Prozess und die dominanten Untergrundprozesse. In diesem Kapitel wird als erster Schritt eine Signal- und eine Untergrunddefinition eingeführt.

2.1 Signaldefinition $W + \gamma + \gamma$

In der $W + \gamma + \gamma$ Analyse wird nach der Produktion von $W + \gamma + \gamma$ Endzuständen in der Proton-Proton Wechselwirkung gesucht. Als Signalereignis werden somit Prozesse bezeichnet, bei denen sich nach der Proton-Proton Wechselwirkungen ein W-Boson und zwei Photonen im Endzustand befinden, wobei die Photonen nicht aus dem Zerfall farbloser Hadronen stammen. Dem W-Boson stehen mehrere sowohl leptonische als auch hadronische Zerfallskanäle zur Verfügung [5], wobei in der $W + \gamma + \gamma$ Analyse nur die leptonischen Zerfallskanäle $W \rightarrow \mu + \bar{\nu}_\mu$ und $W \rightarrow e + \bar{\nu}_e$ betrachtet werden. In dieser Bachelorarbeit wird in der folgenden Analyse nur der Zerfall im Myonkanal $W \rightarrow \mu + \bar{\nu}_\mu$ betrachtet. Es wird folglich nach Ereignissen gesucht, in denen im Detektor zwei Photonen, ein Myon und ein Myon Antineutrino¹ rekonstruiert wurden. In Abbildung 2.1 sind typischen Feynmangraphen des $W + \gamma + \gamma$ Prozesses dargestellt.

Neben der möglichen Photonenabstrahlung der wechselwirkenden Quarks (Initial State Radiation², ISR) oder des Myons im Endzustand (Final State Radiation, FSR³) ist besonders der $WW\gamma\gamma$ Vertex von Interessen (Abbildung 2.1 unten rechts), da hier eine Kopplung von vier Eichbosonen vorliegt.

Die Teilchenrekonstruktion in einem Detektor basiert auf der Wechselwirkung des Teilchens mit dem Detektormaterial, wobei trotz aufwändiger Algorithmen nur mit einer gewissen Wahrscheinlichkeit auf die wahre Identität des Teilchens und dessen Herkunft geschlossen werden kann. Hieraus folgt die Notwendigkeit der Definition eines Signalphotons und eines Untergrundphotons:

- Als Signalphotonen werden im Folgenden solche rekonstruierten Photonkandidaten bezeichnet, welchen der oben definierte Signalprozess zu Grunde liegt.
- Untergrundphotonen gibt es zwei Arten: Zum einen Photonen aus Zerfällen farbloser Hadronen und zum anderen Energiedepositionen anderer Teilchen,

¹Auf Neutrinos wird nur indirekt über fehlende transversale Energie geschlossen

²Englisch für Abstrahlung im Anfangszustand

³Englisch für Abstrahlung im Endzustand

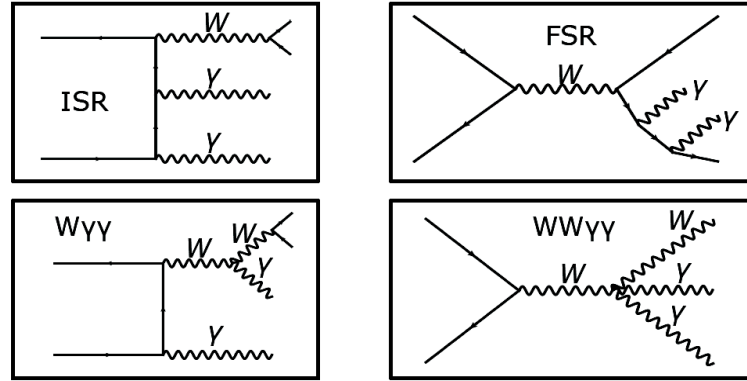


Abbildung 2.1: Typische Feynmangraphen des $W + \gamma + \gamma$ Prozesses [4].

die fälschlicherweise als Photonen rekonstruiert wurden. Untergrundphotonen werden oft als fake⁴ Photonen bezeichnet.

2.2 Untergrunddefinition

Als Untergrund werden alle Ereignisse bezeichnet, in deren Endzustand alle in der $W + \gamma + \gamma$ Analyse gesuchte Teilchen rekonstruiert wurden, obwohl nicht der oben definierte Signalprozess zu Grunde liegt. Alle Untergrundprozesse in der $W + \gamma + \gamma$ Analyse sind reduzibel. Dies bedeutet, dass ein oder mehrere Teilchen fehlrekonstruiert wurden und somit fälschlicherweise auf ein $W + \gamma + \gamma$ Ereignis geschlossen wird. Als häufiges Beispiel dienen Elektronen, die aufgrund ihrer Wechselwirkung mit dem Detektor eine ähnliche Signatur hinterlassen und fälschlicherweise als Photonen rekonstruiert werden können. Um möglichst genaue Aussagen über die Häufigkeit und somit den Wirkungsquerschnitt des gesuchten Prozesses machen zu können, ist es notwendig, eine effiziente Ereignis Selektion durchzuführen und dabei möglichst viele Untergrundereignisse auszuselektieren. Effizient bedeutet in diesem Zusammenhang, dass möglichst viele der Signalereignisse die Selektion überstehen. Prinzipiell lässt sich die Anzahl selektierter reduzibler Untergrundereignisse durch eine Verschärfung der Selektionsschnitte verringern, jedoch muss ein Kompromiss zwischen Untergrundunterdrückung und Effizienz der Selektion gefunden werden. Ziel dieser Arbeit es, eine Untergrundabschätzung für den $W + \gamma + \gamma$ Prozess mit Monte Carlo simulierten Ereignissen zu testen und anschließend auf Daten anzuwenden.

⁴Von engl. to fake deutsch fälschen

2.2.1 Der dominante $W + jets$ Untergrund

Der dominante Untergrund in der $W + \gamma + \gamma$ Analyse setzt sich aus $W + \gamma + jet$ und $W + jet + jet$ Ereignissen zusammen. Hierbei besteht die Möglichkeit, dass jeweils ein oder zwei Jets als Photonen rekonstruiert wurden. Jets, die hochenergetische neutrale Hadronen wie π^0 und η^0 Teilchen enthalten, werden besonders häufig als Photonen rekonstruiert, da sie noch innerhalb des Detektors in zwei Photonen zerfallen. Der Wirkungsquerschnitt dieser Untergrundprozesse [6] liegt um Größenordnungen höher als der erwartete Wirkungsquerschnitt des $W + \gamma + \gamma$ Prozesses, so dass auch bei einer sorgfältigen und gut gewählten Ereignisselektion und trotz einer hohen Unterdrückung von Untergrundereignissen davon auszugehen ist, dass eine nicht vernachlässigbare Anzahl an Untergrundereignissen dieser Art selektiert wird.

2.2.2 Elektroschwacher Untergrund

Aufgrund der elektroschwachen Wechselwirkung zwischen den stoßenden Protonen gibt es neben den $W + \gamma + jet$ und $W + jet + jet$ Ereignissen eine Reihe weiterer Untergrundprozesse, welche im Folgenden als elektroschwacher Untergrund zusammengefasst werden. Beachtet wurden hierbei folgende Prozesse:

- Produktion von $Z + jets$ und $Z + \gamma + jets$ Ereignissen
- Ereignisse mit zwei Bosonen
- Produktion von Ereignissen mit einem Top-Quark oder Top-Antitop-Quark-Paar

Bei diesen Prozessen ist es möglich, dass die rekonstruierten Teilchen nach weiteren Sekundärzerfällen im Endzustand den gesuchten Teilchen in der $W + \gamma + \gamma$ Analyse entsprechen. Ein typisches Beispiel sind $W + Z$ Ereignisse, in denen die Sekundärzerfälle $W \rightarrow \mu\bar{\nu}_\mu$ und $Z \rightarrow e\bar{e}$ möglich sind [5]. In diesem Fall ist es möglich, dass die beiden Elektronen fälschlicherweise als Photonen rekonstruiert werden.

2.3 Verwendeter Datensatz

In dieser Arbeit durchgeführte Analysen mit Daten basieren auf einem Datensatz von Proton-Proton Kollisionseignissen bei einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV, der im ATLAS Experiment im Jahre 2011 aufgenommen wurde. Die integrierte Luminosität beträgt 4644 pb^{-1} mit einer Unsicherheit von 1.8% [7]. Es werden nur Daten berücksichtigt, bei denen alle Subkomponenten des Detektors voll funktionsfähig waren. Die Ereignisse wurde mit Hilfe eines Triggers für Myonen aufgezeichnet.

2.4 Signal und Untergrund Monte Carlo Datensätze

Für die $W + \gamma + \gamma$ Analyse werden verschiedene simulierte Monte Carlo Datensätze benötigt. In Tabelle 2.1 sind die Monte Carlo Samples aufgelistet, in denen Signalprozesse simuliert wurden, Tabelle 2.2 zeigt die Samples des dominanten $W + jets$ Untergrunds und in Tabelle 2.3 sind die Monte Carlo Samples des elektroschwachen Untergrunds aufgelistet. Mögliche Überschneidungen der $W + \gamma + jet$ Untergrundsamples und Signal Samples werden durch ein Veto zweier wahrer Photonen⁵ entfernt.

Prozess	Datenset	Wirkungs- querschnitt (nb)	k-Faktor[4]	Filter
$W + \gamma + \gamma$	12613.WmunugammagammaNp0	$2.18 \cdot 10^{-5}$	1.18	1.0
$W + \gamma + \gamma$	12614.WmunugammagammaNp1	$2.05 \cdot 10^{-5}$	1.18	1.0
$W + \gamma + \gamma$	12615.WmunugammagammaNp2	$1.21 \cdot 10^{-5}$	1.18	1.0

Tabelle 2.1: Die Signal Monte Carlo Samples für die $W + \gamma + \gamma$ Analyse

Prozess	Datenset	Wirkungs- querschnitt (nb)	k-Faktor	Filter
$W + jet$	107690.AlpgeJimmyWmunuNp0	$6921.6 \cdot 10^{-3}$	1.2	1.0
$W + jet$	107691.AlpgeJimmyWmunuNp1	$1304.3 \cdot 10^{-3}$	1.2	1.0
$W + jet$	107692.AlpgeJimmyWmunuNp2	$377.83 \cdot 10^{-3}$	1.2	1.0
$W + jet$	107693.AlpgeJimmyWmunuNp3	$101.88 \cdot 10^{-3}$	1.2	1.0
$W + jet$	107694.AlpgeJimmyWmunuNp4	$25.75 \cdot 10^{-3}$	1.2	1.0
$W + jet$	107695.AlpgeJimmyWmunuNp5	$6.92 \cdot 10^{-3}$	1.2	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$2.13 \cdot 10^{-1}$	1.0	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$5.22 \cdot 10^{-2}$	1.0	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$1.73 \cdot 10^{-2}$	1.0	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$5.33 \cdot 10^{-3}$	1.0	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$1.34 \cdot 10^{-3}$	1.0	1.0
$W + \gamma + jet$	117410.AlpgeJimmyWgammaNp0	$3.38 \cdot 10^{-4}$	1.0	1.0

Tabelle 2.2: Monte Carlo Samples der dominanten $W + jets$ Untergrund Prozesse, jeweils der Zerfall des W-Bosons im Myonkanal.

⁵Bedingung für wahre Photonen: $0 < |ph_truth_mothertype| < 26$

Prozess	Datenset	Wirkungsquerschnitt (nb)	k-Faktor	Filter
$Z + \gamma + jet$	126016.Sherpa_Zy_mu_jet	$1.44 \cdot 10^{-2}$	1	1
$t\bar{t}$	105861.TTbar_PowHeg_Pythia	0.146	1	0.54301
WW	105921.McAtNlo_JIMMY_WpWm_enuenu	$0.50 \cdot 10^{-3}$	1	1
WW	105922.McAtNlo_JIMMY_WpWm_enununu	$0.50 \cdot 10^{-3}$	1	1
WW	105923.McAtNlo_JIMMY_WpWm_enutaunu	$0.50 \cdot 10^{-3}$	1	1
WW	105924.McAtNlo_JIMMY_WpWm_munumunu	$0.50 \cdot 10^{-3}$	1	1
WW	105925.McAtNlo_JIMMY_WpWm_munuenu	$0.50 \cdot 10^{-3}$	1	1
WW	105926.McAtNlo_JIMMY_WpWm_munutaunu	$0.50 \cdot 10^{-3}$	1	1
WW	105927.McAtNlo_JIMMY_WpWm_taututaunu	$0.50 \cdot 10^{-3}$	1	1
WW	105928.McAtNlo_JIMMY_WpWm_tauunuenu	$0.50 \cdot 10^{-3}$	1	1
WW	105929.McAtNlo_JIMMY_WpWm_tauunumunu	$0.50 \cdot 10^{-3}$	1	1
$W \rightarrow \tau + \nu_\tau$	107054.PythiaWtaunu_incl	1.05	1	1
$Z \rightarrow e + \bar{e}$	106046.PythiaZee_no_filter	0.99	1	1
$Z \rightarrow \mu + \bar{\mu}$	106047.PythiaZmumu_no_filter	0.99	1	1
$Z \rightarrow \tau + \bar{\tau}$	106052.PythiaZtautau	0.99	1	1
$t - Quark$	108341.st_tchan_munu_McAtNlo_Jimmy	$6.93 \cdot 10^{-3}$	1	1
$t - Quark$	108342.st_tchan_tauunu_McAtNlo_Jimmy	$6.93 \cdot 10^{-3}$	1	1
$t - Quark$	108344.st_schan_munu_McAtNlo_Jimmy	$0.5 \cdot 10^{-3}$	1	1
$t - Quark$	108345.st_schan_tauunu_McAtNlo_Jimmy	$0.5 \cdot 10^{-3}$	1	1
$t - Quark$	108346.st_Wt_McAtNlo_Jimmy	$15.6 \cdot 10^{-3}$	1	1

Tabelle 2.3: Monte Carlo Samples des elektroschwachen Untergrundes. Die Wahl der betrachteten Prozesse orientiert sich grÖßtenteils an [8].

3 Photonrekonstruktion und -Identifikation im ATLAS Detektor

3.1 Der ATLAS Detektor

Der ATLAS Detektor [3, 9] ermöglicht durch seinen Aufbau die Detektion von den in Kollisionen entstehenden Teilchen über den gesamten Raumwinkel. Um eine effiziente Rekonstruktion und Identifikation der Teilchen zu ermöglichen, sind um den Ort der Wechselwirkung verschiedene Detektorsysteme schichtweise angeordnet, mit deren Hilfe sowohl die Trajektorien¹ als auch die Energiedepositionen der Teilchen rekonstruiert werden können (siehe Abbildung 3.1). Für die Trajektorienrekonstruktion ist der Innere Detektor zuständig, welcher sich aus einem Pixel (engl pixel detector), einem Silizium-Streifen (engl SemiConductor Tracker (SCT)) und einem Übergangsstrahlungsdetektor (engl Transition Radiation Tracker (TRT)) zusammensetzt. Um aus der Teilchentrajektorie auf den Impuls schließen zu können, werden die geladenen Teilchen durch ein 2 T starkes solenoid Magnetfeld abgelenkt.

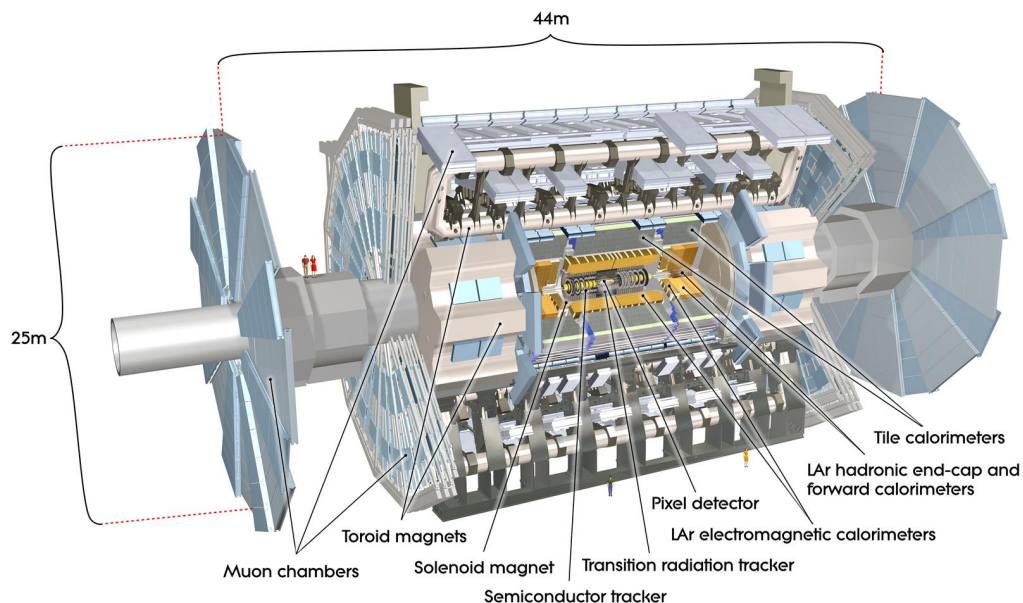


Abbildung 3.1: Der ATLAS Detektor [9].

¹Trajektorienrekonstruktion basiert auf Ionisation und ist daher nur für geladene Teilchen möglich

Der Innere Detektor ermöglicht eine effiziente Rekonstruktion der Teilchentrajektorie bis zu einem Pseudorapiditätsbereich² von $|\eta|=2.5$. Die Messung der Energiedepositionen erfolgt im Kalorimetersystem, welches wiederum aus verschiedenen Subsystemen aufgebaut ist und in einen Pseudorapiditätsbereich von $-4.9 < \eta < 4.9$ Teilchen nachweisen kann. Im ATLAS Detektor werden so genannte Sampling Kalorimeter verwendet, die abwechselnd aus aktivem Material und Absorber bestehen. Das elektromagnetischen Kalorimeter (engl. Electromagnetic Calorimeter) besteht aus einem zylinderförmigen Mittelteil (Bereich $|\eta| < 1.475$), welches aufgrund seiner Form als Barrel (deutsch Fass) bezeichnet wird, sowie den zwei Endkappen die sich in Strahlrichtung anschließen (Bereich $1.375 < |\eta| < 3.2$). Als Absorber und aktives Material dienen hier Blei und flüssiges Argon (siehe Abbildung 3.1). Elektronen und Photonen verlieren ihre gesamte Energie in der Regel durch elektromagnetische Schauer im elektromagnetischen Kalorimeter. Da im Folgenden die Photonrekonstruktion und Identifikation eine zentrale Rolle spielen, wird der Aufbau des elektromagnetischen Kalorimeter genauer vorgestellt:

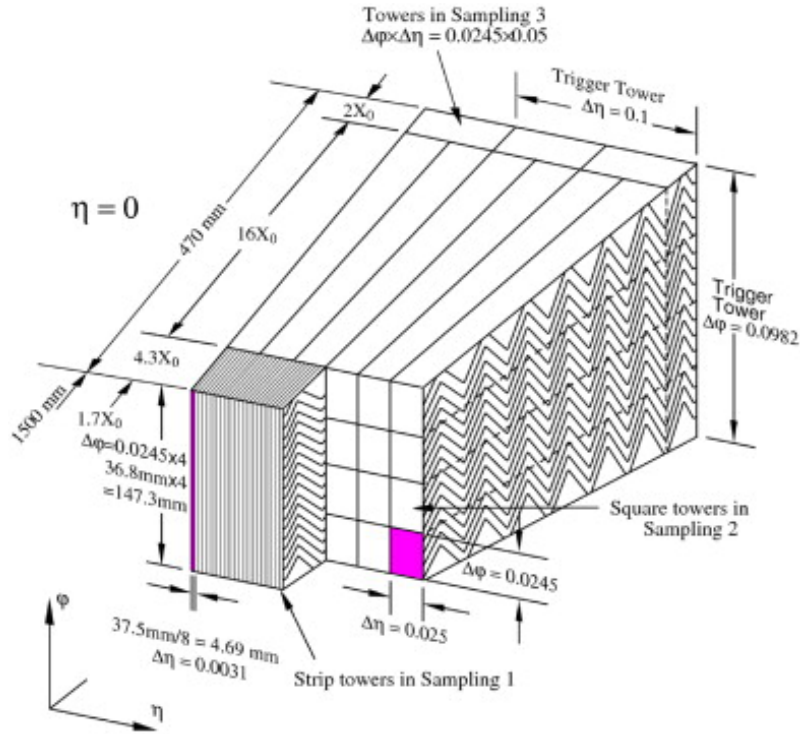


Abbildung 3.2: Schematische Skizze des elektromagnetischen Kalorimeters [3].

²Im ATLAS Experiment wird ein rechtshändiges Koordinatensystem verwendet, dessen Ursprung im nominellen Interaktionspunkt in der Mitte des Detektors liegt. Die z-Achse zeigt in Richtung des Strahlgangs, die x-Achse in Richtung des Zentrums des LHC Rings, und die y-Achse nach oben. Der Azimutalwinkel wird in Bezug auf die x-Achse gemessen. Außerdem wird die Pseudorapidität η über den Polarwinkel θ als $\eta = -\ln \tan \frac{\theta}{2}$ definiert.

In Abbildung 3.2 ist der Aufbau des elektromagnetischen Kalorimeters schematisch skizziert. Es besteht sowohl im $\eta \times \phi$ Raum als auch senkrecht zu dieser Ebene aus feinen Segmenten. Das Kalorimeter gliedert sich in der Ebene senkrecht zum Strahlengang in drei Bereiche, die oft als Front- (oder Strip-³) Middle- und Backlayer⁴ bezeichnet werden [10]. Die Striplayer weist in η Richtung eine besonders feine Granularität auf, so dass die Form der im Kalorimeter auftretenden elektromagnetischen Schauer genau analysiert werden kann.

Das hadronische Kalorimeter besteht aus dem so genannten Tile Kalorimeter und dem Flüssig-Argon-Endkappen-Kalorimeter (englisch Hadronic End-cap Calorimeter, HEC), wobei das Tile Kalorimeter auch wieder aus Mittelteil ($|\eta| < 1.0$) und dem so genannten extended Barrel ($0.8 < |\eta| < 1.7$) besteht. In diesem wechseln sich Stahl als Absorber und aktive Szintillator Kacheln⁵ ab. Das HEC besteht aus Kupferabsorberplatten und flüssigem Argon und deckt den Bereich $1.5 < |\eta| < 3.2$ ab. Im Pseudorapiditätsbereich $3.1 < |\eta| < 4.9$ befindet sich noch das Vorwärtskalorimeter (englisch Forward Calorimeter). Dieses dient hauptsächlich zur Detektion von Jets und besteht aus dem gleichen aktiven Material wie das elektromagnetische Kalorimeter, jedoch aus mehreren verschiedenen Absorbermaterialien.

Da Myonen als minimal ionisierende Teilchen nur einen geringen Bruchteil ihrer Energie im Kalorimetersystem deponieren, besitzt der ATLAS Detektor zur Myonendetektion ein eigenes System, welches ebenfalls aus einem zylindrischen Zentralbereich und zwei Endkappenbereichen besteht und innerhalb eines Toroid Magnetfelds von 0.5 - 1 T liegt. Das Myonensystem deckt den Pseudorapiditätsbereich $|\eta| < 2.7$ ab und ist in der Lage, sowohl die Trajektorie als auch den Impuls der Myonen unabhängig von den anderen Detektorsystemen zu rekonstruieren.

3.2 Photonrekonstruktion

Wie in Abschnitt 2.1 erwähnt, basiert die Teilchenrekonstruktion auf der Wechselwirkung des Teilchens mit dem Detektor. Eine zentrale Rolle für die $W + \gamma + \gamma$ Analyse spielt die Rekonstruktion von Photonen, so dass diese in folgendem Kapitel näher erläutert wird. Ausgangspunkt für die Photonrekonstruktion ist eine Energiedeposition im elektromagnetischen Kalorimeter [11]. Da bei den Kollisionen hochenergetische Photonen entstehen können und zwischen Kollisionspunkt und elektromagnetischen Kalorimeter ca. zwei Strahlungslängen Detektormaterial durchquert werden müssen, ist es wahrscheinlich, dass ein Photon noch innerhalb des Inneren Detektors durch Paarbildung ein Elektron-Positron Paar erzeugt⁶. Daher sind unterschiedliche Rekonstruktionalgorithmen für unkonvertierte⁷ und konvertierte Photonen notwendig.

³Die Bezeichnung Strip (deutsch Streifen) ist der feinen, streifenförmigen Granularität dieser Kalorimeterschicht geschuldet.

⁵Tile bedeutet auf deutsch Kacheln, daher der Name des Tile Kalorimeters

⁶Photonen dieser Art werden als konvertierte Photonen bezeichnet

⁷ Photonen die kein Elektron-Positron Paar vor Erreichen des elektromagnetischen Kalorimeter bilden, bezeichnet man dementsprechend als unkonvertierte Photonen.

3.2.1 Rekonstruktion unkonvertierter Photonen

Energiedepositionen im elektromagnetischen Kalorimeter, welchen weder Konvertierungsvertizes noch Primärvertizes zugeordnet werden können, werden als unkonvertierte Photonen rekonstruiert. Ein Primärvertex gilt als zugeordnet, falls eine vom Kollisionspunkt ausgehende, rekonstruierte Trajektorie innerhalb eines (η, ϕ) Bereichs von (0.05×0.05) um das Zentrum der Energiedeposition liegt. Ein Konvertierungsvertex bezeichnet die Stelle im Detektor, an der die Elektron-Positron Paarbildung stattgefunden hat. Der Primärvertex bezeichnet die Stelle der Primärwechselwirkung. Die Effizienz der Photonrekonstruktion im ATLAS Detektor wurden in [11] zu annähernd 100% bestimmt.

3.2.2 Rekonstruktion konvertierter Photonen

Die Rekonstruktion von konvertierten Photonen gliedert sich in zwei Teile:

- Rekonstruktion von Konvertierungsvertizes im Inneren Detektor
- Zuordnung eines Konvertierungsvertex zu einer Energiedeposition im elektromagnetischen Kalorimeter

Da die Signatur eines Elektrons im Detektor der eines Photons sehr ähnlich ist (Energiedeposition im elektromagnetischen Kalorimeter und rekonstruierte Spur im Inneren Detektor), werden viele konvertierten Photonen als Elektronen rekonstruiert. Es ist ein aufwändiger Algorithmus von Nöten um Elektronen und Photonen zu trennen, der weitere Bedingungen an die rekonstruierten Objekte und deren Trajektorien stellt. Dieser Algorithmus wird in [11] ausführlich beschrieben.

3.3 Photonidentifikation

Wie in [11] beschrieben ist nach der Photonenrekonstruktion davon auszugehen, dass nahezu alle Photonen rekonstruiert wurden. Unter den rekonstruierten Photonkandidaten befindet sich jedoch auch eine große Zahl an Untergrundphotonen. Um diese herauszufiltern, wird ein Identifikationsalgorithmus auf jeden rekonstruierten Photonkandidaten angewandt. Der grundlegende Photonidentifikationsalgorithmus basiert auf Selektionsschnitten [11], welche Informationen aus dem elektromagnetischen und dem hadronischen Kalorimeter verwenden. Die Selektionsvariablen, welche auf der Form der elektromagnetischen Schauer im Kalorimeter beruhen, werden Shower Shape⁸ Variablen genannt. Studien in [11, 10] haben gezeigt, dass die Schauerformen von Photonen, die aus dem Zerfall von elektrisch neutralen Hadronen stammen, sich von den Schauerformen im Streuprozess entstehender Photonen unterscheiden, so dass sich die Shower Shape Variablen zur Unterdrückung von Jet Untergrundereignissen eignen.

⁸deutsch Schauer Form

3.3.1 Kalorimeterbasierte Selektionsvariablen

Man kann die kalorimeterbasierten Selektionsvariablen in 3 Gruppen einteilen [11, 10]:

- Variablen, die sowohl auf Informationen aus dem elektromagnetischen als auch aus dem hadronischen Kalorimeter basieren:
 - **Hadronic Leakage** R_{had1}/R_{had1} : Im Bereich $|\eta| < 0.8$ und $|\eta| > 2.7$ ist der so genannte Hadronic Leakage als das Verhältnis der transversalen Photonenergie in einem Fenster $\Delta\eta \times \Delta\phi = 0.24 \times 0.24$ in der ersten Lage des hadronischen Kalorimeters zur gesamten transversalen Photonenergie-deposition (R_{had1}) definiert. Im Bereich $0.8 < |\eta| < 2.7$ wird nicht nur die transversale Energie in der ersten Lage des hadronischen Kalorimeters verwendet, sondern die gesamte transversale Energie-deposition (R_{had}). Für Photonen erwartet man ausschließlich eine Energie-deposition im elektromagnetischen Kalorimeter, wogegen hadronhaltige Jets einen Teil der Energie im hadronischen Kalorimeter deponieren.
- Variablen, die ausschließlich Informationen aus der mittleren Schicht des elektromagnetischen Kalorimeters (Middle Layer) verwenden:
 - η **Energie Verhältnis** R_η : Das Verhältnis der rekonstruierten Energie-deposition in einem Fenster von $\Delta\eta \times \Delta\phi = 3 \times 7$ zu der rekonstruierten Energie in einem Fenster von $\Delta\eta \times \Delta\phi = 7 \times 7$ um den Mittelpunkt der Deposition ($\Delta\phi$ und $\Delta\eta$ in Einheiten von Kalorimeterzellen).
 - ϕ **Energie Verhältnis** R_ϕ : Das Verhältnis der rekonstruierten Energie-deposition in einem Fenster von $\Delta\eta \times \Delta\phi = 3 \times 3$ zu der rekonstruierten Energie in einem Fenster von $\Delta\eta \times \Delta\phi = 3 \times 7$ um den Mittelpunkt der Deposition ($\Delta\phi$ und $\Delta\eta$ in Einheiten von Kalorimeterzellen).
 - η **laterale Breite** w_2 : Gibt die laterale Breite in η in einem 3×5 Zellen Fenster an und ist definiert als die energiegewichtete Summe über alle Zellen.
$$w_2 = \sqrt{\frac{\sum (E_c \times \eta_c^2)}{\sum E_c} - \left[\frac{\sum (E_c \times \eta_c)}{\sum E_c}\right]^2}$$
wobei E_c die Energie-deposition in der Zelle η_c darstellt.
- Variablen, welche Informationen aus der Strip Layer des elektromagnetischen Kalorimeters nutzen und somit aufgrund der feinen Granularität des Kalorimeters in η -Richtung eine besonders starke Unterdrückung von hochenergetischen Jets, die neutrale Hadronen wie π^0 und η^0 enthalten, ermöglichen. Da π^0 in zwei Photonen zerfallen, die bei hohen Energien im Ruhesystem des Detektors sehr nahe beieinander liegen, erwartet man für ihre Schauer breitere Formen und oft auch zwei Maxima innerhalb des Schauers, im Gegensatz zu den schmaleren Schauern, die von einzelnen direkt aus dem Streuprozess stammenden Photonen ausgelöst werden. Innerhalb eines Fensters der Größe

$\Delta\eta \times \Delta\phi = 0.125 \times 0.2$ um die Strip Layer Zelle⁹ mit der größten transversalen Energiedeposition¹⁰ werden weitere Maxima gesucht, wobei bei mehr als einem Nebenmaximum nur das mit der größten transversalen Energiedeposition betrachtet wird.

- **Zweite Maximum Differenz ΔE :** Gibt die Differenz des Energienebenmaximums und der niedrigsten transversalen Energiedeposition an, welche innerhalb der Strips zwischen Haupt- und Nebenmaximum rekonstruiert wurde. $\Delta E_s = E_{max2} - E_{min}$.
- **Seitliches Energieverhältnis F_{side} :** Das Verhältnis der deponierten Energie außerhalb eines drei Zellen großen Kerns und der deponierten Energie innerhalb dieses Kerns.

$$F_{side} = \frac{[E(\pm 3) - E(\pm 1)]}{E(\pm 1)}$$
, wobei $E(\pm n)$ die rekonstruierte Energiedeposition in $\pm n$ Zellen um die Zelle, in der das Hauptmaximum liegt, darstellt.
- **Laterale Breite w_{s3} :** Ein Maß für die Breite des elektromagnetischen Schauers in einem Bereich von drei Strips um die Zelle, in der das Hauptmaximum liegt.

$$w_{s3} = \sqrt{\frac{\sum E_i \times (i - i_{max})^2}{\sum E_i}}$$
 mit der Stripidentifikationsnummer i und dem Index i_{max} , welcher die Zellenidentifikationsnummer der Zelle mit der maximalen Energiedeposition angibt. w_{s3} wird in Zelleneinheiten angegeben.
- **Totale Breite w_{stot} :** Gibt ein Maß für die gesamte laterale Breite des Schauers, ebenfalls in Einheiten von Zellen. Es wird die gleiche Formel wie für die w_{s3} Berechnung angewandt, jedoch gehen in die Berechnung alle Strips um das Hauptmaximum mit ein, die 2.5 Zellen des Middlelayers abdecken.
- **Relatives Maximaverhältnis E_{ratio} :** Stellt die relative Differenz von Haupt- und Nebenmaximumenergiedeposition innerhalb des Schauers dar.

$$E_{ratio} = \frac{E_{max1} - E_{max2}}{E_{max1} + E_{max2}}$$

3.3.2 *loose* und *loose'* Photonidentifikation

Je nach Art und Anforderungen der Analyse gibt es unterschiedliche Photonidentifikationsgrade, welche sich zwar in der Untergrundunterdrückung, aber auch der Effizienz unterscheiden. Die verschiedenen Identifikationsgrade basieren auf den in Abschnitt 3.3.1 vorgestellten Shower Shape Selektionsvariablen. Eine mögliche Basisidentifikation, genannt *loose*, basiert nur auf den Shower Shape Variablen, welche Informationen aus dem hadronischen und dem Middle Layer des elektromagnetischen Kalorimeters verwenden (R_{had1} , R_{had} , R_η , R_ϕ , w_2). In [12] wurde die *loose*

⁹Kurz: Strip

¹⁰Im folgenden Hauptmaximum genannt

Definition um einen Schnitt auf die Variable w_{stot} , eine der Strip Layer Shower Shape Variablen, zu *loose*’ erweitert.

3.3.3 *tight* und *tight** Photonidentifikation

Erweitert man die *loose* Photonidentifikation um Schnitte, die auf Informationen der Shower Shape Variablen basieren, erhält man die *tight* Identifikation. In [11] wurden die Effizienzen und Jet Unterdrückung der *loose* und *tight* Identifikationen getestet. Hiernach erbringt die *tight* Selektion eine höhere Untergrundunterdrückung gegen Jets, was auf Kosten einer niedrigeren Photon-Identifikationseffizienz geht. Aufgrund der niedrigen Statistik, die bei der $W + \gamma + \gamma$ Analyse vorliegt, wird eine alternative *tight** Identifikation eingeführt und getestet. Hierbei wird im Vergleich zur *tight* Identifikation der Schnitt auf die w_{stot} Variable nicht implementiert.

3.3.4 Kalorimeter Isolation

Zusätzlich zu der Photonidentifikation lassen sich rekonstruierte Photonenkandidaten über ihre Kalorimeter Isolation klassifizieren. Diese Isolation ist definiert als die gesamte im Kalorimeter deponierte Energie des Photons in einem Kegel mit Radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.3$, jedoch außerhalb eines $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ großen Kerns um den Photokandidat [8]. Da Signal und Untergrundphotonen verschiedene Isolationsverteilungen besitzen, ist eine weitere Untergrundunterdrückung möglich. Bei Untergrundphotonen werden in der Regel höhere Isolationswerte als bei Signalphotonen erwartet (siehe Abbildung 3.3). Die Kalorimeter Isolation wird zusätzlich korrigiert, da das Photon Energie außerhalb des oben definierten Kegels deponieren kann, und die Möglichkeit von mehreren Wechselwirkung pro Strahlkreuzung die Messung beeinflussen kann.

3.3.5 Korrelation von Photonidentifikation und Kalorimeter Isolation

In [13] wurde die Korrelation der Shower Shape Variablen und somit der Photonidentifikation mit der Kalorimeter Isolation untersucht. Da viele der Strip Layer Variablen Informationen nutzen, die aus einer Kalorimeterregion innerhalb des in Abschnitt 3.3.4 definierten Kerns stammen, der nicht in die Isolationsberechnung eingeht, zeigen diese Variablen so gut wie keine Korrelation mit der Kalorimeter Isolation. Studien in [13] führten zu einer *loose*’ Definition, die sich nur in den vier Shower Shape Variablen w_{s3} , F_{side} , ΔE und E_{ratio} , welche die geringste Korrelation mit der

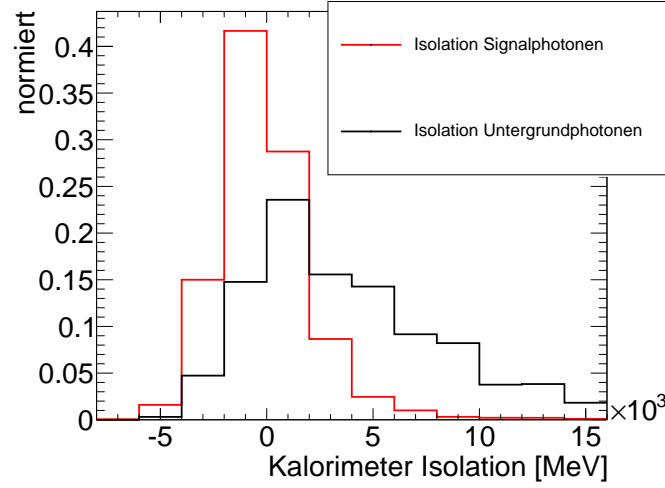


Abbildung 3.3: Dargestellt ist die normierte Isolationsverteilung für die Signal- und Untergrundphotonen, welche die größte transversale Energie haben und eine *loose* Identifikation erfolgreich passiert haben.

Isolation aufweisen, von der *tight* Identifikation unterscheidet. Dadurch unterscheiden sich die Isolationsverteilungen von *loose* und *tight* identifizierten Untergrundphotonen nicht [12]. Aufgrund dieser Ergebnisse wurde die *tight** Identifikation so gewählt, dass sich die Isolationsverteilungen von *tight** und *loose* identifizierten Untergrundphotonen ebenfalls nicht unterscheiden, da diese zwei Identifikationen sich auch in den vier oben genannten Shower Shape Variablen unterscheiden, die die niedrigste Korrelation mit der Isolation aufweisen.

4 Definition der Selektion

In folgendem Kapitel wird die $W + \gamma + \gamma$ Selektion erläutert. Die Selektionsschnitte orientieren sich im Wesentlichen an der in ATLAS durchgeführten und publizierten $W + \gamma$ Analyse [1], wobei Änderungen der zur Zeit laufenden und noch nicht veröffentlichten $W + \gamma$ Analyse [8] einbezogen werden. In der $W + \gamma + \gamma$ Analyse befinden sich im Endzustand der Proton-Proton Wechselwirkung zwei Photonen sowie ein W-Boson, wobei das W-Boson nur indirekt über seine Zerfallsprodukte rekonstruiert wird. Wie in der Einleitung eingeführt wird nur der W-Zerfall im Myon Kanal ($W \rightarrow \mu\bar{\nu}_\mu$) betrachtet. Das im Myonzerfall entstehende Neutrino kann nicht direkt im Detektor nachgewiesen werden, so dass über fehlende transversale Energie der rekonstruierten Teilchen im Endzustand auf die Präsenz eines Neutrinos geschlossen wird. Somit sind in jedem Fall ein Myonkandidat, fehlende transversale Energie und zwei Photonkandidaten in den gesuchten $W + \gamma + \gamma$ Ereignissen vorhanden.

4.1 Vorselektion

Um die Datenmenge zu reduzieren, wird zunächst eine Vorselektion durchgeführt. In der Vorselektion werden solche Ereignisse selektiert, in denen ein Myonkandidat, fehlende transversale Energie und zwei Photonkandidaten im Endzustand rekonstruiert wurden. Um die Anzahl selektierter Untergrundereignisse zu verringern, werden an die rekonstruierten Teilchen weitere Qualitätsmerkmale gestellt, welche in folgenden Abschnitten erläutert werden.

4.1.1 Myon Selektion

Die Myon Rekonstruktion wird mit dem STACO¹ Myonenrekonstruktionsalgorithmus durchgeführt [14]. Dieser Algorithmus verbindet zu Rekonstruktion der Myonentrajektorie sowohl Informationen aus dem Myonenspektrometer als auch dem Inneren Detektor. Es werden alle Ereignisse selektiert, die mindestens einen rekonstruierten Myonkandidat enthalten, der die folgende Qualitätsmerkmale erfüllt:

- transversaler Impuls² $p_T > 25 \text{ GeV}$
- Rekonstruktion in einem Pseudorapiditätsbereich von $|\eta| < 2.4$

¹Statistical Combination of myon spectrometer tracks and inner detector tracks

²Der transversale Impuls ist definiert als: $p_T = \sqrt{p_x^2 + p_y^2}$

- Rekonstruktion einer Spur im Myonenspektrometer, die mit einer Spur im Inneren Detektor hinsichtlich Impuls und Richtung übereinstimmt:
 $mu_staco_isCombinedMuon == true$
- höchste Rekonstruktionsqualität: $mu_staco_tight == true$
- Abstand von dem Primärvertex der Interaktion kleiner als 1 mm in der longitudinalen und der transversalen Ebene
- weitere Qualitätsmerkmale an die rekonstruierte Spur im Inneren Detektor:
 - $(!mu_staco_expectBLayerHit || (mu_staco_nBLHits > 0))$
 - $(mu_staco_nPixHits + mu_staco_nPixelDeadSensors) > 1$
 - $(mu_staco_nSCTHits + mu_staco_nSCTDeadSensors) >= 6$
 - $(mu_staco_nPixHoles + mu_staco_nSCTHoles) < 2$
 - im Bereich $|\eta| < 1.9$: $N_{TRT}^{hits} >= 6$ und $\frac{N_{TRT}^{outliers}}{N_{TRT}^{hits}} < 0.9$ mit $N_{TRT} = N_{TRT}^{hits} + N_{TRT}^{outliers}$
 - im Bereich $|\eta| > 1.9$ und $N_{TRT}^{hits} >= 6$: $\frac{N_{TRT}^{outliers}}{N_{TRT}^{hits}} < 0.9$
- Myonkandidat ist isoliert: Die Summe der transversalen Impulse aller rekonstruierten Spuren in einem Kegel mit Radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ (mit Ausnahme der Spur des Myons selbst) normiert auf den transversalen Myonenimpuls ist kleiner als 0.15: $\sum p_T^{tracks(\Delta R < 0.3)} / p_T < 0.15$
- Abstand zu allen rekonstruierten Jetkandidaten von $\Delta R^3 > 0.4$

4.1.2 Fehlende transversale Energie

Da Neutrinos weder der starken noch der elektromagnetischen Wechselwirkung unterliegen sind sie nicht direkt im ATLAS Detektor nachweisbar. Daher wird die Summe der Transversalenergie aller rekonstruierter Teilchen im Endzustand gebildet und nach Verletzungen der Energieerhaltung gesucht. Die fehlende transversale Energie (MET) wird analog zu der $W + \gamma$ Analyse mit $MET_RefFinal$ bestimmt. Eine genauere Beschreibung der MET Rekonstruktion findet sich in [15]. Hierbei werden alle Energiedepositionen in den Kalorimeterzellen mit $|\eta| > 4.9$, sowie die Energie rekonstruierter Myonen im Myonensystem berücksichtigt. Es wird eine fehlende transversale Energie von $MET > 35$ GeV gefordert.

4.1.3 Photon Selektion

In der Vorselektion werden nur Events mit mindestens zwei *loose* identifizierten Photonkandidaten ausgewählt. Außerdem müssen die Photonkandidaten weitere Merkmale erfüllen:

$$^3 \Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

- rekonstruierte transversale Energie $E_T > 15$ GeV
- Rekonstruktion in einem Pseudorapiditätsbereich von $|\eta| < 1.37$ oder $1.52 < |\eta| < 2.37$
- keine Rekonstruktion in Kalorimeterbereichen mit Auslesefehlern oder defekter Kalorimiterelektronik
- Abstand zu allen rekonstruierten Jetkandidaten von $\Delta R > 0.4$
- Abstand zum den rekonstruierten Myonkandidaten von $\Delta R > 0.4$
- Abstand der beiden Photonkandidaten von $\Delta R > 0.3$

Die Abstände der rekonstruierten Photonkandidaten zu Jets, Myonen und untereinander sind nötig, um die Bestimmung der Kalorimeter Isolation nicht zu beeinflussen.

4.1.4 Unterdrückung von elektroschwachen Untergrundprozessen

Um die elektroschwachen Untergrundprozesse zu reduzieren, wird die Selektion um zwei weitere Bedingungen erweitert [4, 1]:

- Forderung nach genau einem Myon
- Veto von Ereignissen, in denen ein Jet rekonstruiert wurde, der aus dem Zerfall eines B-Quarks stammt. Dieses Veto dient zur Reduktion von Ereignissen mit einem Top-Quark, da dieses in ein W-Boson und ein B-Quark zerfällt.

4.2 Finale Selektion

In der finalen Selektion, welche die Signalregion der $W + \gamma + \gamma$ Analyse definiert, werden an beide Photonkandidaten eine weitere Identifikationsbedingung und eine kalorimeterbasierte Isolationsbedingung gestellt. In der $W + \gamma$ Analyse [8] ist der übliche Identifikationsschnitt die *tight* Identifikation und die Bedingung für die Kalorimeter Isolation liegt bei $E_T^{iso} < 6$ GeV für den rekonstruierten Photonkandidaten. Alternativ wird im Folgenden aufgrund der niedrigen Statistik des $W + \gamma + \gamma$ Prozesses der in Abschnitt 3.3.3 eingeführte *tight** Identifikationsschnitt und voneinander unabhängig variierende Isolationsbedingung für die beiden Photon von $E_T^{iso} < 3-7$ GeV eingeführt. Im Folgenden wird das Photon mit dem höheren transversalen Impuls leading Photon, und das Photon mit dem niedrigeren transversalen Impuls subleading Photon genannt.

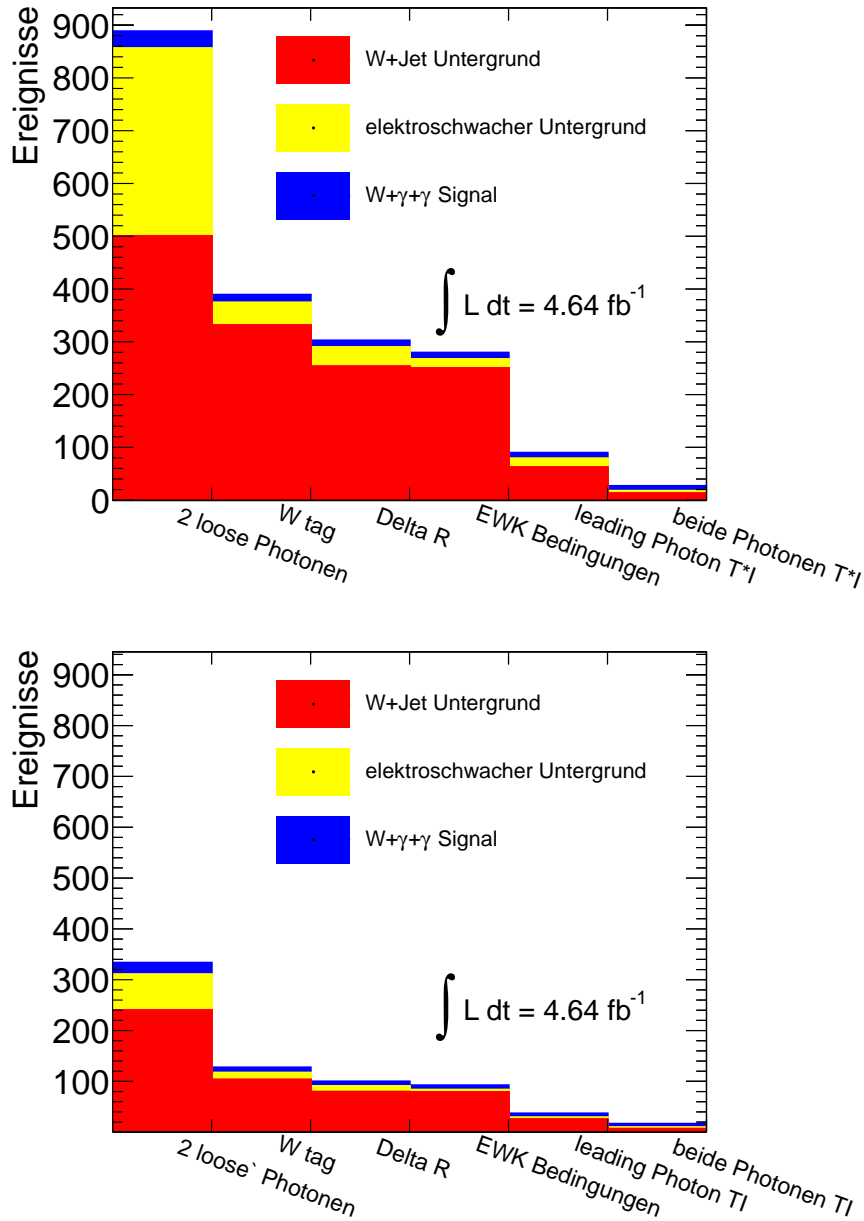


Abbildung 4.1: Zu sehen ist die Anzahl der selektierten Ereignisse. Die Forderung nach mindestens einem Myon, welches die Qualitätsmerkmale erfüllt und der MET wird als *Wtag* zusammengefasst. Die Bedingungen zur Unterdrückung der elektroschwachen Untergrundprozessen wird mit EWK Bedingungen abgekürzt. *Oben*: Definition der Signalregion: *tight** Identifikation und Isolation < 6 GeV für beide Photonen. *Unten*: Erweiterung der Vorselektion um eine *loose'* Bedingung und *tight* Identifikation und Isolation < 6 GeV für beide Photonkandidaten in der Signalregion.

5 Untergrundbestimmung

5.1 Die 2×2 D Sideband Methode

Der dominante Untergrund der $W + \gamma + \gamma$ Analyse setzt sich aus $W + jet + \gamma$ und $W + jet + jet$ Ereignissen zusammen, in denen ein, beziehungsweise zwei Jets als Photon rekonstruiert und identifiziert, sowie alle Qualitätsanforderungen an die Photonkandidaten erfüllt werden. Da dieser Prozess in Monte Carlo Simulationen nicht verlässlich modelliert werden kann [12], wird eine datenbasierte Untergrundabschätzung mit Hilfe der 2×2 D Sideband Methode durchgeführt, welche in [12] als Erweiterung der 2D Sideband Methode¹ [16] für zwei Photon Endzustände eingeführt wurde.

Die Grundidee besteht darin, für beide Photonen mit Hilfe von jeweils drei Kontrollregionen die Anzahl an Untergrundereignissen in eine Signalregion zu extrapolieren. Im nächsten Abschnitt wird die in [12] entwickelte Methode vorgestellt.

In der 2×2 D Sideband Methode können die Untergrundabschätzung des elektroschwachen und des $W + jets$ Untergrunds nicht gleichzeitig durchgeführt werden, da die Verteilung der elektroschwachen Untergrundereignisse auf Signal- und Kontrollregion nicht der Verteilung der $W + jet + \gamma$ und $W + jet + jet$ Ereignisse entspricht. Zur Abschätzung des elektroschwachen Untergrunds wird eine auf Monte Carlo Simulationen basierende Abschätzung durchgeführt, indem die $W + \gamma + \gamma$ Selektion auf die Monte Carlo Samples der elektroschwachen Prozesse angewandt, und somit der Anzahl an Ereignissen bestimmt wird, welche die Selektion überstehen.

5.1.1 Definition von Signal und Kontrollregionen

In Abschnitt 4 wurde die Vorselektion der $W + \gamma + \gamma$ Ereignisse vorgestellt. Die vorselektierten Ereignisse werden in der 2×2 D Sideband Methode zunächst durch Einführung zweier weiterer Selektionsschnitten auf die Photon-Identifikation und Kalorimeter Isolation in vier Feldern aufgeteilt, wobei im ersten Schritt nur das leading Photon betrachtet wird.

In Abbildung 5.1 (oben) ist veranschaulicht, wie in der zweidimensionalen Ebene, die von Identifikation und Isolation aufgespannt wird, vier Felder durch die Schnitte definiert werden. In Feld A werden beide zusätzliche Bedingungen erfüllt, so dass dieses Feld die Signalregion darstellt. In den Feldern B und C werden jeweils eine, und in Feld D beide zusätzlichen Bedingungen nicht erfüllt, womit diese Felder als Untergrund Kontrollregionen dienen. Im zweiten Schritt werden in der 2×2 D Sideband Methode nur Ereignisse betrachtet, die in der Signalregion A liegen. In diesen

¹Oft als „ABCD Methode“ bekannt

Ereignissen werden an das subleading Photon anschließend wieder die zusätzliche Identifikations- und Isolationsbedingung gestellt. In Abbildung 5.1 unten ist wieder die zweidimensionale Darstellung in der Identifikations- Isolationsebene dargestellt. Es ergeben sich vier Felder, wobei in Feld A' das subleading Photon beide zusätzlichen Bedingungen erfolgreich erfüllt (Signalregion) und in den Regionen B', C' und D' eine beziehungsweise beide Bedingungen nicht erfüllt werden. In der graphischen Darstellung ergeben sich also zwei zweidimensionale Ebenen, womit sich auch der Name der 2×2D Sideband Methode erklärt.

Im Folgenden werden der Einfachheit halber vorselektierte Photonen, welche die zusätzliche Identifikationsbedingung nicht erfüllen, als *loose* bezeichnet, wogegen Photonen, die den zusätzlichen Identifikationsschnitt erfolgreich passieren als *tight* identifiziert bezeichnet werden.

Außerdem werden folgende Rechnungen frei von einer Festlegung auf eine konkrete Isolationsbedingung sein. Es wird nur die Rede von isolierten Photonen sein, falls die zusätzliche Isolationsbedingung erfüllt wird, oder entsprechend von nicht isolierten Photonen, falls die Isolationsbedingung nicht erfüllt wird. Für die 2×2D Sideband Methode ist es unerheblich, welche konkreten Variablen die zusätzlichen Schnitte und somit die Signal- und die Kontrollregionen definieren, jedoch ist es notwendig, dass die gewählten Variablen keine Korrelation aufweisen. Daher wird in dem sich anschließenden Abschnitt 5.1.6 erläutert, in wie weit die in Abschnitt 3.3.2 und 3.3.3 eingeführten Identifikationsgrade und die Kalorimeter Isolation aus Abschnitt 3.3.5 verwendet werden können.

Wie oben erläutert setzt sich die Signalregion A' aus Photonenpaaren zusammen, welche beide das zusätzliche Identifikationskriterium und die Isolationsbedingung erfüllen. Zur Verdeutlichung der verwendeten Notation folgt nun die Zusammensetzung der Ereignisse in den Signalfeldern A und A':

$$N_A = W_{tot}^{TI} = W_{\gamma\gamma}^{TI} + W_{\gamma j}^{TI} + W_{j\gamma}^{TI} + W_{jj}^{TI} \quad (5.1)$$

$$N'_A = W_{tot}^{TITI} = W_{\gamma\gamma}^{TITI} + W_{\gamma j}^{TITI} + W_{j\gamma}^{TITI} + W_{jj}^{TITI} \quad (5.2)$$

Die ersten beiden hochgestellten und der erste tiefgestellte Index beziehen sich auf das leading Photon. Unter den hochgestellten Indizes bedeutet **T**, dass die zusätzliche Identifikationsbedingung erfüllt wurde und **I** kennzeichnet die erfüllte Isolationsbedingung. Der tiefgestellte Indize zeigt an, ob es sich bei dem betrachtete Photonkandidat wirklich um ein Signalphoton handelt, oder ob es sich um einen Jet² handelt³. Für das subleading Photon gilt die gleiche Notation, jedoch beziehen sich auf dieses die letzten beiden hochgestellten und der zweite tiefgestellte Index.

²In den meisten Fällen wird nicht der Jet selbst als Photon rekonstruiert und identifiziert, sondern es handelt sich um Photonen aus sekundären Zerfallsprozesse hadronischer Jets

³Die Einteilung nach Signal- und Untergrundphotonkandidaten dient der folgenden Rechnung, in der Praxis kann diese Unterscheidung natürlich nicht durchgeführt werden

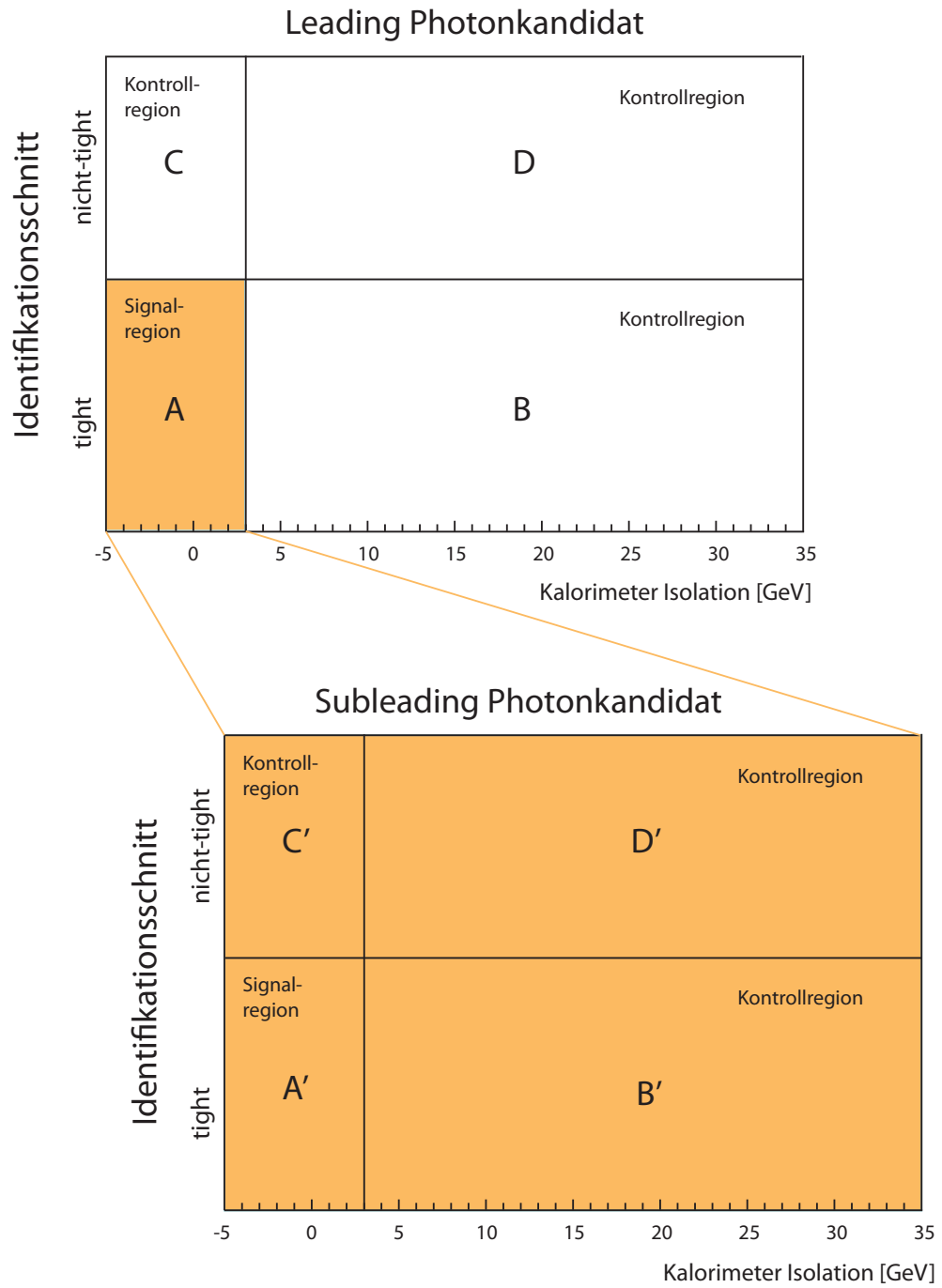


Abbildung 5.1: Signal- und Kontrollregionen der 2×2D Sideband Methode. Erstellt nach einem Vorbild in [12].

5.1.2 Anwendung der 2D Sideband Methode auf das Photon mit höherem Transversalimpuls

Zuerst wird die 2D Sideband Methode auf das leading Photon angewandt. Dies bedeutet, alle vorselektierten Ereignisse werden nach den in Abschnitt 5.1.1 eingeführten Schnitten in eine Signalregion und drei Untergrund Kontrollregionen eingeteilt. Obwohl die Selektion so konstruiert ist, dass sich in Region A am meisten Signalereignisse befinden, wird trotzdem ein gewisser Teil an Signalereignissen in den Kontrollfeldern, und eine nicht zu vernachlässigende Anzahl an Untergrundergebnissen in die Signalregion fallen. Die Ereignisse in den Feldern setzen sich nun wie folgt zusammen:

$$N_A = N_A^{sig} + N_A^{bkg} \quad (5.3)$$

$$N_B = N_B^{sig} + N_B^{bkg} \quad (5.4)$$

$$N_C = N_C^{sig} + N_C^{bkg} \quad (5.5)$$

$$N_D = N_D^{sig} + N_D^{bkg} \quad (5.6)$$

Unter der Annahme, dass die Isolationsverteilung und die Identifikation von Untergrundphotonen unkorreliert sind, lässt sich die Anzahl an Untergrundereignissen in der Signalregion über die Anzahl der Untergrundereignisse in den Kontrollregionen bestimmen:

$$\frac{N_A^{bkg}}{N_B^{bkg}} = \frac{N_C^{bkg}}{N_D^{bkg}} \quad (5.7)$$

somit folgt für die Anzahl der Untergrundereignisse in A:

$$N_A^{bkg} = N_B^{bkg} \cdot \frac{N_C^{bkg}}{N_D^{bkg}} \quad (5.8)$$

Setzt man nun Gleichung 5.8 in Gleichung 5.7 ergibt sich für die Anzahl an Signalereignissen in A:

$$N_A^{sig} = N_A - N_B^{bkg} \cdot \frac{N_C^{bkg}}{N_D^{bkg}} \quad (5.9)$$

5.1.3 Korrektur aufgrund von Korrelationen und Signalereignissen in den Kontrollregionen

Gleichung 5.9 ist nur gültig, wenn, wie oben schon erwähnt, Identifikations- und Isolationsschnitt vollkommen unkorreliert sind. Zwar wird der Identifikationsschnitt so gewählt, dass keine zusätzlichen Korrelationen mit der Isolation erwartet wird, jedoch ist a priori nicht auszuschließen, dass doch eine Korrelation vorliegt. Daher

wird hier der Einfluss einer möglichen Korrelation durch die Definition der Größe R^{bkg} eingeführt:

$$R^{bkg} = \frac{N_D^{bkg}}{N_C^{bkg}} \frac{N_A^{bkg}}{N_B^{bkg}} \quad (5.10)$$

Mit der eingeführten Korrelation verändert sich Gleichung 5.9 zu:

$$N_A^{sig} = N_A - N_B^{bkg} \cdot \left[\frac{N_C^{bkg}}{N_D^{bkg}} \right] R^{bkg} \quad (5.11)$$

Falls Identifikations- und Isolationsschnitt vollkommen unkorreliert sind, ergibt sich $R^{bkg} = 1$ und Gleichung 5.11 geht wieder in Gleichung 5.9 über.

Außerdem muss berücksichtigt werden, dass sich je nach Definition der Signalregion in den Kontrollregionen nicht nur Untergrundereignisse, sondern auch Signalereignisse befinden. Dies wird als Signal Leakage⁴ bezeichnet. Um dies zu beschreiben werden folgenden Parameter definiert:

$$c_1 = \frac{N_B^{sig}}{N_A^{sig}} \quad (5.12)$$

$$c_2 = \frac{N_C^{sig}}{N_A^{sig}} \quad (5.13)$$

$$c_3 = \frac{N_D^{sig}}{N_A^{sig}} \quad (5.14)$$

Nun lässt sich die Anzahl an Untergrundereignissen in den Kontrollregionen mit Hilfe der Signal Leakage, der Anzahl an Signalereignissen im Signalfeld A und der gemessenen Anzahl an Ereignissen in den jeweiligen Kontrollregionen ausdrücken. Hierzu wird die Ereignisanzahl in den Kontrollfeldern um die Anzahl an Signalereignissen reduziert.

$$N_B^{bkg} = N_B - c_1 \cdot N_A^{sig} \quad (5.15)$$

$$N_C^{bkg} = N_C - c_2 \cdot N_A^{sig} \quad (5.16)$$

$$N_D^{bkg} = N_D - c_3 \cdot N_A^{sig} \quad (5.17)$$

Schließlich werden in Gleichung 5.11 die Ausdrücke für die Untergrundereignisse in den Kontrollregionen N_B^{bkg} , N_C^{bkg} und N_D^{bkg} mit Hilfe von den Gleichungen 5.15 ff. ersetzt und es folgt ein nicht linearer Ausdruck für N_A^{sig}

$$N_A^{sig} = N_A - \left[(N_B - c_1 N_A^{sig}) \frac{N_C - c_2 N_A^{sig}}{N_D - c_3 N_A^{sig}} \right] R^{bkg} \quad (5.18)$$

⁴Von englisch leakage deutsch Verlust

Die Lösungen dieser quadratischen Gleichung haben die übliche Form

$$N_A^{sig} = \frac{b \pm \sqrt{b^2 - 4ac}}{2a} \quad (5.19)$$

wobei folgende Substitutionen vorgenommen wurden:

$$a = c_3 - c_1 c_2 R^{bkg} \quad (5.20)$$

$$b = N_D + c_3 N_A - (c_1 N_C + c_2 N_B) R^{bkg} \quad (5.21)$$

$$c = N_D N_A - N_C N_B R^{bkg} \quad (5.22)$$

Auch wenn die Mathematik zwei Lösungen ergibt, ist die folgende die einzig physikalisch sinnvolle:

$$N_A^{sig} = \frac{b - \sqrt{b^2 - 4ac}}{2a} \quad (5.23)$$

5.1.4 Anwendung der 2D Sideband Methode auf das zweite Photon

Innerhalb des Feldes A, in dem sich die Ereignisse befinden, in denen der leading Photonkandidat *tight* identifiziert wurde und die Isolationsbedingung erfüllt, wird nun auf den subleading Photonkandidat analog zum leading Photon die 2D Sideband Methode angewandt. Hierzu wird das Feld A durch Forderung von einem zusätzlichen Identifikationsschnitt und einer Isolationsbedingung an das subleading Photon in eine Signalregion A' und die drei Untergrundkontrollregionen B', C' und D' eingeteilt. Wieder führt die Tatsache einer nicht vollkommen effizienten Selektion dazu, dass sich sowohl Signalereignisse in den Kontrollregionen, als auch Untergrundergebnisse in der Signalregion befinden. Zur Berechnung der Signalereignisse $N_A^{sig'}$ in Region A' werden analog zu 5.1.2 neben den Gesamtereigniszahlen N'_A, N'_B, N'_C und N'_D in den vier Feldern die folgenden Parameter benötigt:

- Die Korrelation von Isolationschnitt und Identifikationsschnitt der Untergrundergebnisse:

$$R^{bkg'} = \frac{N_D^{bkg'} N_A^{bkg'}}{N_C^{bkg'} N_B^{bkg'}} \quad (5.24)$$

- Die Signal Leakage in die Kontrollregionen:

$$c'_1 = \frac{N_B^{sig'}}{N_A^{sig'}} \quad (5.25)$$

$$c'_2 = \frac{N_C^{sig'}}{N_A^{sig'}} \quad (5.26)$$

$$c'_3 = \frac{N_D^{sig'}}{N_A^{sig'}} \quad (5.27)$$

Führt man nun die gleiche Rechnung wie oben durch, ergibt sich wieder eine quadratische Lösung für die Anzahl an Signalereignissen in A':

$$N_A^{sig'} = N'_A - \left[(N'_B - c'_1 N_A^{sig'}) \frac{N'_C - c'_2 N_A^{sig'}}{N'_D - c'_3 N_A^{sig'}} \right] R^{bkg'} \quad (5.28)$$

welche eine zu 5.23 analoge Lösung hervorbringt.

$$N_A^{sig'} = \frac{b' - \sqrt{b'^2 - 4a'c'}}{2a'} \quad (5.29)$$

mit den Substitutionen

$$a' = c'_3 - c'_1 c'_2 R^{bkg'} \quad (5.30)$$

$$b' = N'_D + c'_3 N'_A - (c'_1 N'_C + c'_2 N'_B) R^{bkg'} \quad (5.31)$$

$$c' = N'_D N'_A - N'_C N'_B R^{bkg'} \quad (5.32)$$

5.1.5 Berechnung der Anzahl an Signal- und Untergrundereignissen in der Signalregion A'

Wie in 5.1.1 eingeführt, setzt sich die Gesamtheit der Ereignisse, in denen der leading Photonkandidaten die *tight* und Isolationsbedingung erfüllt, wie folgt zusammen:

$$W_{tot}^{TI} = W_{\gamma\gamma}^{TI} + W_{\gamma j}^{TI} + W_{j\gamma}^{TI} + W_{jj}^{TI} \quad (5.33)$$

Im Vergleich dazu ist die Zusammensetzung aller Ereignisse nach der Basisidentifikation, in denen beide Photonen *loose* identifiziert wurden:

$$W_{tot}^{LL} = W_{\gamma\gamma}^{LL} + W_{\gamma j}^{LL} + W_{j\gamma}^{LL} + W_{jj}^{LL} \quad (5.34)$$

Führt man die Größe ϵ ein, die die Wahrscheinlichkeit angibt, mit der das leading Photon⁵ die Selektionsschnitte passiert, sowie die Größen f , welche die Wahrscheinlichkeit angibt, dass ein fake Photon, welches als leading Photonkandidat rekonstruiert wurde⁶, die Selektionsschnitte erfolgreich passiert, lassen sich Gleichung 5.33 und Gleichung 5.34 verbinden.

⁵Es wird nicht unterschieden, ob das leading Photon aus einem $\gamma\gamma$ oder einem γj Ereignis stammt

⁶Wieder sei es nicht relevant, ob es sich um ein $j\gamma$ oder ein γj Ereignis handelt

$$W_{\gamma\gamma}^{TI} = \epsilon W_{\gamma\gamma}^{LL} \quad (5.35)$$

$$W_{\gamma j}^{TI} = \epsilon W_{\gamma j}^{LL} \quad (5.36)$$

$$W_{j\gamma}^{TI} = f W_{j\gamma}^{LL} \quad (5.37)$$

$$W_{jj}^{TI} = f W_{jj}^{LL} \quad (5.38)$$

$$(5.39)$$

Mit der Einführung der entsprechenden Größen ϵ' und f' für das subleading Photon gilt folgender Zusammenhang:

$$W_{\gamma\gamma}^{TITI} = \epsilon\epsilon' W_{\gamma\gamma}^{LL} = \epsilon' W_{\gamma\gamma}^{TI} \quad (5.40)$$

$$W_{\gamma j}^{TITI} = \epsilon f' W_{\gamma j}^{LL} = f' W_{\gamma j}^{TI} \quad (5.41)$$

$$W_{j\gamma}^{TITI} = f \epsilon' W_{j\gamma}^{LL} = \epsilon' W_{j\gamma}^{TI} \quad (5.42)$$

$$W_{jj}^{TITI} = f f' W_{jj}^{LL} = f' W_{jj}^{TI} \quad (5.43)$$

Das Ziel ist nun, die Größen $W_{\gamma\gamma}^{TITI}$, $W_{\gamma j}^{TITI}$, $W_{j\gamma}^{TITI}$ und W_{jj}^{TITI} , welche die relevanten Größen für die $W + \gamma + \gamma$ Analyse darstellen, aus den Ergebnissen der 2D Sideband Methode für leading und subleading Photon zu erhalten.

Dazu muss als erstes die Zusammensetzung der N_A^{sig} , N_A^{bkg} , $N_A^{sig'}$ und $N_A^{bkg'}$ aus 5.1.2 und 5.1.4 geklärt werden:

$$N_A^{sig} = W_{\gamma\gamma}^{TI} + W_{\gamma j}^{TI} \quad (5.44)$$

$$N_A^{bkg} = W_{j\gamma}^{TI} + W_{jj}^{TI} \quad (5.45)$$

$$N_A^{sig'} = W_{\gamma\gamma}^{TITI} + W_{j\gamma}^{TITI} \quad (5.46)$$

$$N_A^{bkg'} = W_{\gamma j}^{TITI} + W_{jj}^{TITI} \quad (5.47)$$

An dieser Stelle wird klar, warum es *nicht* ausreicht, die Anzahl an Ereignissen in der Signalregion $N_A^{sig'}$ zu bestimmen. Da die 2×2D Sideband Methode sukzessive für leading und subleading Photon durchgeführt wird, ist jeweils keine Aussage über das andere Photon möglich. Dies bedeutet in der Ereignismenge $N_A^{sig'}$ befinden sich Ereignisse, in denen das leading Photon ein Jet ist. Gleiche Argumentation gilt auch für die Ereignisse in N_A^{sig} .

Nun lässt sich mit den Gleichungen 5.44 ff. und 5.40 ff. das zu lösende Gleichungssystem formulieren:

$$N_A^{sig} = \frac{1}{\epsilon'} W_{\gamma\gamma}^{TITI} + \frac{1}{f'} W_{j\gamma}^{TITI} \quad (5.48)$$

$$N_A^{bkg} = N_A - N_A^{sig} = \frac{1}{\epsilon'} W_{j\gamma}^{TITI} + \frac{1}{f'} W_{jj}^{TITI} \quad (5.49)$$

$$N_A^{sig'} = W_{\gamma\gamma}^{TITI} + W_{j\gamma}^{TITI} \quad (5.50)$$

$$N_A^{bkg'} = N'_A - N_A^{sig'} = W_{j\gamma}^{TITI} + W_{jj}^{TITI} \quad (5.51)$$

Um die Größe f zu bestimmen, werden Gleichung 5.35ff. mit Gleichung 5.44 und 5.45 verbunden:

$$N_A^{sig} = \epsilon W_{\gamma\gamma}^{LL} + \epsilon W_{j\gamma}^{LL} \quad (5.52)$$

$$N_A^{bkg} = f W_{j\gamma}^{LL} + f W_{jj}^{LL} \quad (5.53)$$

Es ergibt sich:

$$f = \frac{N_A^{bkg}}{W_{j\gamma}^{LL} + W_{jj}^{LL}} \quad (5.54)$$

$$f = \frac{N_A - N_A^{sig}}{N_A + N_B + N_C + N_D - \frac{1}{\epsilon} N_A^{sig}} \quad (5.55)$$

wobei in 5.55 ausgenutzt wurde, dass die Summe aller Ereignisse sich aus Signal- und Untergrundereignissen zusammensetzt und die Summe aus $W_{j\gamma}^{LL}$ und W_{jj}^{LL} den Untergrund darstellt.

Die Größe f' kann direkt aus Gleichung 5.48 und 5.49 bestimmt werden:

$$f' = \frac{N'_A - N_A^{sig'}}{N_A - \frac{1}{\epsilon} N_A^{sig'}} \quad (5.56)$$

wobei das resultierende Gleichungssystem unterbestimmt ist. Daher wird eine neue Variable α eingeführt. Diese Variable bezeichnet das Verhältnis von Untergrundergebnissen, in denen der leading Photonkandidat ein Jet ist, zu der Summe von allen Untergrundereignissen, in denen sich ein Signalphoton und ein fake Photon befinden.

$$\alpha = \frac{W_{j\gamma}^{TITI}}{W_{j\gamma}^{TITI} + W_{\gamma j}^{TITI}} \quad (5.57)$$

Die Effizienzen ϵ und ϵ' folgen direkt aus der Signal Leakage:

$$\epsilon = \frac{1}{1 + c_1 + c_2 + c_3} \quad (5.58)$$

$$\epsilon' = \frac{1}{1 + c'_1 + c'_2 + c'_3} \quad (5.59)$$

Nun ist es möglich alle benötigten Parameter zu bestimmen und es folgt die Lösung des Gleichungssystems 5.48 - 5.51:

$$W_{\gamma\gamma}^{TITI} = \frac{\epsilon'(\alpha f' N_A^{sig} + (\alpha - 1) N_A^{sig'})}{(\alpha - 1)\epsilon' + \alpha f'} \quad (5.60)$$

$$W_{\gamma j}^{TITI} = \frac{(\alpha - 1)f'(\epsilon' N_A^{sig} - N_A^{sig'})}{(\alpha - 1)\epsilon' + \alpha f'} \quad (5.61)$$

$$W_{j\gamma}^{TITI} = \frac{\alpha f'(N_A^{sig'} - \epsilon' N_A^{sig})}{(\alpha - 1)\epsilon' + \alpha f'} \quad (5.62)$$

$$W_{jj}^{TITI} = \frac{f'(\alpha \epsilon' f' N_A + (\alpha - 1)\epsilon'^2(N_A - N_A^{sig} - \alpha f' N_A^{sig'}))}{\epsilon'((\alpha - 1)\epsilon' + \alpha f')} \quad (5.63)$$

Die Summe aller Untergrundereignisse wird mit W_{bkg}^{TITI} bezeichnet und berechnet sich nach:

$$W_{bkg}^{TITI} = W_{\gamma j}^{TITI} + W_{j\gamma}^{TITI} + W_{jj}^{TITI} \quad (5.64)$$

Der analytische Fehler berechnet sich nach Gaußscher Fehlerfortpflanzung, wobei

$$N_A = N'_A + N'_B + N'_C + N'_D$$

gesetzt wird, damit die verwendete Größen voneinander unabhängig sind :

$$\Delta W_{XX}^{TITI} = \sqrt{\sum_i \left(\frac{\partial W_{XX}^{TITI}}{\partial N_i} \Delta N_i \right)^2}, \quad i \in \{B, C, D, A', B', C', D'\} \quad (5.65)$$

Hierbei steht XX für $\gamma\gamma$, γj , jj oder bkg . Um mit der 2×2D Sideband Methode eine Untergrundabschätzung durchführen zu können, müssen also die Anzahl an Einträgen in den 8 definierten Feldern ermittelt, sowie 9 Parameter, im Folgenden Monte Carlo Input Parameter genannt, mit Hilfe von Monte Carlo simulierten Ereignissen bestimmt werden:

- Die Signal Leakage Werte $c_1, c_2, c_3, c'_1, c'_2, c'_3$
- Die Korrelationen $R^{bkg}, R^{bkg'}$
- Das Verhältnis α

5.1.6 Wahl der Identifikations- und Isolationsbedingung

In Abschnitt 3.3.2 und 3.3.3 wurden die vier Photonidentifikationen *loose*, *loose'*, *tight*, und *tight** eingeführt. Es wurde in 3.3.5 erläutert, dass diese Identifikationen so definiert wurden, dass bei einem Übergang von *loose'* zu *tight*, sowie von *loose*

zu *tight** auf diejenigen Shower Shape Variablen geschnitten wird, welche wenig Korrelation mit der Kalorimeter Isolation aus Abschnitt 3.3.4 aufweisen. Da die zwei Ebenen in der 2×2D Sideband Methode (siehe Abbildung 5.1) von zwei unkorrelierten Variablen aufgespannt werden müssen, gibt es zwei Möglichkeiten, den Identifikationsschnitt zu definieren:

- *loose* Vorselektion und *tight** Identifikation als zusätzliche Bedingung
- *loose'* Vorselektion und *tight* Identifikation als zusätzliche Bedingung

Als zusätzliche Isolationsbedingung wird die Kalorimeter Isolation im Folgenden für das leading und subleading Photon zwischen 3 GeV und 7 GeV variiert. Alle Isolations- und Identifikationskombinationen werden getestet, so dass insgesamt 50 verschiedene Signalregiondefinitionen für die $W + \gamma + \gamma$ Analyse getestet werden.

5.2 Bestimmung der Monte Carlo Input Parameter

Zur Bestimmung der benötigten Monte Carlo Input Parameter werden die in Abschnitt 2 eingeführten Monte Carlo Samples verwendet. Monte Carlo simulierte Ereignisse bieten die Möglichkeit zu überprüfen, ob es sich bei rekonstruierten Photonen um Signalphotonen oder Untergrundphotonen handelt.⁷ Erst durch diese Überprüfung ist es möglich, die Monte Carlo Input Parameter zu bestimmen, da diese die genaue Kenntnis der Herkunft der selektierten Photonen voraussetzen. Mit Ausnahme der Korrelationen R^{bkg} und $R^{bkg'}$ sind die Inputparameter abhängig von der gewählten Isolations- und Identifikationsbedingung. In Abbildung 5.2 sind die Ergebnisse für die Korrelationen R^{bkg} und $R^{bkg'}$ in Abhängigkeit der Isolationsbedingung für das leading und das subleading Photon dargestellt. Als zusätzliche Identifikationsbedingung wurde eine *tight* Identifikation für die Signal- und eine *loose'* Identifikation als Vorselektion gefordert. Zu Beginn der Herleitung der 2×2D Sideband Methode wurde angenommen, dass der zusätzliche Identifikationsschnitt keine zusätzliche Korrelation mit der Kalorimeter Isolation mit sich bringt, gleichbedeutend mit $R^{bkg} = 1$ und $R^{bkg'} = 1$ unabhängig vom gewählten Isolationsschnitt. In Abbildung 5.2 ist zu sehen, dass innerhalb der Unsicherheiten die bestimmten Korrelationen mit 1 verträglich sind, jedoch kann aufgrund der geringen Monte Carlo Statistik R^{bkg} und $R^{bkg'}$ nur im Rahmen einer sehr großen Unsicherheit untersucht werden. Bei der Anwendung der 2×2D Sideband Methode auf Daten werden $R^{bkg} = R^{bkg'} = 1$ gesetzt und der Einfluss einer möglichen Abweichung wird untersucht sowie im systematischen Fehler berücksichtigt. Der systematische Fehler für R^{bkg} und $R^{bkg'}$ wird abgeschätzt, indem die mittlere Abweichung der in 5.2 bestimmten Korrelationen von 1 gebildet wird. Die Ergebnisse sind in Tabelle 5.1 zu finden. Die Ergebnisse für die weiteren Input Parameter $\alpha, c_1, c_2, c_3, c'_1, c'_2, c'_3$ sowie der Verlauf von R^{bkg} und $R^{bkg'}$ bei einer *tight** Identifikation befinden sich im Anhang.

⁷Bedingung für Signalphotonen: $0 < |ph_truth_mothertype| < 26$

Parameter	systematischer Fehler	
	<i>tight</i> Identifikation	<i>tight*</i> Identifikation
R^{bkg}	± 0.35	± 0.26
$R^{bkg'}$	± 0.39	± 0.31

Tabelle 5.1: Als systematischer Fehler wird die mittlere Abweichung der in 5.2 bestimmten Werte für R^{bkg} und $R^{bkg'}$ von 1 angenommen. In der linken Spalte werden die Ergebnisse bei einer *tight*, in der rechten Spalte bei einer *tight** Identifikationsbedingung an die Photonen in der Signalregion angegeben.

5.3 Monte Carlo Closure Test

Als erster Test der 2×2D Sideband Methode wird ein so genannter Closure Test durchgeführt, mit dessen Hilfe sichergestellt wird, dass die Methode fehlerfrei implementiert wurde. Zunächst muss jedoch geklärt werden, wie sich das verwendete Monte Carlo Sample aus den in Abschnitt 2.4 eingeführten Samples zusammen setzt, und wie die Unsicherheiten der Monte Carlo Ereigniszahlen zu verstehen sind.

5.3.1 Zusammenstellung des Monte Carlo Sample und Monte Carlo Unsicherheiten

Es gilt die Monte Carlo Samples zunächst auf eine Luminosität von 4.64 fb^{-1} zu normieren, um eine mit dem Datensample vergleichbare Ereigniszusammensetzung zu erhalten. In den einzelnen Samples ist eine feste Anzahl an Ereignissen unterschiedlicher Prozesse simuliert, die sich in ihrem Wirkungsquerschnitt und somit in ihrer erwarteten Häufigkeit unterscheiden. Um eine realistische Zusammenstellung des gesamten Sample zu bekommen, ist eine Gewichtung der Ereignisse gemäß ihres Wirkungsquerschnitts notwendig. Für die Anzahl an Ereignissen N gilt bei vorgegebener integrierter Luminosität L und Wirkungsquerschnitt σ :

$$N = \sigma \cdot L \quad (5.66)$$

Der Wirkungsquerschnitt ist bekannt und die Anzahl an Ereignissen N durch die Simulation vorgegeben. Eine Normierung auf die gewünschte integrierte Luminosität L^w erfolgt über die Einführung eines Gewichts w :

$$w = \frac{L^w \cdot \sigma}{N} \quad (5.67)$$

Nach der Gewichtung kommen alle Prozesse so häufig in dem Monte Carlo Sample vor, wie sie bei der gewählten integrierten Luminosität und des entsprechenden Wirkungsquerschnitts erwarten werden, unabhängig davon welche absolute Anzahl an simulierten Ereignissen eines Prozesses vorliegen. Für die folgenden Monte Carlo

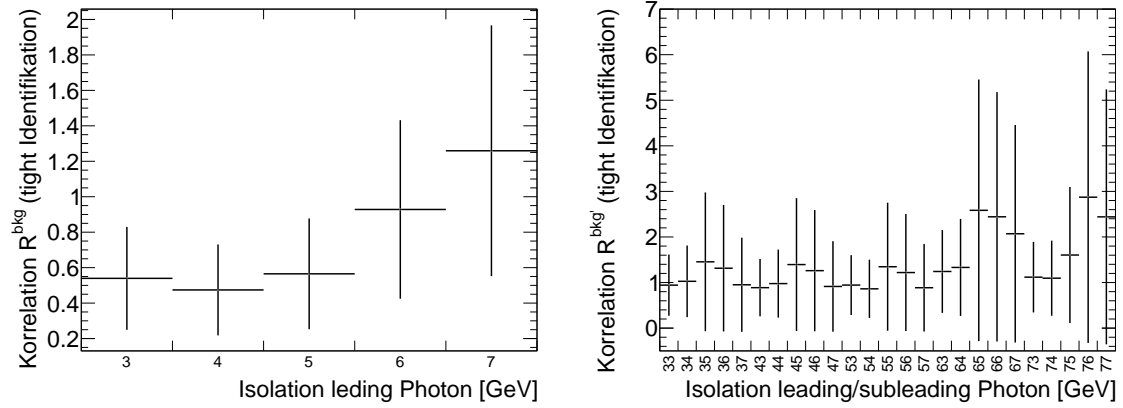


Abbildung 5.2: Korrelation zwischen Identifikation und Kalorimeter Isolation für das leading Photon (*links*) und das subleading Photon (*rechts*). Die Zahlenpaare auf der y-Achse geben die Isolationsbedingungen in GeV an, die für das leading und das subleading Photon gefordert werden und definieren somit die Signalregionen und die Kontrollregionen in der 2×2D Sideband Methode.

Studien werden die Monte Carlo Samples des dominanten $W + jets$ Untergrunds und Signal Monte Carlo Samples aus Abschnitt 2.4 zusammen genommen und alle Ereignisse werden nach obiger Formel gewichtet. Wird der Begriff Monte Carlo Sample im Folgenden verwendet, bezeichnet dies somit stets eine Mischung von simulierten Signal- und Untergrundereignissen.

Der relative Fehler der poissonverteilten Ereigniszahlen beträgt $\frac{\Delta N}{N} = \frac{1}{\sqrt{N}}$. Die Monte Carlo Unsicherheit durch die limitierte Anzahl von simulierten Ereignissen wird so gewählt, dass der relative Fehler nicht durch die Gewichtung verändert wird. Die Unsicherheit einer Menge von N Ereignissen ist wie folgt definiert [17]:

$$\Delta N = \sqrt{\sum_i w_i^2} \quad (5.68)$$

wobei i über alle Ereignisse in der betrachteten Menge summiert. Somit ist sichergestellt, dass die statistische Genauigkeit vor und nach der Gewichtung gleich ist.

5.3.2 Durchführung des Monte Carlo Closure Test

Um die 2×2D Sideband Methode anwenden zu können, müssen im nächsten Schritt alle Ereignisse im Monte Carlo Sample der Vorselektion unterzogen werden. Im Anschluss ist es möglich, an das leading und das subleading Photon in den rekonstruierten Ereignissen die zusätzliche Identifikations- und Isolationsbedingung zu stellen, die in der 2×2D Sideband Methode die Signalregion definieren. Alle oben eingeführten

Möglichkeiten der Signalregiondefinition werden getestet, indem jeweils die Anzahl der Ereignisse in Signal- und Untergrundregionen selektiert werden und mit den in Abschnitt 5.2 bestimmten Monte Carlo Input Parametern die Anzahl an Signal- und Untergrundereignissen mit Hilfe von Gleichung 5.60 bis 5.64 berechnet werden. Für den Closure Test werden zur Berechnung zunächst auch die in Abschnitt 5.2 bestimmten Korrelationen R^{bkg} und $R^{bkg'}$ verwendet, da es vorrangig darum geht die Implementierung der Methode zu überprüfen. Da die Monte Carlo Input Parameter und die Ereignisse in den einzelnen Feldern der $2 \times 2D$ Sideband Methode von den gleichen Monte Carlo Samples stammen, wird als Ergebnis des Closure Tests erwartet, dass die berechnete Zahl $W_{\gamma\gamma}^{TITI}$ an Signalereignissen mit der wahren Anzahl an Signalereignissen $W_{\gamma\gamma, true}^{TITI}$ in der Signalregion übereinstimmt⁸. Dies bestätigt sich in Abbildung 5.3, in der die erwartete und berechnete Anzahl, sowie deren Differenz dargestellt sind. Die Unsicherheiten stammen aus der analytischen Fehlerrechnung mit Gleichung 5.65, wobei als Einzelfehler der Größen N_B bis N_D' die oben definierten Monte Carlo Unsicherheit verwendet werden. Die großen Unsicherheiten resultieren aus der begrenzten Anzahl an simulierten Ereignissen, so dass eine Monte Carlo Vorhersage auch nur innerhalb einer relativ großen Unsicherheit möglich ist.

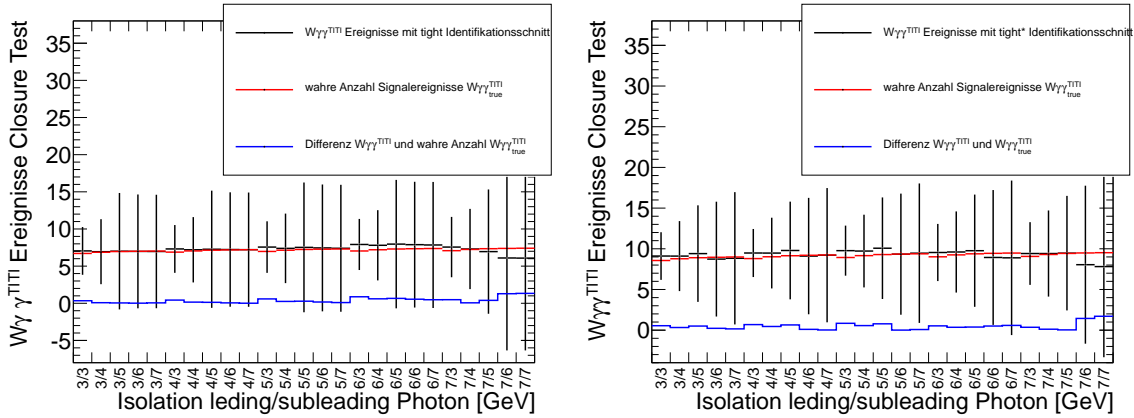


Abbildung 5.3: Dargestellt sind sowohl die berechnete als auch die wahre Anzahl an Signalereignissen in der TITI⁹ Region und deren Differenz für die verschiedenen Isolationsbedingungen. Zusätzliche Identifikationsbedingung *links: tight* ($R^{bkg} = 0.56$, $R^{bkg'} = 1.34$), *rechts: tight** ($R^{bkg} = 1.22$, $R^{bkg'} = 1.03$).

Wie erwartet führt die strengere *tight* Identifikationsbedingung dazu, dass weniger Signalereignisse in die TITI Region fallen, jedoch wird auch die Zahl an Untergrundereignissen in der Signalregion A' deutlich verringert. Es fällt auf, dass die berechnete Unsicherheit stark von der Isolationsbedingung des subleading Photons abhängt und weniger von der des leading Photon. Dies bedeutet der analytisch

⁸Die Anzahl der wahren Signalereignissen wird wieder mit $0 < |ph_truth_mothertype| < 26$ abgefragt

berechnete Fehler ist umso kleiner, je mehr Ereignisse in den Kontrollfeldern, insbesondere in B', C' und D' liegen.

Die Durchführung des Monte Carlo Closure Test zeigt, dass die $2 \times 2D$ Sideband Methode fehlerfrei implementiert wurde und bei Bekanntheit der Monte Carlo Input Parameter die Anzahl wahrer Signalereignisse in der Signalregion exakt bestimmt werden kann.

5.3.3 Einfluss der Korrelationen R^{bkg} und $R^{bkg'}$ auf den Closure Test

Wie oben erwähnt werden bei Anwendung der $2 \times 2D$ Sideband Methode auf Daten die Korrelationen $R^{bkg} = R^{bkg'} = 1$ gesetzt und mögliche Abweichungen gehen in den systematischen Fehler ein. Im Folgenden wird der Closure Test mit $R^{bkg} = R^{bkg'} = 1$ wiederholt um die Auswirkung der Korrelation auf das Ergebnis zu untersuchen. In Abbildung 5.4 sind die Ergebnisse der analytischen Berechnungen dargestellt.

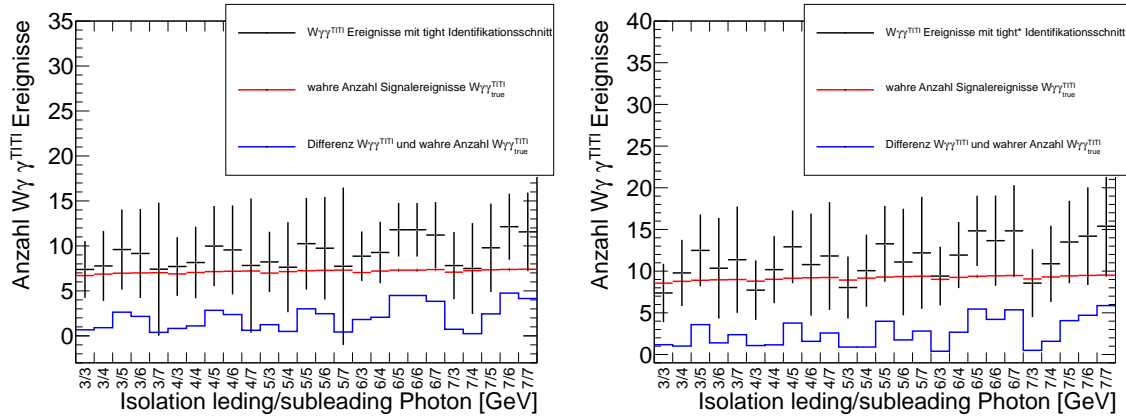


Abbildung 5.4: Dargestellt sind wieder berechnete und wahre Anzahl an Signalereignissen in der TITI Region sowie deren Differenz für die verschiedenen Isolationsbedingungen. Zusätzliche Identifikationsbedingung *links*: $tight$ ($R^{bkg} = R^{bkg'} = 1$), *rechts*: $tight^*$ ($R^{bkg} = R^{bkg'} = 1$).

Im Vergleich zum Closure Test zeigen sich nun für bestimmte Isolationsschnitte deutlich größere Abweichungen von berechneter und wahrer Anzahl an Signalereignissen in der TITI Region. Der Vergleich von Abbildung 5.2 und 5.4 zeigt, dass die Differenz für die Isolationsschnitte am kleinste ist, bei denen die Korrelationen R^{bkg} und $R^{bkg'}$ am wenigsten von eins abweichen. Um die Abweichung der berechneten Signalereigniszahlen zu höheren Ergebnissen zu verstehen, muss der Zusammenhang von $W_{\gamma\gamma}^{TITI}$ und R^{bkg} beziehungsweise $R^{bkg'}$ analysiert werden. In Abbildung 5.5 ist $W_{\gamma\gamma}^{TITI}$ in Abhängigkeit der beiden Korrelationen am Beispiel für die Isolationsbedingungen 5 GeV für beide Photonen und $tight$ Identifikation in der Signalregion dargestellt. Der Vergleich von Abbildung 5.4 und 5.3 zeigt, dass die Anzahl der

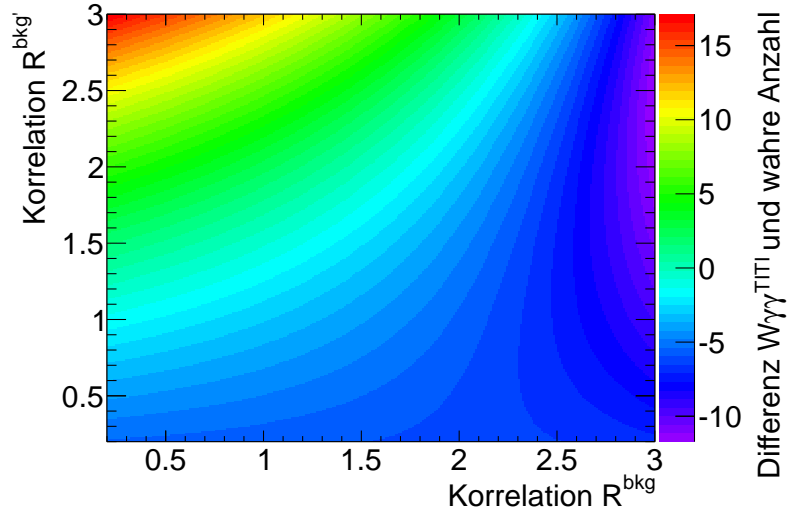


Abbildung 5.5: Abhängigkeit der berechneten Anzahl Signalereignisse $W_{\gamma\gamma}^{TITI}$ von R^{bkg} und $R^{bkg'}$ am Beispiel des Isolationsschnitt 5/5 GeV.

berechneten $W_{\gamma\gamma}^{TITI}$ von 7.5 auf 10.2 steigt falls $R^{bkg} = R^{bkg'} = 1$ gesetzt werden. Betrachtet man den Verlauf der Korrelationen in Abbildung 5.2, fällt auf dass, im Closure Test $R^{bkg} = 0.56$ und $R^{bkg'} = 1.34$ verwendet wurde. In Abbildung 5.5 ist nun deutlich zu sehen, wie die Erhöhung von R^{bkg} in Kombination mit der Herabsetzung von $R^{bkg'}$ auf eins zu der Erhöhung der berechneten $W_{\gamma\gamma}^{TITI}$ Ereigniszahl führt. Außerdem sieht man in Abbildung 5.5, dass sich kleine Änderungen von $R^{bkg'}$ im Bereich $R^{bkg} = R^{bkg'} = 1$ sich stärker als Änderungen von R^{bkg} auf das Rechenergebnis auswirken.

Mit der Durchführung dieser Studie zur Untersuchung des Einflusses einer möglichen Korrelation von Identifikationsschnitt und Kalorimeter Isolation wird gezeigt, dass sich in diesem Falle mit der $2 \times 2D$ Sideband Methode zwar größere Abweichungen von berechneter und wahrer Anzahl an Signalereignissen ergeben, jedoch liegt immer noch eine gute Übereinstimmung vor und im Rahmen der Unsicherheit sind die berechneten Werte mit der wahren Anzahl an Signalereignissen in der Signalregion konsistent.

5.4 Signal- und Untergrund- Wahrscheinlichkeitsdichteverteilungen

In der Teilchenphysik sind Vorhersagen immer nur im Rahmen einer statistischen Unsicherheit möglich, da es sich bei den beobachteten Größen um poissonverteilte Zufallsgrößen handelt. Dies bedeutet, dass bei mehrmaliger Wiederholung des Experiments nicht immer die gleiche Anzahl an Ereignissen in den definierten Feldern gemessen wird. Stattdessen ergibt sich eine Häufigkeitsverteilung mit Erwartungswert

N und Varianz ΔN . Normiert man diese Verteilung auf die Fläche eins, erhält man eine Wahrscheinlichkeitsdichtefunktion, kurz PDF¹⁰, welche die Wahrscheinlichkeit angibt, mit der ein bestimmter Wert $W_{\gamma\gamma}^{TITI}$ im Intervall $\Delta W_{\gamma\gamma}^{TITI}$ bei einmaliger Durchführung des Experiments gemessen wird.

Im Folgenden wird nun die Wahrscheinlichkeitsdichtefunktion der berechneten Anzahl an Signalereignissen $W_{\gamma\gamma}^{TITI}$ und Untergrundereignissen W_{bkg}^{TITI} in der Signalregion bestimmt, indem als beste Schätzwerte für die Erwartungswerte der Ereigniszahlen in allen Feldern 2×2D Sideband Methode die in Monte Carlo Studien selektierten Ereigniszahlen verwendet werden. Dafür werden zwei verschiedene Methoden implementiert und verglichen:

- Analytische Berechnung des Erwartungswertes $W_{\gamma\gamma}^{TITI}$ und des Untergrunds W_{bkg}^{TITI} mit Gleichung 5.60 - 5.64 und Bestimmung der Breite der Verteilung durch analytischer Fehlerrechnung
- Bestimmung der $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} PDF mit Hilfe von sogenannten Toy Monte Carlo Studien

Der Einfluss der Monte Carlo Ungenauigkeit aus Abschnitt 5.3.2, welcher die Genauigkeit dieser Erwartungswerte einschränkt, wird später diskutiert.

5.4.1 Analytische Bestimmung der Wahrscheinlichkeitsdichteverteilungen

Mit Gleichung 5.60 bis 5.64 werden die Anzahl der $W_{\gamma\gamma}^{TITI}$, $W_{\gamma j}^{TITI}$, $W_{j\gamma}^{TITI}$ und W_{jj}^{TITI} Ereignisse in der Signalregion ausgerechnet, indem die selektierten Monte Carlo Ereignisanzahlen und Monte Carlo Input Parameter eingesetzt werden. Der errechnete $W_{\gamma\gamma}^{TITI}$ Wert wird als Erwartungswert der $W_{\gamma\gamma}^{TITI}$ Verteilung angenommen. Die Untergrund PDF setzt sich aus der Summe von $W_{\gamma j}^{TITI}$, $W_{j\gamma}^{TITI}$ und W_{jj}^{TITI} zusammen. Zur Bestimmung der Breite der PDF wird die analytische Fehlerformel 5.65 verwendet. Dazu wird als Einzelfehler für jeden Eintrag in den 7 unabhängigen Feldern der 2×2D Sideband Methode die Wurzel der Varianz einer Poissonverteilung¹¹ als symmetrische Standardabweichung verwendet. Die Gaußsche Fehlerfortpflanzung 5.65 ergibt somit auch einen symmetrischen Fehler für $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} , welcher als Breite einer gaußverteilten Wahrscheinlichkeitsdichtefunktion interpretiert wird (siehe Abbildung 5.6). Die Annahme der Wurzel der Varianz als symmetrische Standardabweichung ist für poissonverteilte Größen mathematisch nur für hohe Erwartungswerte gültig, da dann die Poissonverteilung in eine Gaußverteilung mit Erwartungswert N und Breite $\sigma = \sqrt{N}$ übergeht.

¹⁰Von der englischen Übersetzung **P**robability **D**ensity **F**unction

¹¹ Dies bedeutet bei N selektierten Ereignissen beträgt der Fehler $\Delta N = \sqrt{N}$

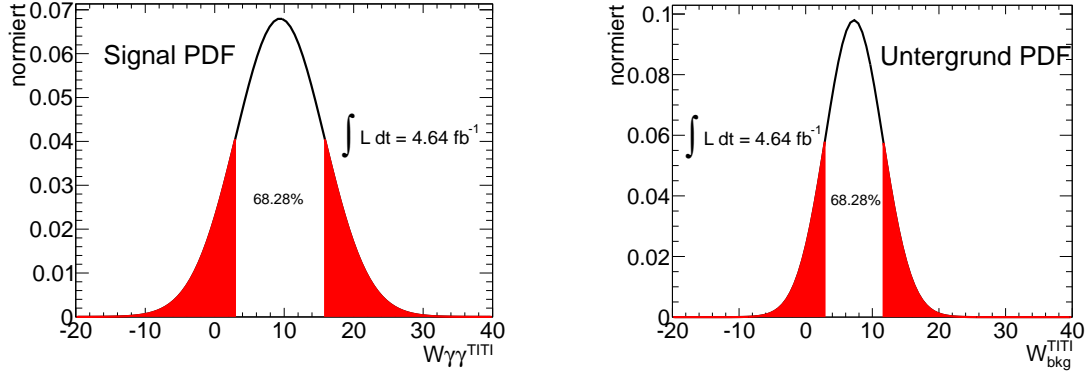


Abbildung 5.6: Wahrscheinlichkeitsdichtefunktion für Signal (*links*) und Untergrund (*rechts*) bei *tight** Identifikationsbedingung und 6/3 GeV Isolationsbedingung für leading bzw. subleading Photon. Die PDF basieren auf analytischer Rechnung. Die rot gefärbten Bereiche entsprechen jeweils 15.86% der Gesamtfläche der PDF.

5.4.2 Bestimmung der Wahrscheinlichkeitsdichteverteilungen mit Toy Monte Carlo

Bei den hier vorliegenden niedrigen Anzahl an selektierten Ereignissen wird eine alternative Fehlerabschätzung mit Hilfe von Toy Monte Carlo implementiert. Hierbei werden die $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} Berechnung 1000 mal durchgeführt, wobei bei jeder Rechnung die Ereignissanzahlen N_B bis N'_D in den unabhängigen Feldern B bis D' im Rahmen einer Poissonverteilung zufällig variiert werden. Dadurch ergibt sich zwei Verteilungen, welche durch Flächennormierung auf eins die Wahrscheinlichkeitsdichtefunktion von Signal und Untergrund darstellen. Um den Maximalwert der Verteilung zu bestimmen, wird bei der Signal PDF eine Gaußfunktion und bei der Untergrund PDF eine Landaufunktion um das Maximum gefittet. Da die Verteilungen an den Rändern nicht durch die Fitfunktionen beschrieben werden können, wird zur Bestimmung des 1σ Vertrauensbereich der Bereich bestimmt, indem 68.28% der Gesamtfläche liegen. Dazu werden iterativ von beiden Rändern der Verteilung die Bereiche bestimmt, deren Integral jeweils 15.86% der Gesamtfläche ausmachen. Da die Verteilungen nicht symmetrisch sind, ergeben sich asymmetrische Fehlerintervalle. Es ergibt sich, dass bei der Signal PDF der größere Fehler zu kleineren Werten zeigt, wogegen bei der Untergrund PDF mehr Fluktuationen nach oben auftreten. Außerdem stimmen Maximum und Erwartungswert der Verteilungen aufgrund der Asymmetrie nicht überein. Um zu verstehen, warum die Maxima der beiden Verteilungen nicht mit der gleichen Wahrscheinlichkeit angenommen werden, sind in Abbildung 5.8 die Signal- und Untergrundverteilung zweidimensional dargestellt. Die Berechnung einer $W_{\gamma\gamma}^{TITI}$ Signalanzahl ist nicht eindeutig einer Anzahl an Untergrundereignissen zuzuordnen, sondern es gibt verschiedene mögliche Kombinationen, die unterschiedlich oft eintreten. Da die Signal PDF eine größere

Breite aufweist, können nur weniger Ergebnisse den Maximalwert annehmen.

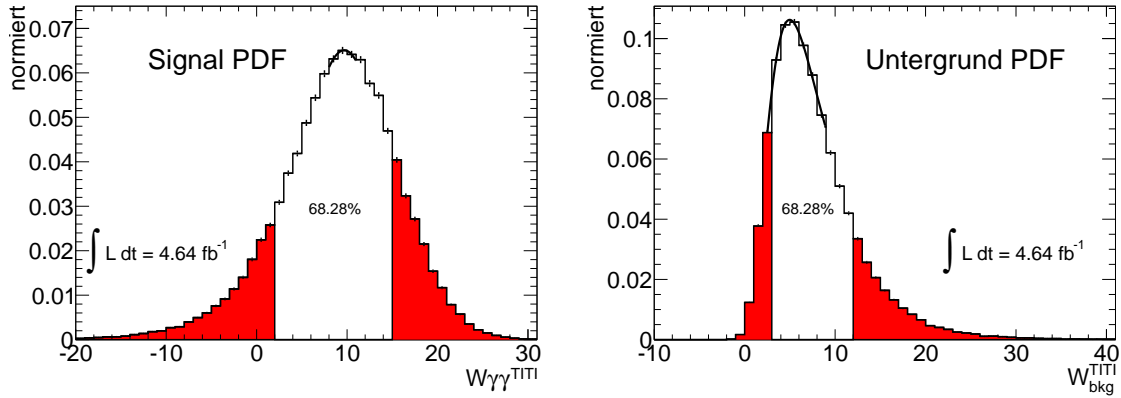


Abbildung 5.7: PDF Signal (*links*) und Untergrund (*rechts*), welche auf Toy Monte Carlo Studien basieren, wieder bei *tight** Identifikationsbedingung und 6/3 GeV Isolationsbedingung für leading bzw. subleading Photon.

5.4.3 Vergleich von analytischer Berechnung und Toy Monte Carlo

Mit den oben bestimmten Wahrscheinlichkeitsdichtefunktionen lässt sich für alle definierten Isolationsschnitte und beide eingeführten Identifikationsbedingungen das wahrscheinlichste Resultat der $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} Bestimmung und der 1σ Vertrauensbereich angegeben. Da die durch analytische Rechnung bestimmte PDF symmetrisch ist, fallen das Maximum und der Erwartungswert der Verteilung zusammen. Addiert man die wahrscheinlichsten Signal- und Untergrundereigniszahlen, ergibt sich als Summe die Anzahl der Ereignisse N'_A im Signalfeld A' . Somit ist zu erklären, dass in Abbildung 5.9 zu hoch berechnete Werte bei der Signalanzahl in der linken Abbildung durch eine zu niedrig berechnete Anzahl Untergrundereignissen (rechts) ausgeglichen wird.

In der Toy Monte Carlo Studie liegen die wahrscheinlichsten $W_{\gamma\gamma}^{TITI}$ Ergebnisse für alle Isolationsschnitte höher als die Ergebnisse der analytischen Berechnung, wogegen die Maxima der Untergrundverteilung bei der Toy Monte Carlo Methode deutlich unter den Resultaten der analytischen Berechnung liegen. Um diese Tatsache zu verstehen, müssen die Maximalwerte der PDF in Zusammenhang mit den Breiten und somit dem angegebenen Fehlerbereich gesehen werden. Oben wurde definiert, dass die Toy Monte Carlo Methode als Ergebnis den Bereich angibt, in dem 68.28% der Ergebnisse um den wahrscheinlichsten Wert der PDF liegen. Werden an das leading Photon eine strenge Isolationsbedingung und an das subleading Photon eine großzügigere Isolationsbedingung gestellt, führt dies dazu, dass insbesondere in den Feldern B' und D' nur sehr wenige Ereignisse liegen. So befinden sich bei einer

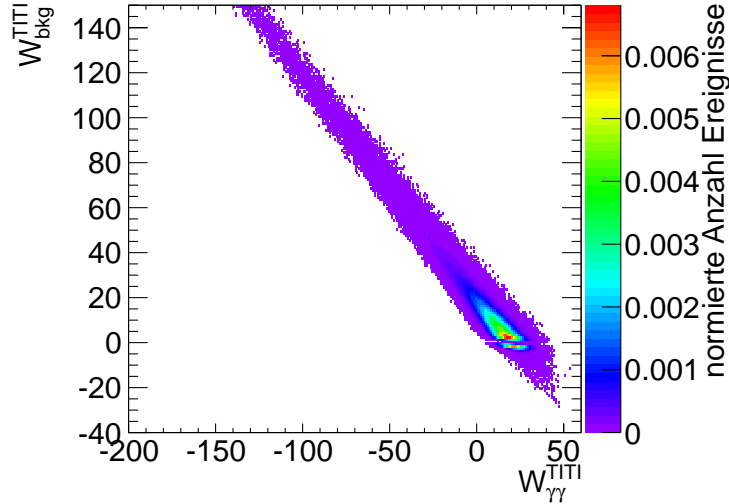


Abbildung 5.8: Zweidimensionale Darstellung von berechneter Anzahl an Signal- und Untergrundereignissen für jeden Durchgang in der Toy Monte Carlo Studie. Die Signalregion wird durch eine *tight** Identifikation und und 3/7 GeV Isolationsbedingung definiert.

3/7 GeV Kombination nur 2 Ereignisse in Feld B', so dass eine kleine Variation eine große absolute Änderung mit sich bringt. Es resultiert eine breite Wahrscheinlichkeitsdichteverteilung für den Untergrund, so dass der Maximalwert für sich alleine keine große Aussagekraft hat. In Abbildung 5.9 ist zu sehen, dass bei der 3/7 GeV Isolationsbedingungskombination zwar der Maximalwert der Signal PDF niedriger als der analytisch berechnete Wert ist, jedoch ist in Abbildung 5.8 eine breitere Ausdehnung zu höheren Werten sichtbar. Bei der Signalverteilung ist das Verhalten genau entgegengesetzt: Im Vergleich zur analytischen Berechnung ein größerer Maximalwert und dafür eine Tendenz zu kleineren Werten. Die Asymmetrie der Wahrscheinlichkeitsdichteverteilungen resultiert aus der Poissonverteilung der Ereigniszahlen. Bei dieser sind keine negativen Werte möglich, so dass bei niedrigen Erwartungswerten eine asymmetrische Verteilung vorliegt und Fluktuationen nach oben wahrscheinlicher sind als nach unten. Dies wirkt sich auch bei der $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} Berechnung aus, so dass asymmetrische PDFs entstehen. Durch die Bestimmung der statistischen Unsicherheit mit Hilfe von Toy Monte Carlo Studien wird erreicht, dass diese bei niedriger Statistik auftretende Asymmetrie in der Fehlerabschätzung berücksichtigt werden kann.

Toy Monte Carlo bestimmte Ereigniszahlen sind im Folgenden stets der Form: Maximalwert PDF $\pm 68.28\%$ Vertrauensbereich. Es ist aus oben erläuterten Gründen *nicht* verwunderlich, wenn die Summe von Maximalwert Signal- und Untergrundverteilung nicht wie bei der analytischen Berechnung der Ereigniszahl N_A' im Signalfeld A' entspricht. Für die Summe der Erwartungswerte der Verteilungen ist dies jedoch gültig, was in Abbildung 5.10 zu sehen ist und im Anhang rechnerisch begründet

wird.

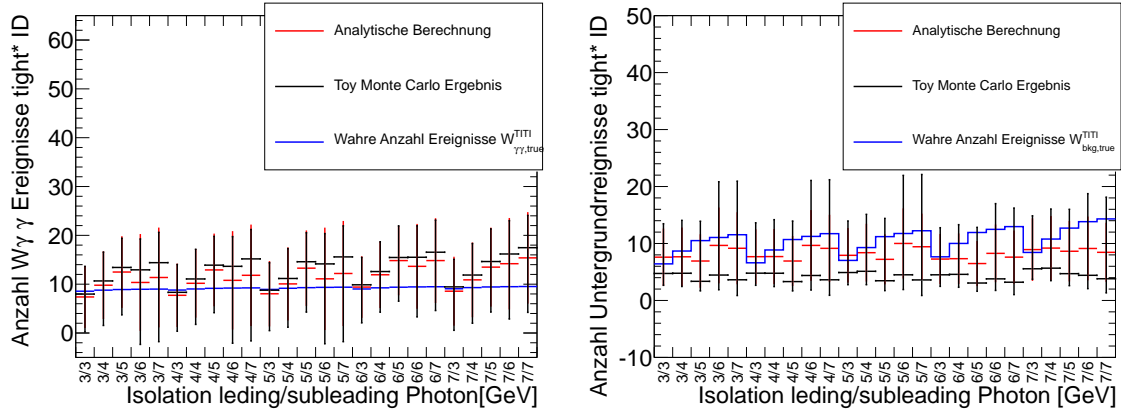


Abbildung 5.9: Vergleich von analytisch und mit Toy Monte Carlo Methode bestimmte Maximalwerte der Signal- (*links*) und Untergrund PDF (*rechts*) sowie der wahren Anzahl selektierter Ereignisse.

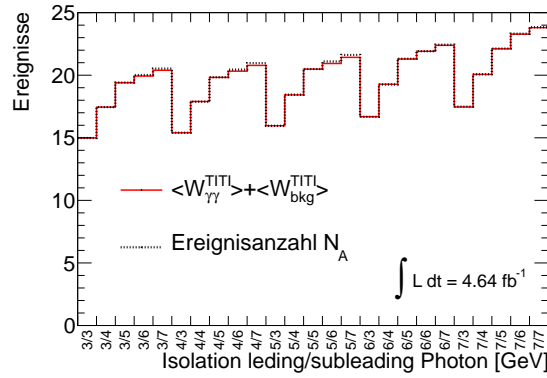


Abbildung 5.10: Vergleich von der Summe der Erwartungswerte von Signal- und Untergrund PDF mit der Anzahl an Ereignissen N'_A .

5.4.4 Vergleich von analytischer Berechnung und Toy Monte Carlo bei hoher Statistik

In diesem Abschnitt wird der Vergleich der Wahrscheinlichkeitsdichtefunktionen für Signal und Untergrund bei höherer Statistik wiederholt. Mit obiger Argumentation wird erwartet, dass die mit der Toy Monte Carlo Methode bestimmten Verteilungen sich mit höheren Erwartungswerten für die Feldeinträge den gaußförmig analytisch bestimmten Verteilungen annähern. Um den Vergleich durchführen zu können wird das Monte Carlo auf zunächst auf 20 fb^{-1} und anschließend um den Faktor 100 auf 464 fb^{-1} skaliert. In Tabelle 5.2 sind die Ergebnisse des Vergleichs bei einer

Signalregiondefinition von *tight** Identifikation und 6/3 GeV Isolationsbedingung dargestellt. Es wird deutlich, dass die Abweichung von analytischer Berechnung und

$\int L dt$	$4.64 fb^{-1}$	$20 fb^{-1}$	$464 fb^{-1}$
Signalereignisse			
- analytische Methode	9.4 ± 6.4	40.0 ± 13.1	940.7 ± 63.48
- Toy Monte Carlo Methode	$9.7^{+5.6}_{-7.1}$	$39.6^{+13.0}_{-13.6}$	$941.2^{+62.8}_{-63.6}$
Untergrundereignisse			
- analytische Methode	7.3 ± 4.3	30.9 ± 8.8	726.7 ± 42.8
- Toy Monte Carlo Methode	$5.0^{+7.1}_{-1.6}$	$28.4^{+11.8}_{-6.0}$	$727.2^{+44.0}_{-41.9}$

Tabelle 5.2: Anzahl der bestimmten Signal- und Untergrundereignisse in Abhängigkeit der gewählten Luminosität.

Toy Monte Carlo ein Effekt niedriger Statistik ist. Bei ausreichend hoher Statistik ergibt sich aus der Toy Monte Carlo Methode eine symmetrische PDF für Signal- und Untergrundereignisse, deren Breite und Maximalwert mit der analytische bestimmten PDF übereinstimmt. Ausreichend hohe Statistik liegt bei der Signal PDF schon bei der im Jahr 2012 erwarteten integrierten Luminosität von $20 fb^{-1}$ vor, wobei die Untergrund PDF noch asymmetrische Fehler ausweist. Bei einer integrierten Luminosität von $464 fb^{-1}$ ist auch die Untergrund PDF nahezu symmetrisch, was in Abbildung 5.11 zu sehen ist.

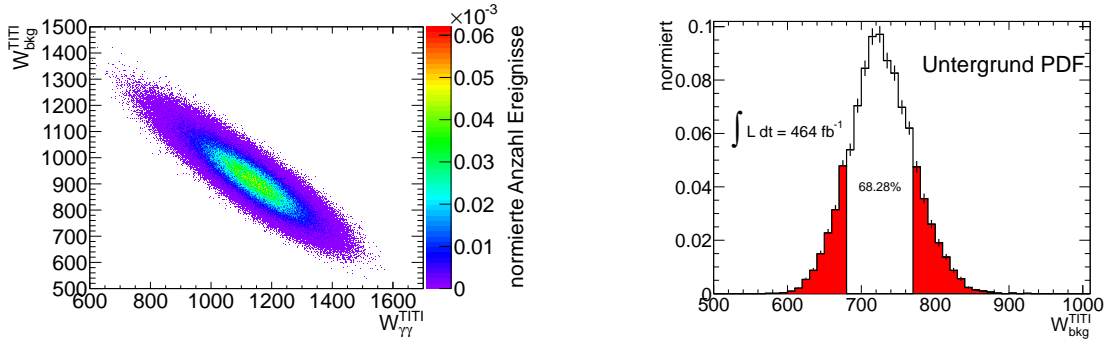


Abbildung 5.11: Bei hoher Statistik liegen auch für die Isolationsschnitte 3/7 GeV symmetrische Signal- und Untergrundverteilungen vor. Dargestellt sind die Verteilungen bei einer integrierten Luminosität von $464 fb^{-1}$. Rechts: Untergrund PDF bei hoher Statistik (*tight** Identifikation und 6/3 GeV Isolationsbedingung).

5.5 Einfluss der Monte Carlo Unsicherheit auf die Vorhersage des relativen Fehlers

In Abschnitt 5.3 wurde bereits erläutert, dass aufgrund von begrenzter Monte Carlo Statistik Vorhersagen nur innerhalb einer Unsicherheit möglich sind. Für die Bestimmung der Wahrscheinlichkeitsdichtefunktionen werden als Schätzwerte für die Anzahl an Ereignissen in den Feldern der 2×2 D Sideband Methode die Ereigniszahlen verwendet, die in den Monte Carlo Samples selektiert werden. Zur Untersuchung des Einflusses der Monte Carlo Unsicherheit auf den verwendeten Schätzwert und somit auf den statistischen Fehler der Berechnung der Signal- und Untergrundereignisse wird eine weitere Toy Monte Carlo Studie durchgeführt. Dazu wird die Toy Monte Carlo Studie aus Abschnitt 5.4.2 10000 mal durchgeführt, wobei bei jeder Ausführung die Schätzwerte für die Feldeinträge im Rahmen der Monte Carlo Unsicherheit variiert werden. Somit kann untersucht werden, welchen Einfluss es auf die Breite der $W_{\gamma\gamma}^{TITI}$ Verteilung hat, falls als beste Schätzwerte nicht genau die selektierten Ereigniszahlen verwendet werden, sondern Werte, die innerhalb der Monte Carlo Unsicherheit verträglich sind. In Abbildung 5.12 ist für jede Kombination von Schätzwerten der relative Fehler der $W_{\gamma\gamma}^{TITI}$ Bestimmung eingetragen. Der relative Fehler wird als Quotient von Breite und Maximum der resultierenden Toy Monte Carlo PDF berechnet. Abbildung 5.12 ist auf die Fläche 1 normiert. Zum Vergleich wurde der Test bei hoher Statistik wiederholt, wobei deutlich wird, dass die breite Verteilung des relativen Fehlers nur bei niedriger Statistik auftritt.

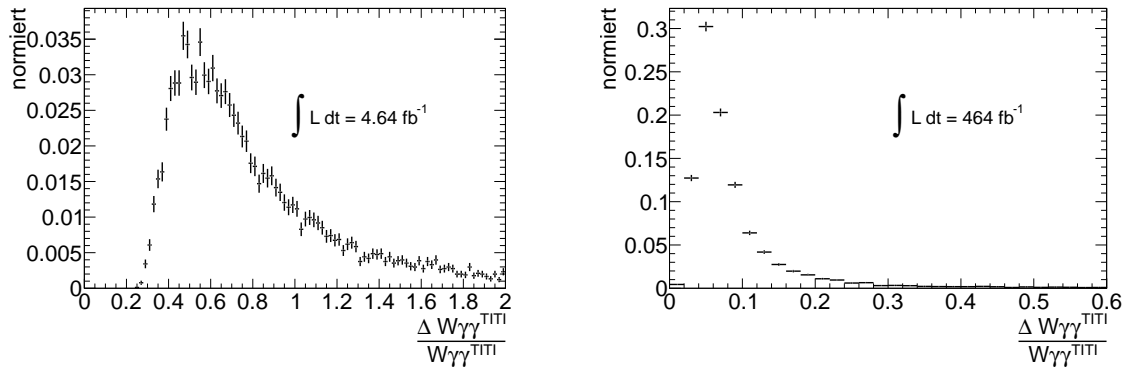


Abbildung 5.12: Verteilung des erwarteten relativen Fehlers für die $W_{\gamma\gamma}^{TITI}$ Berechnung bei integrierter Luminosität von 4.64 fb^{-1} *links* und 464 fb^{-1} *rechts*.

5.6 Studie des systematischen Fehlers

Am Beispiel der Signalregiondefinition *tight** Identifikationsschnitt und Kalorimeter Isolation $E_T^{iso} < 6$ GeV für das leading Photon und $E_T^{iso} < 3$ GeV für das subleading Photon wird in folgendem Abschnitt die Bestimmung des systematische Fehlers diskutiert, welcher sich aus den in der 2×2D Sideband Methode benötigten Monte Carlo Input Parametern ergibt. Dazu wird wieder sowohl eine analytische Fehlerrechnung als auch eine auf Toy Monte Carlo basierende Methode eingeführt.

5.6.1 Analytische Bestimmung des systematischen Fehlers

Zunächst wird der Einfluss der einzelnen Monte Carlo Input Parameter auf das Ergebnis der Berechnung von Signal und Untergrundereignissen mit Gleichungen 5.60 bis 5.64 bestimmt, indem jeder Parameter einzeln um seine Monte Carlo Unsicherheit erhöht und erniedrigt wird. Anschließend wird mit dem variierten Parameter die Signal- und Untergrundberechnung durchgeführt. Die Ergebnisse sind in Tabelle 5.3 dargestellt. Da die Gleichungen 5.60 bis 5.64 nicht linear von den Input Parametern abhängen, ergeben sich unterschiedliche Abweichungen für die Variationen der Parametern zu höheren und tieferen Werten.

<i>tight*</i> Identifikation, 6/3 Isolation			systematischer Fehler	
Parameter	verwendeter Wert	MC Unsicherheit	$W_{\gamma\gamma}^{TITI}$	W_{bkg}^{TITI}
R^{bkg}	1	$\pm 26\%$	$\begin{pmatrix} +0.4 \\ -0.4 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.4 \\ -0.4 \end{pmatrix}^{sys}$
$R^{bkg'}$	1	$\pm 31\%$	$\begin{pmatrix} +2.4 \\ -2.7 \end{pmatrix}^{sys}$	$\begin{pmatrix} +2.7 \\ -2.4 \end{pmatrix}^{sys}$
α	0.44	$\pm 27\%$	$\begin{pmatrix} +1.1 \\ -1.4 \end{pmatrix}^{sys}$	$\begin{pmatrix} +1.4 \\ -1.1 \end{pmatrix}^{sys}$
c_1	0.025	$\pm 50\%$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$
c_2	0.106	$\pm 50\%$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$
c_3	0.003	$\pm 50\%$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$
c'_1	0.068	$\pm 50\%$	$\begin{pmatrix} +0.5 \\ -0.5 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.5 \\ -0.5 \end{pmatrix}^{sys}$
c'_2	0.155	$\pm 50\%$	$\begin{pmatrix} +0.4 \\ -0.4 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.4 \\ -0.4 \end{pmatrix}^{sys}$
c'_3	0.015	$\pm 50\%$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$	$\begin{pmatrix} +0.1 \\ -0.1 \end{pmatrix}^{sys}$

Tabelle 5.3: Wie oben schon erläutert wird, $R^{bkg} = R^{bkg'} = 1$ gesetzt und die ermittelte Unsicherheit aus Abschnitt 5.2 für die Bestimmung des systematischen Fehlers verwendet. Die ermittelten Unsicherheiten der c_i und c'_i sind deutlich kleiner als 50%, jedoch wird konservativ vorgegangen und großzügig variiert.

Der dominante Fehler geht von der großen Ungenauigkeit des Parameters $R^{bkg'}$ aus. Leider ist im Rahmen der begrenzten Monte Carlo Statistik keine genauere Vorhersage möglich. Es fällt auf, dass eine Erhöhung der Zahl an Signalereignissen durch

eine Parametervariation zu einer Erniedrigung der Anzahl an Untergrundereignissen im gleichen Maße führt, was dadurch zu erklären ist, dass die Gesamtanzahl an Ereignissen erhalten bleibt. In einer ersten analytischen Bestimmung des systematischen Gesamtfehlers wird die Gaußsche Fehlerfortpflanzung von Gleichung 5.60 und 5.64 angewandt:

$$\Delta W_{XX}^{TITI} = \sqrt{\sum_i \left(\frac{\partial W_{XX}^{TITI}}{\partial \chi_i} \Delta \chi_i \right)^2} \quad (5.69)$$

wobei χ für alle verwendeten Monte Carlo Input Parameter steht.

Das Ergebnis der analytischen Berechnung des systematischen Fehlers¹² beträgt:

$$W_{\gamma\gamma}^{TITI} = 9.4 \pm 4.3^{sys} \quad (5.70)$$

$$W_{bkg}^{TITI} = 7.3 \pm 4.3^{sys} \quad (5.71)$$

Voraussetzung für die Anwendung von Formel 5.69 ist, dass sich die Funktionen W_{XX}^{TITI} sich im Bereich ihrer Unsicherheit annähernd linear verhält, was bei Betrachtung von Tabelle 5.3 nicht für alle Parameter der Fall ist.

5.6.2 Bestimmung des systematischen Fehlers mit Toy Monte Carlo

Alternativ zur analytischen Fehlerrechnung wird analog zur Abschnitt 5.4.2 eine Toy Monte Carlo Studie durchgeführt. Dazu werden zunächst die Parameter einzeln variiert, wodurch sich eine asymmetrische PDF ergibt. Zur Fehlerbestimmung wird wieder iterativ von beiden Rändern der normierten Verteilungen der Bereich bestimmt, in dem 15.86% der Gesamtfläche liegen. Somit liegen im angegebenen Intervall 68.28% aller Ergebnisse um den Maximalwert der PDF. Die Resultate finden sich in Tabelle 5.4.

Zur Bestimmung des systematischen Gesamtfehlers mit der Toy Monte Carlo Methode werden alle Monte Carlo Input Parameter in einer einzigen Toy Monte Carlo Studie variiert. In Abbildung 5.6.2 ist deutlich zu sehen, wie die Verteilung von Signal und Untergrundereignissen durch Spiegelung einander über gehen. Dies wird erwartet, da die Summe von Signal und Untergrund eine Konstante, nämlich die Anzahl an Ereignissen in Feld A', darstellt. Eine Erhöhung der berechneten Signalanzahl führt daher zu einer Erniedrigung der Untergrundberechnung im gleichen Maße. Somit folgen für die systematischen Gesamtfehler von Signal und Untergrund¹³:

$$W_{\gamma\gamma}^{TITI} = 9.7^{(+1.8)}_{(-4.5)}^{sys} \quad (5.72)$$

$$W_{bkg}^{TITI} = 5.0^{(+4.5)}_{(-1.8)}^{sys} \quad (5.73)$$

¹²Die Werte für $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} stammen in diesem Fall auch aus analytischen Rechnung.

¹³Hier stammen die Werte für $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} aus Toy MC Studien.

<i>tight</i> * Identifikation, 6/3 Isolation	systematischer Fehler	
Parameter	$W_{\gamma\gamma}^{TITI}$	W_{bkg}^{TITI}
R^{bkg}	$\begin{pmatrix} +0.6 \\ -0.3 \end{pmatrix}_{sys}$	$\begin{pmatrix} +0.2 \\ -0.9 \end{pmatrix}_{sys}$
$R^{bkg'}$	$\begin{pmatrix} +2.2 \\ -2.9 \end{pmatrix}_{sys}$	$\begin{pmatrix} +2.9 \\ -2.2 \end{pmatrix}_{sys}$
α	$\begin{pmatrix} +0.7 \\ -2.2 \end{pmatrix}_{sys}$	$\begin{pmatrix} +2.4 \\ -0.6 \end{pmatrix}_{sys}$
$c_1-c'_3$	$\begin{pmatrix} +0.7 \\ -1.5 \end{pmatrix}_{sys}$	$\begin{pmatrix} +1.4 \\ -0.7 \end{pmatrix}_{sys}$

Tabelle 5.4: Ergebnisse der Toy Monte Carlo Studie, für einzeln variierte Parameter. Es wurden die gleichen Werte und Unsicherheiten wie in Tabelle 5.3 verwendet. Die Variation der Parameter c wurde zusammengefasst, da die Variation eines einzigen c zu einer Verteilung führt, welche zu schmal ist um mit der verwendeten Methode ausgewertet zu werden.

5.7 Zusammenfassung der Monte Carlo Studie und Entdeckungswahrscheinlichkeit

Die oben beschriebene Studien der $2 \times 2D$ Sideband Methode bei niedrigerer Statistik zeigen, dass bei der Wahl der Signalregionendefinitionen A und A' folgende Aspekte beachtet werden müssen:

- möglichst geringe Isolationsbedingung für das subleading Photon, da der statistische Fehler mit steigender Isolationsbedingung steigt (siehe Abbildung 5.9)
- keine Isolationsbedingung, für die eine Ereigniszahl in einem der Felder nahe bei 0 erwartet wird (Abbildung 5.8)
- möglichst hohe Gesamtanzahl an Ereignissen, da der relative Fehler mit steigender Anzahl an Ereignissen sinkt und Wahrscheinlichkeitsdichteverteilungen zunehmend symmetrisch werden (Tabelle 5.2, Abbildung 5.11)
- hohe Gesamtanzahl an Ereignissen, um Monte Carlo Input Parameter mit geringer Unsicherheit bestimmen zu können (Tabelle 5.3 und 5.4).

Um die unteren beiden Punkte zu erfüllen, bietet sich eine *loose* Vorselektion und eine *tight** Identifikationsbedingung zur Signalregiondefinition an. Bei einer *tight** Identifikation zeigt die durchgeführte Studie für die Isolationsbedingungen 6 GeV für das leading und 3 GeV für das subleading Photon die vielversprechendsten Ergebnisse, da hier auch die oberen beiden Punkten erfüllt sind. Daher werden in folgendem Abschnitt die Ergebnisse für diese Signalregiondefinition zusammengefasst und anschließend mit Ergebnissen verglichen, welche aus Daten ermittelt werden. Um der momentan üblichen Identifikationsbedingung zu folgen, bietet sich bei einer *tight* Bedingung für die Signalregion eine Isolationsbedingung von 4 GeV für das

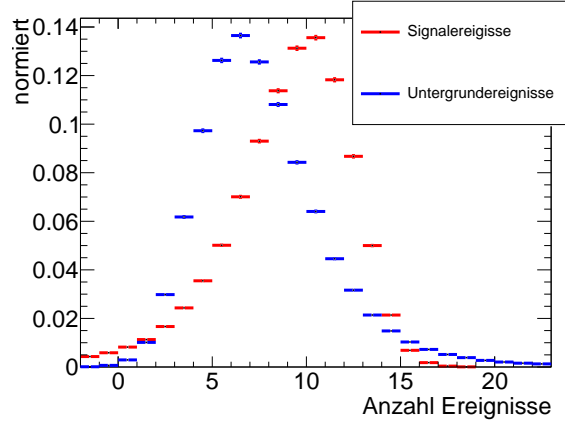


Abbildung 5.13: Dargestellt sind die PDF für Signal und Untergrund, welche sich in einer Toy Monte Carlo Studie aus der Variation der Monte Carlo Input Parameter ergibt. Als Signalregiondefinition ist in diesem Beispiel eine *tight** Identifikation und eine 6/3 GeV Isolationsbedingung gewählt.

leading und 3 GeV für das subleading Photon an. Die Ergebnisse dieser Variante werden im Anhang aufgeführt. Außerdem wird mit Hilfe der Untergrund PDF eine Wahrscheinlichkeit gegeben, mit der die selektierte Ereigniszahl in der Signalregion ausschließlich aus Untergrundereignissen besteht.

5.7.1 Monte Carlo Ergebnisse

In diesem Abschnitt sind die Ergebnisse der Toy Monte Carlo Studien für die in vorherigem Abschnitt gewählte Signalregiondefinition zusammengefasst. In Tabelle 5.5 sind die auf dem Monte Carlo Sample selektierten Ereigniszahlen in den Feldern der 2×2 D Sideband Methode und ihre Monte Carlo Unsicherheit aufgelistet. Diese Werte dienen wie in Abschnitt 5.4.2 beschrieben zur Bestimmung der Signal- und Untergrundwahrscheinlichkeitsdichteverteilungen. Aus diesen PDFs wurden die Signal- und Untergrundereigniszahlen aus Tabelle 5.7 ermittelt, wobei die benötigten Monte Carlo Input Parameter in Tabelle 5.6 zu finden sind. Da die Abschätzung des elektroschwachen Untergrunds ausschließlich auf Monte Carlo simulierten Ereignissen basiert, wird in diesem Abschnitt in Tabelle 5.8 die Anzahl elektroschwacher Untergrundereignisse gezeigt.

5.7.2 Entdeckungswahrscheinlichkeit in Abhängigkeit der Luminosität

Mit Hilfe der Wahrscheinlichkeitsdichteverteilungen für den Untergrund lässt sich eine Entdeckungswahrscheinlichkeit für den $W + \gamma + \gamma$ Prozess angeben. Die Berechnung eines offiziellen CL_s Limits [18] übersteigt den Rahmen dieser Bachelorarbeit,

<i>tight*</i> Identifikation, 6/3 Isolation		
Ereignisfeld	selektierter Wert	MC Unsicherheit
-in Monte Carlo-		
N_A	72.8	± 5.8
N_B	17.0	± 2.8
N_C	58.7	± 6.0
N_D	40.3	± 4.4
N'_A	16.7	± 1.8
N'_B	8.1	± 1.8
N'_C	19.9	± 2.9
N'_D	28.2	± 4.3

Tabelle 5.5: Selektierten Ereigniszahlen in den Feldern der 2×2 Sideband Methode und ihrer Monte Carlo Unsicherheit.

jedoch wird als vereinfachte Variante die Ausschlusswahrscheinlichkeit der Hypothese, dass sich in den gemessenen Daten nur Untergrundereignisse befinden, gegeben. Dazu wird als Null-Hypothese angenommen, dass sich die Signalregion A' ausschließlich aus Untergrundereignisse zusammensetzt. Die Wahrscheinlichkeitsdichteverteilung des Untergrunds enthält die Information, mit welcher Wahrscheinlichkeit eine Untergrundanzahl gemessen wird, die der gemessenen Ereigniszahl N'_A entspricht oder höher liegt (der so gennante p-Value). Dazu wird die PDF von $W_{bkg}^{TITI} = N'_A$ bis unendlich integriert.

Um sowohl die statistische als auch die systematische Unsicherheiten zu erhalten, wird für diese Berechnung die Untergrund PDF bestimmt, indem in einer Toy Monte Carlo Studie sowohl die Feldeinträge, als auch die Monte Carlo Input Parameter variiert werden. Mit der Variation der Feldeinträge wird die statistische Unsicherheit berücksichtigt, die Monte Carlo Input Parameter decken die systematische Unsicherheit ab. Die resultierende PDF für eine *tight** Identifikation und 6/3 GeV Isolation ist in Abbildung 5.14 dargestellt. Der ermittelte p-Value bei einer Gewichtung der Monte Carlo Samples auf eine integrierte Luminosität von 4.64 fb^{-1} lautet:

$$p = 0.112$$

Um eine Prognose für höhere integrierte Luminositäten zu geben, ist in Abbildung 5.15 *links* der Verlauf des erwarteten p-Values in Abhängigkeit der Gewichtung der Ereignisse dargestellt. Der p-Value fällt wie erwartet mit steigender Luminosität, jedoch nähert er sich schon ab einer integrierten Luminosität von 25 fb^{-1} einem Grenzwert an. Um dieses Verhalten zu verstehen, ist in Abbildung 5.15 *rechts* der Verlauf des systematischen und des statistischen Fehlers in Abhängigkeit der Luminosität dargestellt. Bei niedrigen Luminositäten dominiert der statistische Fehler, wohingegen bei höheren Luminositäten der systematische Fehler den größeren Einfluss hat. Dies ist damit zu erklären, dass der systematische Fehler linear mit der Luminosität steigt, wogegen der statistische Fehler proportional zur Wurzel der be-

<i>tight*</i> Identifikation, 6/3 Isolation		
Parameter	bestimmter Wert	MC Unsicherheit
R^{bkg}	1.22	$\pm 29\%$
$R^{bkg'}$	1.03	$\pm 45\%$
α	0.44	$\pm 27\%$
c_1	0.025	$\pm 14\%$
c_2	0.106	$\pm 6\%$
c_3	0.003	$\pm 46\%$
c'_1	0.068	$\pm 9\%$
c'_2	0.155	$\pm 6\%$
c'_3	0.016	$\pm 20\%$

Tabelle 5.6: Die zugehörigen Monte Carlo Input Parameter für ID und 6/3 Iso Signalregiondefinition. Zur Signal- und Untergrundberechnung wird $R^{bkg} = R^{bkg'} = 1$ gesetzt und mögliche Abweichungen werden im systematischen Fehler mit der in Tabelle 5.1 angegebenen Unsicherheit berücksichtigt. Wie oben erwähnt, werden die c_i bei der Bestimmung des systematischen Fehlers um 50% variiert.

rechneten Ereigniszahl und somit zur Wurzel der Luminosität steigt¹⁴. Die limitierte Monte Carlo Statistik schränkt eine Entdeckungswahrscheinlichkeit des $W + \gamma + \gamma$ Prozesses ein.

5.8 Anwendung der 2×2D Sideband Methode auf Daten

In diesem Abschnitt wird die 2×2D Sideband Methode auf Daten angewandt. Als Signalregionendefinitionen werden die *tight** Identifikationsbedingung und die 6/3 GeV Isolationsbedingungen für das leading beziehungsweise subleading Photon gewählt, da diese Variante nach der durchgeführten Monte Carlo Studie sich bei der niedrigen erwarteten Statistik am besten eignet. Wie schon im vorherigen Abschnitt werden im Anhang die Ergebnisse für die *tight* 4/3 GeV Isolation Variante der Signalregiondefinition aufgeführt.

5.8.1 Reduzierung des elektroschwachen Untergrunds

Da die 2×2D Sideband Methode wie oben erwähnt nicht für verschiedene Untergründe gleichzeitig angewandt werden kann, wird zunächst eine Korrektur der selektier-

¹⁴Werden die Monte Carlo Ereignisse auf eine höhere Luminosität gewichtet, steigen alle Ereigniszahlen linear mit der veränderten gewählten Luminosität (siehe Abschnitt 5.3.1) Somit steigt auch die berechnete Ereigniszahl $W_{\gamma\gamma}^{TITI}$ linear mit der Luminosität

<i>tight*</i> Identifikation, 6/3 Isolation			
Ereignisart	Wert(Toy MC) -in Monte Carlo-	Toy MC Unsicherheit statistisch	Toy MC Unsicherheit systematisch
$W_{\gamma\gamma}^{TITI}$	9.7	$(+5.6)_{stat}$ (-7.1)	$(+1.8)_{sys}$ (-4.5)
$W_{\gamma j}^{TITI}$	2.0	$(+3.9)_{stat}$ (-0.7)	$(+2.1)_{sys}$ (-0.8)
W_{jj}^{TITI}	1.6	$(+3.1)_{stat}$ (-0.5)	$(+3.3)_{sys}$ (-0.5)
W_{bkg}^{TITI}	0.9	$(+1.5)_{stat}$ (-0.5)	$(+1.3)_{sys}$ (-0.3)
	5.0	$(+7.1)_{stat}$ (-1.6)	$(+4.5)_{sys}$ (-1.8)
p-Value	0.112		

Tabelle 5.7: Die Ergebnisse der Signal- und Untergrundberechnung mit der 2×2D Sideband Methode, wobei hier wieder der wahrscheinlichste Wert und der 68.28% Vertrauensbereich mit Toy Monte Carlo bestimmt wurde.

ten Ereignisse in den Signal und Kontrollfeldern durchgeführt, indem die gemessene Ereigniszahl um die Anzahl der erwarteten elektroschwachen Untergrundereignisse reduziert wird. Da die Bestimmung der elektroschwachen Untergrundereignisse auch nur innerhalb der Monte Carlo Unsicherheit erfolgen kann, wird dies bei der Bestimmung des systematischen Fehlers berücksichtigt. Analog zu den Monte Carlo Input Parametern wird dabei die Anzahl elektroschwacher Untergrundereignisse in der Toy Monte Carlo Studie innerhalb der Monte Carlo Unsicherheit variiert. In Tabelle 5.8 sind die zur Ereignisreduktion benötigten Ereigniszahlen aufgeführt.

5.8.2 Datenergebnisse

In Tabelle 5.9 sind die Ereigniszahlen aufgelistet, welche in den Feldern der 2×2D Sideband Methode gemessen wurden. Außerdem wurden in der rechten Spalte die Ereigniszahlen um die Anzahl an elektroschwachen Untergrundprozessen reduziert, welche durch Anwendung der Selektion auf das elektroschwache Monte Carlo Sample erhalten wurde. Tabelle 5.10 zeigt die mit Hilfe der 2×2D Sideband Methode berechnete Anzahl an Signal- und Untergrundereignissen, wobei die angegebenen Werte und Unsicherheiten auf Toy Monte Carlo Studien basieren. Der Maximalwert der mit Daten bestimmten PDF liegt etwas höher als in der Monte Carlo Studie vorhergesagt. Innerhalb des Vertrauensbereich sind die Werte konsistent.

<i>tight</i> * Identifikation, 6/3 Isolation		
Ereignisfeld	selektierter Wert	MC Unsicherheit
-in Monte Carlo-		
N_A	9.1	± 1.4
N_B	1.8	± 0.6
N_C	2.1	± 0.7
N_D	3.6	± 0.8
N'_A	4.4	± 1.0
N'_B	0.9	± 0.4
N'_C	1.3	± 0.5
N'_D	2.5	± 0.7

Tabelle 5.8: Erwartete Anzahl an Untergrundereignissen aus den elektroschwachen Untergrund Monte Carlo Samples.

Feld	selektierte Ereignisanzahl	reduzierte Ereignisanzahl
-in Daten-		
N_A	88 ± 9.4	$78.9 \pm 9.4^{stat} \pm 1.4^{sys}$
N_B	21 ± 4.7	$19.2 \pm 4.7^{stat} \pm 0.6^{sys}$
N_C	58 ± 7.6	$55.9 \pm 7.6^{stat} \pm 0.7^{sys}$
N_D	43 ± 6.6	$39.4 \pm 6.6^{stat} \pm 0.8^{sys}$
N'_A	28 ± 5.3	$23.6 \pm 5.3^{stat} \pm 1.0^{sys}$
N'_B	15 ± 3.9	$14.1 \pm 3.9^{stat} \pm 0.4^{sys}$
N'_C	17 ± 4.1	$15.7 \pm 4.1^{stat} \pm 0.5^{sys}$
N'_D	28 ± 4.3	$25.5 \pm 4.3^{stat} \pm 0.7^{sys}$

Tabelle 5.9: Die auf Daten selektierten Ereigniszahlen in den Feldern der 2×2D Sideband Methode. Die Monte Carlo Unsicherheit der elektroschwachen Untergrundereignisse wird hier als systematischer Fehler angegeben.

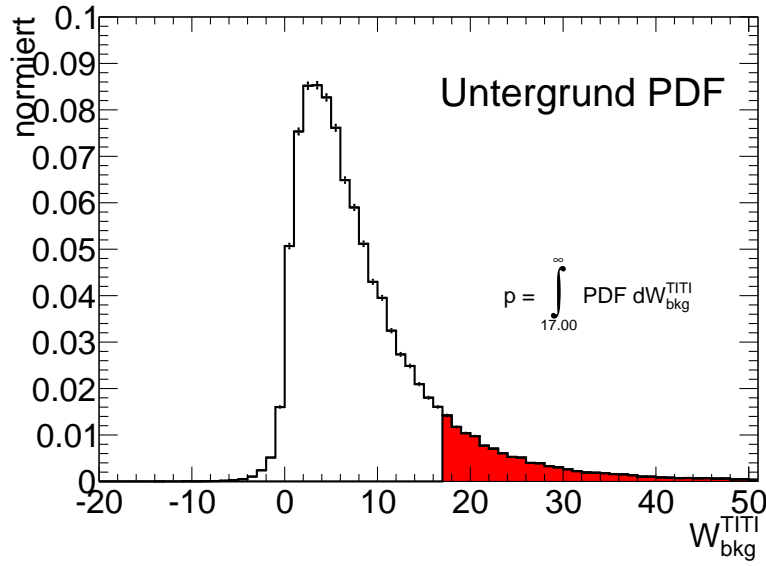


Abbildung 5.14: Die Verteilung stellt die Untergrund PDF bei einer integrierten Luminosität von 4.64 fb^{-1} dar, welche sich durch gleichzeitige Variation von erwarteten Feldeinträgen und Monte Carlo Input Parameter ergibt. Integration von $W_{bkg}^{TITI} = N'_A$ bis unendlich ergibt den p-Value.

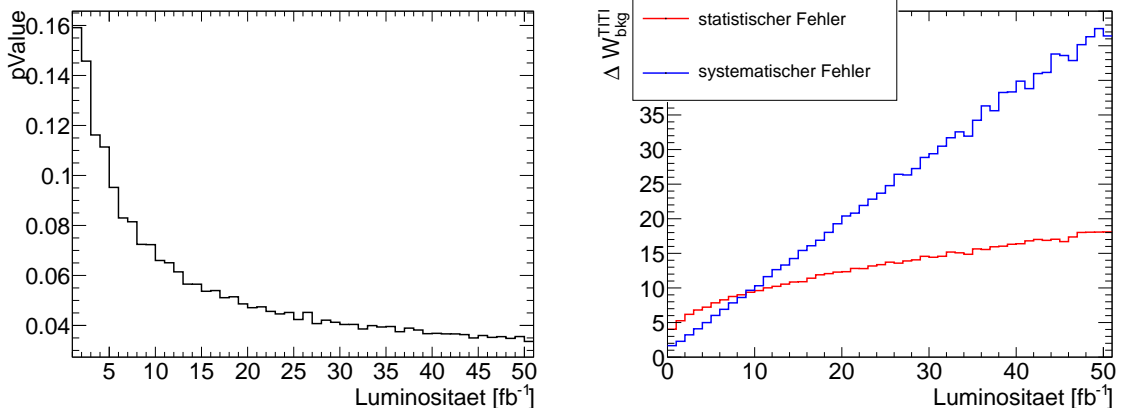


Abbildung 5.15: *Links*: Berechneter Verlauf des p-Values in Abhängigkeit der integrierten Luminosität bei einer *tight** Identifikation und 6/3 GeV Isolation Signalregiondefinition. *Rechts*: Für die gleiche Signalregiondefinition der Verlauf von systematischer und statistischer Unsicherheit der Anzahl an Untergrundereignissen. Es wird konservativ die größere Unsicherheit verwendet.

<i>tight</i> * Identifikation, 6/3 Isolation			
	Wert (Toy MC)	Toy MC Unsicherheit	Toy MC Unsicherheit
	-in Daten-	statistisch	systematisch
W^{TITI}	15.6	$(+5.4)_{stat}$ (-11.5)	$(+2.8)_{sys}$ (-7.0)
$W^{\gamma\gamma}_{TITI}$	2.6	$(+5.8)_{stat}$ (-0.9)	$(+3.4)_{sys}$ (-1.2)
$W^{\gamma j}_{TITI}$	2.1	$(+4.6)_{stat}$ (-0.8)	$(+5.0)_{sys}$ (-0.6)
W^{jj}_{TITI}	1.2	$(+2.1)_{stat}$ (-0.6)	$(+1.6)_{sys}$ (-0.5)
W^{TITI}_{bkg}	6.9	$(+10.4)_{stat}$ (-2.0)	$(+7.1)_{sys}$ (-2.4)
p-Value	0.070		

Tabelle 5.10: Anzahl an Signal -und Untergrundereignisse in Daten, berechnet mit der 2×2 D Sideband Methode. Angegebene Werte und Unsicherheit basieren auf Toy Monte Carlo Studien.

6 Zusammenfassung

In dieser Bachelorarbeit wurde die Ereignisselektion der $W + \gamma + \gamma$ Analyse vorgestellt. Außerdem wurde eine auf Daten basierende Untergrundbestimmung für den dominanten $W + jets$ Untergrund, sowie eine auf Monte Carlo Studien basierende Abschätzung für weitere elektroschwache Untergrundprozesse präsentiert.

Zur Bestimmung der Anzahl an $W + jets$ Untergrundereignissen in der Signalregion der $W + \gamma + \gamma$ Analyse wurde die $2 \times 2D$ Sideband Methode implementiert und die korrekte Funktionsweise in einem Monte Carlo Closure Test gezeigt. Zur Bestimmung der statistischen Unsicherheit wurden eine analytische Fehlerrechnung sowie alternativ Toy Monte Carlo Studien durchgeführt. Mit Einführung der Toy Monte Carlo Studien konnten Abweichungen der Signal- und Untergrundwahrscheinlichkeitsdichtefunktionen von einer symmetrischen Verteilung bei niedrigen Ereigniszahlen berücksichtigt werden. Vergleiche ergaben, dass bei hoher Statistik gaußverteilte Wahrscheinlichkeitsdichtefunktionen vorliegen und somit analytische Fehlerrechnung und Toy Monte Carlo Methode zum gleichen Ergebnis führen.

Da in dem verwendeten Datensatz aus dem Jahr 2011 insgesamt nur wenige Ereignisse die $W + \gamma + \gamma$ Selektion überstehen, wurden verschiedene Singalregiondefinitionen eingeführt. In Monte Carlo Studien wurde der Einfluss auf die statistische Unsicherheit der Ergebnisse der $2 \times 2D$ Sideband Methode getestet.

In einer Optimierungsstudie wurde zur Definition verschiedener Signalregionen eine neue Photonidentifikationsvariable, $tight^*$, eingeführt und unterschiedliche Kalorimeter Isolationsbedingungen an die beiden Photonkandidaten gestellt. Die Einführung der neuen Photonidentifikation $tight^*$ führt dazu, dass insgesamt mehr Ereignisse in den Feldern der $2 \times 2D$ Sideband Methode selektiert werden und somit insbesondere bei der derzeitigen integrierten Luminosität die statistische Unsicherheit von bestimmter Signal- und Untergrundereignisszahl reduziert wird. Für zukünftige Datensätze, die einer höheren integrierten Luminosität entsprechen, wird die Verwendung der herkömmlichen $tight$ Identifikation empfohlen, da in Monte Carlo Studien die Einführung der $tight^*$ Identifikation nur eine geringe Reduzierung der statistischen Unsicherheit erbrachte. Im Vergleich zur $tight$ Identifikation liegt jedoch eine deutlich geringere Unterdrückungsfähigkeit gegen sowohl $W + jets$ als auch elektroschwache Untergrundereignisse vor. Außerdem wird das Problem niedrig besetzter Felder in der $2 \times 2D$ Sideband Methode für zukünftige Datensätze nicht mehr erwartet. Anwendung der $2 \times 2D$ Sideband Methode in Daten zeigt im Vergleich keine Verbesserung der statistischen Unsicherheit durch die Einführung der $tight^*$ Identifikation.

Zur Bestimmung des systematischen Fehlers wurden ebenfalls sowohl eine analytische Fehlerrechnung als auch eine alternative Abschätzung mit Toy Monte Carlo

Studien eingeführt. Der Einfluss der Unsicherheit der Monte Carlo Input Parameter, bedingt durch limitierte Monte Carlo Statistik, wurde studiert. Es wurde gezeigt, dass der systematische Gesamtfehler von der hohen Unsicherheit auf die Korrelation $R^{bkg'}$ dominiert wird. Monte Carlo Prognosen zeigen, dass schon ab einer integrierten Luminosität von 10 fb^{-1} systematischer Fehler und statistischer Unsicherheit ungefähr gleich groß sind, wobei bei höheren integrierten Luminositäten der systematische Fehler dominiert.

Die durchgeführten Monte Carlo Studien zeigen, dass sich die $2 \times 2\text{D}$ Sideband Methode zur Untergrundbestimmung des dominanten $W + jets$ Untergrunds für zukünftige Datensätze gut eignet. Um eine präzise Untergrundbestimmung durchführen zu können, muss es jedoch gelingen, die Unsicherheiten der Monte Carlo Input Parameter und somit den systematischen Gesamtfehler zu senken. Bei den momentan vorliegenden Ereigniszahlen aus dem Jahr 2011 sind Berechnungen durch statistische Unsicherheiten limitiert.

Anwendung der Untergrundabschätzung auf die Daten aus dem Jahr 2011 ergibt eine Anzahl von $15.6^{(+5.4)}_{(-11.5)}{}^{stat}({}^{+2.8}_{(-7.0)}{}^{sys})$ selektierter Signalereignisse in der optimierten Signalregion, sowie ein Ausschluss der Hypothese, dass die Daten nur aus Untergrundereignissen bestehen, von 93.0%.

A Monte Carlo Input Parameter

In diesem Kapitel sind alle nicht im Haupttext aufgeführten Monte Carlo Input Parameter in Abhängigkeit der Isolationsbedingungen für das leading und subleading Photon sowohl für *tight* als auch *tight** Identifikationsbedingung in der Signalregion dargestellt.

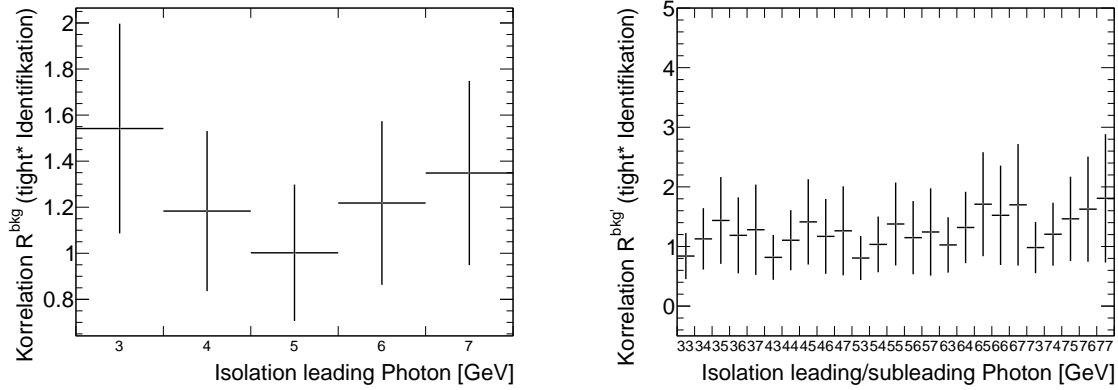


Abbildung A.1: Korrelation R^{bkg} und $R^{bkg'}$ in Abhängigkeit der Isolationsbedingung (*tight** Identifikationsbedingung).

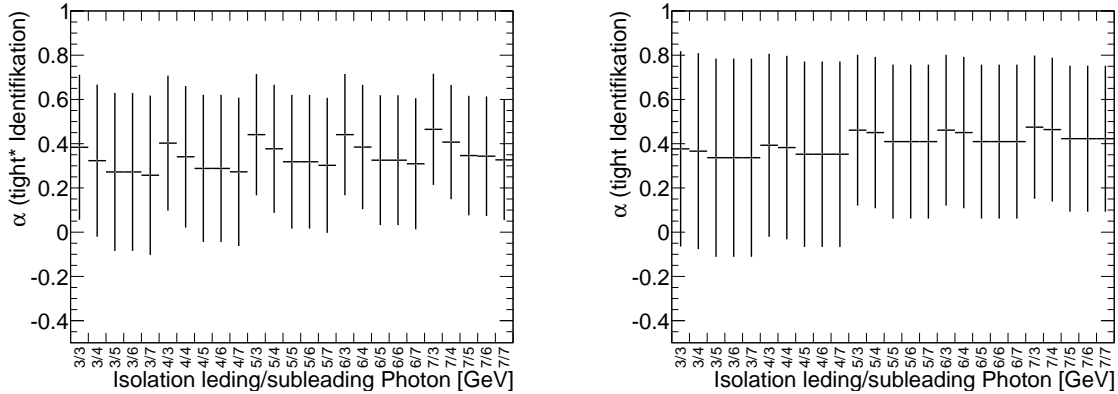


Abbildung A.2: Verlauf des Parameter α in Abhängigkeit der Isolationsbedingungen. *Links:* *tight** Identifikationsbedingung in der Signalregion. *Rechts:* *tight* Identifikationsbedingung.

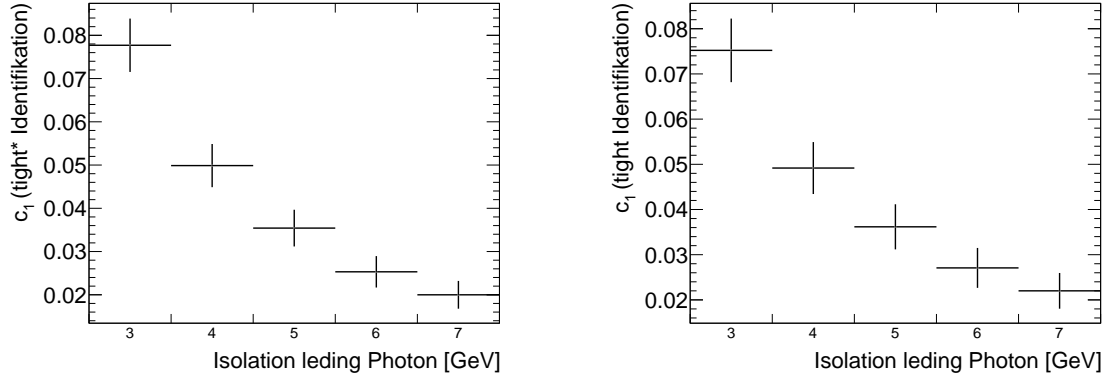


Abbildung A.3: Verlauf des Parameter c_1 in Abhängigkeit der Isolationsbedingungen. *Links*: *tight** Identifikationsbedingung in der Signalregion. *Rechts*: *tight* Identifikationsbedingung.

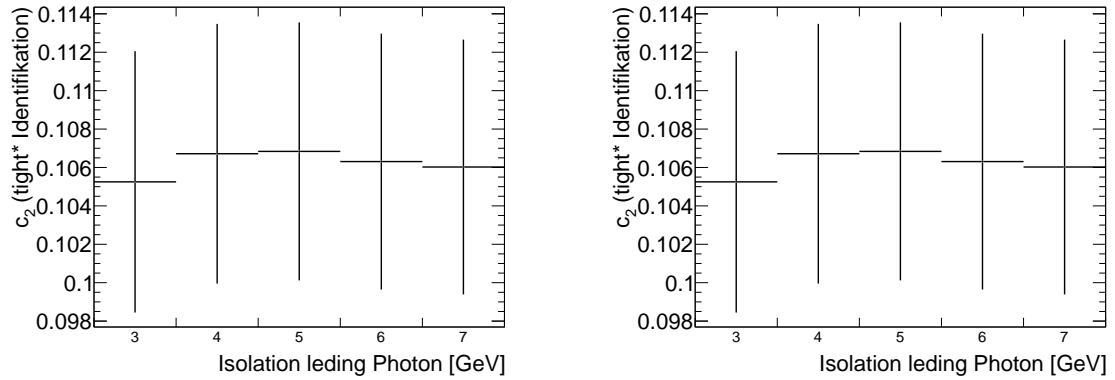


Abbildung A.4: Verlauf des Parameter c_2 in Abhängigkeit der Isolationsbedingungen. *Links*: *tight** Identifikationsbedingung in der Signalregion. *Rechts*: *tight* Identifikationsbedingung.

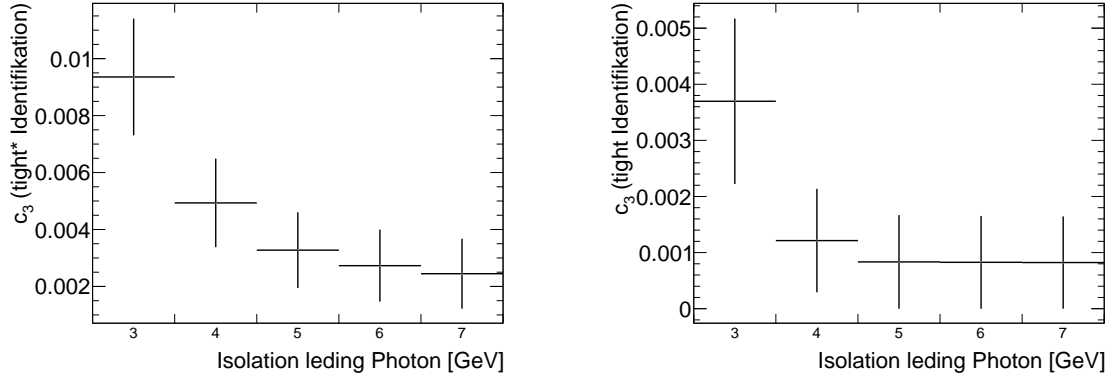


Abbildung A.5: Verlauf des Parameter c_3 in Abhängigkeit der Isolationsbedingungen. *Links: tight** Identifikationsbedingung in der Signalregion. *Rechts: tight* Identifikationsbedingung.

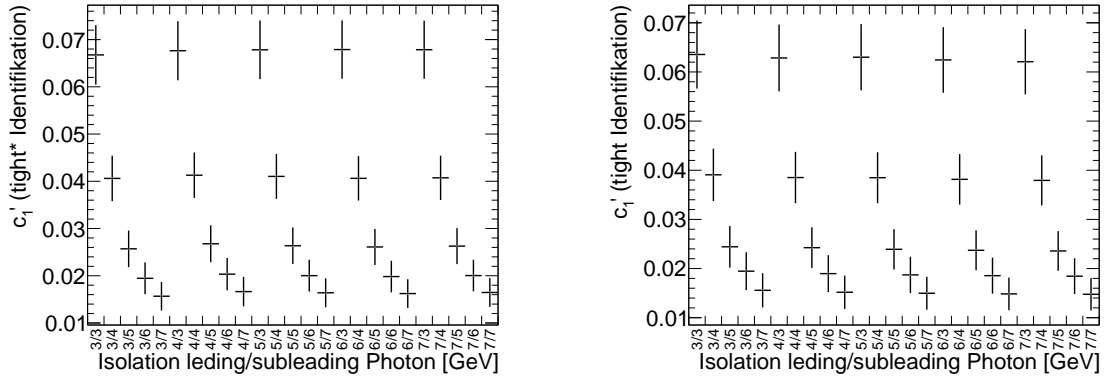


Abbildung A.6: Verlauf des Parameter c_1' in Abhängigkeit der Isolationsbedingungen. *Links: tight** Identifikationsbedingung in der Signalregion. *Rechts: tight* Identifikationsbedingung.

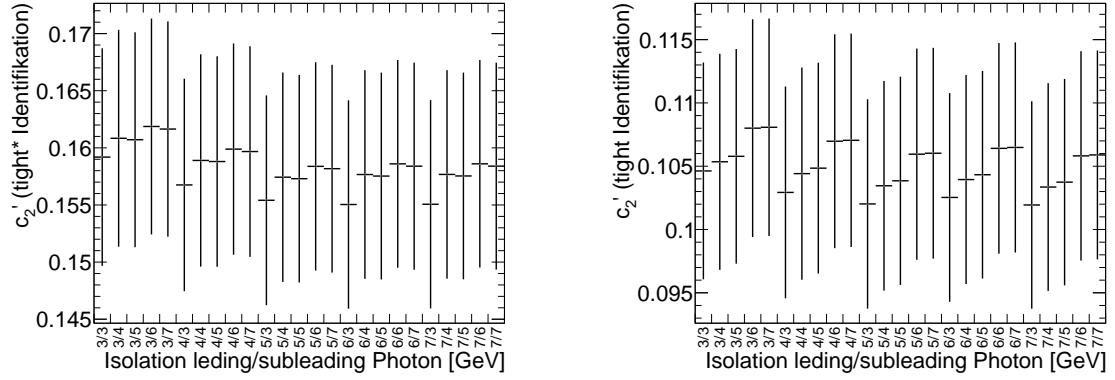


Abbildung A.7: Verlauf des Parameter c'_2 in Abhängigkeit der Isolationsbedingungen. *Links*: *tight** Identifikationsbedingung in der Signalregion. *Rechts*: *tight* Identifikationsbedingung.

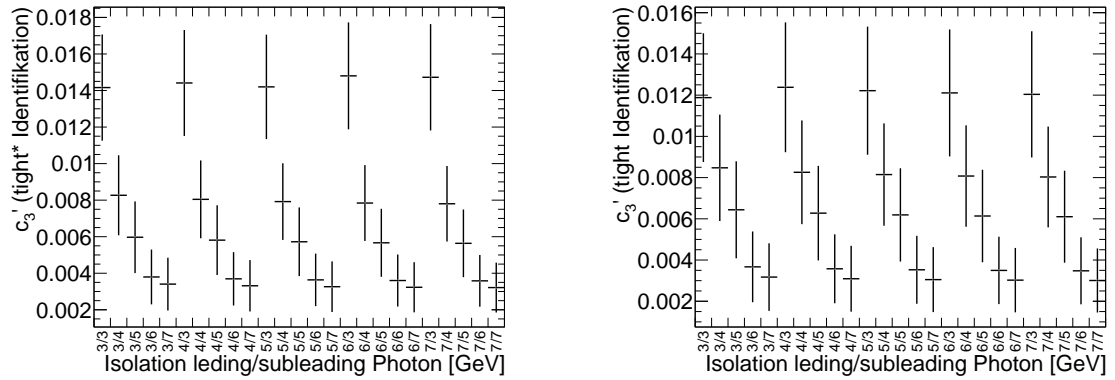


Abbildung A.8: Verlauf des Parameter c'_3 in Abhängigkeit der Isolationsbedingungen. *Links*: *tight** Identifikationsbedingung in der Signalregion. *Rechts*: *tight* Identifikationsbedingung.

B Monte Carlo Studie für alternative Signalregiondefinition

In diesem Kapitel werden die Ergebnisse der 2×2 D Sideband Methode analog zu 5.7.1 mit einer alternativ gewählten Signalregiondefinition aufgeführt. Die Bedingung für die Signalregion setzt sich aus einer *tight* Identifikationsbedingung bei einer *loose* Vorselektion und einer Isolationsbedingung von 4/3 GeV für das leading/subleading Photon zusammen. In Tabelle B.1 sind die selektierten Ereigniszahlen in den Feldern der 2×2 D Sideband Methode aufgelistet, Tabelle B.2 zeigt die benötigten Monte Carlo Input Parameter und in Tabelle B.3 sind die mit Toy Monte Carlo Studien bestimmten Ergebnisse der Ereignisanzahlen von Signal- und Untergrund. Zwar können die Monte Carlo Input Parameter bei einer *tight* Bedingung in der Signalregion nur innerhalb einer größeren Unsicherheit bestimmt werden, jedoch ist der systematische Fehler auch von der Gesamtzahl der selektierten Ereignissen abhängig, so dass sich trotzdem ein kleiner systematischer Fehler im Vergleich zur *tight** Identifikationsbedingung in der Signalregion ergibt.

<i>tight</i> Identifikation, 4/3 Isolation		
Ereignisfeld	selektierter Wert	MC Unsicherheit
-in Monte Carlo-		
N_A	28.5	± 3.0
N_B	10.3	± 3.2
N_C	13.1	± 2.5
N_D	7.9	± 1.9
N'_A	11.1	± 1.3
N'_B	2.6	± 1.0
N'_C	9.0	± 1.9
N'_D	5.8	± 1.5

Tabelle B.1: Für die alternative *tight* ID 4/3 GeV Iso Signalregiondefinitionen sind hier die selektierten Ereigniszahlen in den Signal und Kontrollfeldern aufgelistet.

<i>tight</i> Identifikation, 4/3 Isolation		
Parameter	bestimmter Wert	MC Unsicherheit
R^{bkg}	0.47	$\pm 54\%$
$R^{bkg'}$	0.89	$\pm 71\%$
α	0.39	$\pm 41\%$
c_1	0.049	$\pm 12\%$
c_2	0.066	$\pm 9\%$
c_3	0.001	$\pm 76\%$
c'_1	0.063	$\pm 11\%$
c'_2	0.103	$\pm 8\%$
c'_3	0.124	$\pm 25\%$

Tabelle B.2: Die zugehörigen Monte Carlo Input Parameter. Zur Signal- und Untergrundberechnung wird wie oben $R^{bkg} = R^{bkg'} = 1$ gesetzt und mögliche Abweichungen werden im systematischen Fehler mit der in Tabelle 5.1 angegebenen Unsicherheit berücksichtigt. Auch hier werden die c_i bei der Bestimmung des systematischen Fehlers großzügig variiert.

<i>tight</i> Identifikation, 4/3 Isolation			
Ereignisart	Wert (Toy MC) -in Monte Carlo-	Toy MC Unsicherheit statistisch	Toy MC Unsicherheit systematisch
W^{TITI}	9.3	$(+3.7)_{stat}$ (-9.2)	$(+1.3)_{sys}$ (-3.1)
$W^{\gamma\gamma}_{TITI}$	0.3	$(+4.4)_{stat}$ (-1.3)	$(+1.9)_{sys}$ (-0.9)
$W^{j\gamma}_{TITI}$	0.2	$(+2.9)_{stat}$ (-1.0)	$(+1.7)_{sys}$ (-0.5)
W^{jj}_{TITI}	0.3	$(+4.5)_{stat}$ (-0.4)	$(+2.2)_{sys}$ (-0.9)
W^{jj}_{bkg}	0.3	$(+9.6)_{stat}$ (-0.3)	$(+3.2)_{sys}$ (-1.2)
p-Value	0.146		

Tabelle B.3: Die Ergebnisse der Signal und Untergrundberechnung mit der $2 \times 2D$ Sideband Methode, wobei hier wieder der wahrscheinlichste Wert und der 68.28 Vertrauensbereich mit Toy Monte Carlo bestimmt wurde.

C Alternative Studie auf Daten

In den folgenden Tabellen C.2 und C.3 befinden sich die Ergebnisse der 2×2D Sideband Studie mit alternativer Signalregiondefinition angewandt auf Daten: *tight* Identifikation und Isolation leading/subleading Photon $< 4/3$ GeV. Tabelle C.2 zeigt die Ereigniszahlen in den Feldern der 2×2D Sideband Methode, Tabelle C.3 die Ergebnisse der Signal- und Untergrundbestimmung. Die benötigte Anzahl an elektroschwachen Untergrundereignissen stehen in Tabelle C.1.

<i>tight</i> Identifikation, 4/3 Isolation		
Ereignisfeld	selektierter Wert	MC Unsicherheit
-in Monte Carlo-		
N_A	3.3	± 0.9
N_B	1.1	± 0.5
N_C	0.1	± 0.1
N_D	0.1	± 0.2
N'_A	2.5	± 0.7
N'_B	0.4	± 0.3
N'_C	0.4	± 0.3
N'_D	0.1	± 0.1

Tabelle C.1: Erwartete Anzahl an Untergrundereignissen aus den elektroschwachen Untergrund Monte Carlo Samples.

Feld	selektierte Ereignisanzahl -in Daten-	reduzierte Ereignisanzahl
N_A	40 ± 6.3	$36.7 \pm 6.3^{stat} \pm 0.9^{sys}$
N_B	8 ± 2.8	$6.9 \pm 2.8^{stat} \pm 0.5^{sys}$
N_C	8 ± 2.8	$8.0 \pm 2.8^{stat} \pm 0.1^{sys}$
N_D	7 ± 2.6	$6.8 \pm 2.6^{stat} \pm 0.2^{sys}$
N'_A	21 ± 4.6	$18.5 \pm 4.6^{stat} \pm 0.7^{sys}$
N'_B	10 ± 3.2	$9.6 \pm 3.2^{stat} \pm 0.3^{sys}$
N'_C	3 ± 1.7	$2.6 \pm 1.7^{stat} \pm 0.3^{sys}$
N'_D	6 ± 2.4	$6.0 \pm 2.4^{stat} \pm 0.1^{sys}$

Tabelle C.2: In Daten selektierten Ereigniszahlen in den Signal und Kontrollfeldern für die alternative *tight* ID und 4/3 Iso Signalregiondefinition. Die Monte Carlo Unsicherheit der elektroschwachen Untergründereignisse wird als systematischer Fehler angegeben.

<i>tight</i> Identifikation, 4/3 Isolation			
Ereignisart	Wert (Toy MC) - in Daten-	Toy MC Unsicherheit statistisch	Toy MC Unsicherheit systematisch
W^{TITI}	17.8	$(+5.3)^{stat}$ (-9.9)	$(+2.4)^{sys}$ (-5.2)
$W^{\gamma TITI}$	0.3	$(+4.9)^{stat}$ (-1.0)	$(+3.3)^{sys}$ (-0.6)
$W^{j TITI}$	0.1	$(+3.4)^{stat}$ (-0.7)	$(+2.5)^{sys}$ (-0.4)
$W^{jj TITI}$	0.4	$(+0.4)^{stat}$ (-2.3)	$(+0.2)^{sys}$ (-1.4)
$W^{jj TITI}_{bkg}$	0.5	$(+7.6)^{stat}$ (-2.3)	$(+4.9)^{sys}$ (-2.3)
p-Value	0.106		

Tabelle C.3: Die Ergebnisse der Signal- und Untergrundberechnung mit der 2×2D Sideband Methode angewandt auf Daten für die alternative *tight* ID und 4/3 Iso Signalregiondefinition.

D Monte Carlo Vorhersage für 2012

Als Vorhersage für das Jahr 2012 wird das verwendete Monte Carlo Sample auf eine integrierte Luminosität von 20 fb^{-1} gewichtet. Somit wird die Studie aus 5.7.1 mit der erwarteten Statistik der Daten im Jahr 2012 wiederholt. In Tabelle D.1 sind die selektierten Ereigniszahlen und ihre Monte Carlo Unsicherheit zu finden, die Ergebnisse der Signal- und Untergrundbestimmung befinden sich in Tabelle D.2. Ein Vergleich mit Abschnitt 5.7.1 und Abschnitt B zeigt, wie die Gesamtbreite des 68.28% Vertrauensbereichs bei der Bestimmung des statistischen Fehlers proportional zur Wurzel der Luminosität steigt. Die Bestimmung des systematischen Fehlers zeigt eine direkte Proportionalität von Vertrauensbereichs und Luminosität.

Ereignisse	<i>tight*</i> Identifikation, 6/3 Isolation		<i>tight</i> Identifikation, 4/3 Isolation	
	selektierter Wert -in Monte Carlo-	Unsicherheit	selektierter Wert -in Monte Carlo-	Unsicherheit
N_A	309.6	± 24.7	121.1	± 12.5
N_B	72.4	± 11.9	43.9	± 13.8
N_C	249.4	± 25.4	55.8	± 10.4
N_D	171.3	± 18.9	33.7	± 8.2
N'_A	70.9	± 7.7	47.4	± 5.5
N'_B	34.4	± 7.8	11.1	± 4.3
N'_C	84.5	± 12.4	38.0	± 8.1
N'_D	119.8	± 18.4	24.6	± 6.6

Tabelle D.1: Selektierten Ereigniszahlen in den Signal- und Kontrollfeldern bei einer Gewichtung der Ereignisse auf eine integrierte Luminosität von 20 fb^{-1} .

Ereignisart	<i>tight</i> * Identifikation, 6/3 Isolation		<i>tight</i> Identifikation, 4/3 Isolation	
	Wert (Toy MC) -in Monte Carlo-	Unsicherheit	Wert(Toy MC) -in Monte Carlo-	Unsicherheit
W^{TITI}	39.4	$(+11.8)_{stat} (+6.9)_{sys}$ $(-14.4)_{stat} (-19.9)_{sys}$	34.9	$(+9.3)_{stat} (+4.8)_{sys}$ $(-15.8)_{stat} (-13.2)_{sys}$
$W^{\gamma\gamma}_{TITI}$	12.3	$(+6.6)_{stat} (+8.3)_{sys}$ $(-2.7)_{stat} (-4.0)_{sys}$	0.9	$(+8.1)_{stat} (+7.8)_{sys}$ $(-2.4)_{stat} (-3.5)_{sys}$
$W^{\gamma j}_{TITI}$	9.7	$(+5.2)_{stat} (+14.1)_{sys}$ $(-2.2)_{stat} (-1.7)_{sys}$	0.6	$(+5.2)_{stat} (+6.3)_{sys}$ $(-1.6)_{stat} (-1.4)_{sys}$
W^{jj}_{TITI}	5.2	$(+2.9)_{stat} (+3.7)_{sys}$ $(-0.7)_{stat} (-2.0)_{sys}$	6.7	$(+8.6)_{stat} (+9.2)_{sys}$ $(-1.9)_{stat} (-2.9)_{sys}$
W^{TITI}_{bkg}	28.4	$(+11.7)_{stat} (+19.8)_{sys}$ $(-6.1)_{stat} (-7.0)_{sys}$	9.7	$(+15.7)_{stat} (+13.3)_{sys}$ $(-3.5)_{stat} (-4.7)_{sys}$
p-Value	0.047		0.066	

Tabelle D.2: Die Ergebnisse der Signal und Untergrundberechnung mit der 2×2D Sideband Methode bei einer Gewichtung der Ereignisse auf eine integrierte Luminosität von 20 fb⁻¹.

E Zusätzliche Rechnung

Es gilt zu zeigen, dass die Summe der Erwartungswerte von $W_{\gamma\gamma}^{TITI}$ und W_{bkg}^{TITI} dem Erwartungswert der Ereigniszahl N'_A entspricht.

In der Toy Plot Studie gilt bei jeder Durchführung der Berechnung die Beziehung:

$$W_{\gamma\gamma,i}^{TITI} + W_{bkg,i}^{TITI} = N'_{A,i} \quad (E.1)$$

denn die Summe von Signalereignissen und Untergrundereignissen ergibt die Anzahl Ereignisse in A', die sich jedoch auch in jedem Durchgang im Rahmen der Poissonstatistik ändert.

Die Erwartungswerte von Signal und Untergrund werden definiert als:

$$\langle W_{\gamma\gamma}^{TITI} \rangle = \sum_i P(W_{\gamma\gamma,i}^{TITI}) \cdot W_{\gamma\gamma,i}^{TITI} \quad (E.2)$$

$$\langle W_{bkg}^{TITI} \rangle = \sum_i P(W_{bkg,i}^{TITI}) \cdot W_{bkg,i}^{TITI} \quad (E.3)$$

wobei i über alle berechneten Ergebnisse der Signal- und Untergrundberechnung summiert. $P(W_{\gamma\gamma,i}^{TITI})$ und $P(W_{bkg,i}^{TITI})$ geben die Wahrscheinlichkeiten an, mit der die Ergebnisse $W_{\gamma\gamma,i}^{TITI}$ bzw. $W_{bkg,i}^{TITI}$ erhalten werden.

Die Wahrscheinlichkeiten berechnen sich nach:

$$P(W_{\gamma\gamma,i}^{TITI}) = \frac{N(W_{\gamma\gamma,i}^{TITI})}{N_{ges}} \quad (E.4)$$

$$P(W_{bkg,i}^{TITI}) = \frac{N(W_{bkg,i}^{TITI})}{N_{ges}} \quad (E.5)$$

Wobei $N(W_{\gamma\gamma,i}^{TITI}) \in \mathbb{N}$ die Anzahl angibt, mit der das Ergebnis $W_{\gamma\gamma,i}^{TITI}$ berechnet wurde. Analog für den Untergrund. N_{ges} ist die Anzahl der durchgeführten Berechnungen. Für die Summe der Erwartungswerte gilt folglich:

$$\langle W_{\gamma\gamma}^{TITI} \rangle + \langle W_{bkg}^{TITI} \rangle = \sum_i P(W_{\gamma\gamma,i}^{TITI}) \cdot W_{\gamma\gamma,i}^{TITI} + P(W_{bkg,i}^{TITI}) \cdot W_{bkg,i}^{TITI} \quad (E.6)$$

$$= \frac{1}{N_{ges}} \sum_i N(W_{\gamma\gamma,i}^{TITI}) \cdot W_{\gamma\gamma,i}^{TITI} + N(W_{bkg,i}^{TITI}) \cdot W_{bkg,i}^{TITI} \quad (E.7)$$

$$= \frac{1}{N_{ges}} \sum_i N'_{A,i} = \langle N'_A \rangle \quad (E.8)$$

Die letzte Zeile folgt, da jeder Summand $W_{\gamma\gamma,i}^{TITI} + W_{bkg,i}^{TITI}$ nach E.1 $N'_{A,i}$ ergibt.

F Verzeichnisse

F.1 Abbildungsverzeichnis

2.1	Typische $W + \gamma + \gamma$ Feynmangraphen [4]	10
3.1	Der ATLAS Detektor [9]	14
3.2	Das elektromagnetische Kalorimeter des ATLAS Detektors [3]	15
3.3	Isolationsverteilung Photonen	21
4.1	Anzahl Ereignisse nach den Selektionsschnitten der $W + \gamma + \gamma$ Selektion	25
5.1	Signal- und Kontrollregionen der 2×2D Sideband Methode (Nach Vorlage aus [12])	28
5.2	Korrelation R^{bkg} und $R^{bkg'}$ in Abhängigkeit der Isolationsbedingung (tight Identifikationsbedingung)	38
5.3	Monte Carlo Closure Test	39
5.4	Monte Carlo Closure Test mit $R^{bkg} = R^{bkg'} = 1$	40
5.5	Abhängigkeit der berechneten Anzahl Signalereignisse $W_{\gamma\gamma}^{TITI}$ von R^{bkg} und $R^{bkg'}$ am Beispiel des Isolationsschnitt 5/5 GeV.	41
5.6	Analytisch bestimmte PDF für Signal- und Untergrundereignisse am Beispiel der tight* ID und 6/3 GeV Iso Signalregiondefinition	43
5.7	Mit Toy Monte Carlo bestimmte PDF für Signal- und Untergrundergebnisse am Beispiel der tight* ID und 6/3 GeV Iso Signalregiondefinition	44
5.8	Zweidimensionale Darstellung von Signal- und Untergrundereignissen in der Toy Monte Carlo Studie am Beispiel der tight* ID und 3/7 GeV Iso Signalregiondefinition	45
5.9	Vergleich von analytisch und mit Toy Monte Carlo Methode bestimmte Maximalwerte der Signal- (<i>links</i>) und Untergrund PDF (<i>rechts</i>) sowie der wahren Anzahl selektierter Ereignisse	46
5.10	Vergleich von der Summe der Erwartungswerte von Signal- und Untergrund PDF mit der Anzahl an Ereignissen N'_A	46
5.11	<i>Links</i> : Zweidimensionale Darstellung von Signal- und Untergrundergebnissen in der Toy Monte Carlo Studie am Beispiel der tight* ID und 3/7 GeV Iso Signalregiondefinition bei einer integrierten Luminosität von 464 fb^{-1} , <i>Rechts</i> : Untergrund PDF bei hoher Statistik (<i>tight*</i> Identifikation und und 6/3 GeV Isolationsbedingung)	47

5.12	Vergleich der Wahrscheinlichkeitsverteilung des relativen Fehlers der $W_{\gamma\gamma}^{TITI}$ Berechnung bei integrierter Luminosität von 4.64 fb^{-1} <i>links</i> und 464 fb^{-1} <i>rechts</i>	48
5.13	Signal- und Untergrund PDF, die sich in systematischer Toy Monte Carlo Studie ergeben am Beispiel der <i>tight*</i> ID und eine 6/3 GeV Iso Signalregiondefinition	52
5.14	P-Value Bestimmung am Beispiel der <i>tight*</i> ID und 6/3 GeV Iso Signalregiondefinition	57
5.15	Erwarteter Verlauf von p-Value, statistischem und systematischem Fehler in Abhängigkeit der integrierten Luminosität bei einer <i>tight*</i> ID und 6/3 GeV Iso Signalregiondefinition	57
A.1	Korrelation R^{bkg} und $R^{bkg'}$ in Abhängigkeit der Isolationsbedingungen (<i>tight*</i> Identifikationsbedingung)	61
A.2	Verlauf des Parameter α in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	61
A.3	Verlauf des Parameter c_1 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	62
A.4	Verlauf des Parameter c_2 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	62
A.5	Verlauf des Parameter c_3 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	63
A.6	Verlauf des Parameter c'_1 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	63
A.7	Verlauf des Parameter c'_2 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	64
A.8	Verlauf des Parameter c'_3 in Abhängigkeit der Isolationsbedingungen. <i>Links:tight*</i> Identifikationsbedingung in der Signalregion. <i>Rechts:tight*</i> Identifikationsbedingung	64

F.2 Tabellenverzeichnis

2.1	Die Signal Monte Carlo Samples für die $W + \gamma + \gamma$ Analyse	12
2.2	Monte Carlo Samples dervdominanten $W + jets$ Untergrund Prozesse	12
2.3	Monte Carlo Samples des elektroschwachen Untergrundes	13
5.1	Bestimmung des systematischen Fehlers der Korrelationen R^{bkg} und $R^{bkg'}$	37

5.2	Anzahl der bestimmten Signal- und Untergrundereignisse in Abhängigkeit der gewählten Luminosität.	47
5.3	Analytische Bestimmung des systematischen Fehlers	49
5.4	Bestimmung des systematischen Fehlers mit Toy Monte Carlo Studien	51
5.5	Selektierten Monte Carlo Ereigniszahlen in den Signal und Kontrollfeldern für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	53
5.6	Monte Carlo Input Parameter für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	54
5.7	Die Ergebnisse der Signal und Untergrundberechnung mit der 2×2D Sideband Methode angewandt auf Monte Carlo Samples für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	55
5.8	Erwartete Anzahl an elektroschwachen Untergrundereignissen für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	56
5.9	In Daten selektierten Ereigniszahlen in den Signal und Kontrollfeldern für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	56
5.10	Die Ergebnisse der Signal- und Untergrundberechnung mit der 2×2D Sideband Methode angewandt auf Daten für die <i>tight*</i> ID und 6/3 Iso Signalregiondefinition	58
B.1	Selektierten Monte Carlo Ereigniszahlen in den Signal- und Kontrollfeldern für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	65
B.2	Monte Carlo Input Parameter für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	66
B.3	Ergebnisse der Signal und Untergrundberechnung mit der 2×2D Sideband Methode angewandt auf Monte Carlo Samples für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	66
C.1	Erwartete Anzahl elektroschwacher Untergrundereignissen für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	67
C.2	In Daten selektierten Ereigniszahlen in den Signal und Kontrollfeldern für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	68
C.3	Die Ergebnisse der Signal- und Untergrundberechnung mit der 2×2D Sideband Methode angewandt auf Daten für die <i>tight</i> ID und 4/3 Iso Signalregiondefinition	68
D.1	Selektierten Monte Carlo Ereigniszahlen in den Signal- und Kontrollfeldern bei einer Gewichtung der Ereignisse auf eine integrierte Luminosität von 20 fb ⁻¹	69
D.2	Ergebnisse der Signal- und Untergrundberechnung mit der 2×2D Sideband Methode angewandt bei einer Gewichtung der Monte Carlo Ereignisse auf eine integrierte Luminosität von 20 fb ⁻¹	70

G Literaturverzeichnis

- [1] G. Aad et al. Measurement of $W\gamma$ and $Z\gamma$ production cross sections in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous triple gauge couplings with the ATLAS detector. oai:cds.cern.ch:1448436. Technical Report arXiv:1205.2531. CERN-PH-EP-2012-059, CERN, Geneva, May 2012.
- [2] Lyndon Evans and Philip Bryant (editors). LHC Machine. 2008. JINST 3 S08001.
- [3] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [4] V. Scharf. private Kommunikation.
- [5] J. Beringer and others (Particle Data Group). *Phys. Rev.*, D86, 2012.
- [6] Georges Aad et al. Measurement of the production cross section for W-bosons in association with jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys.Lett.*, B698:325–345, 2011.
- [7] Georges Aad et al. Luminosity Determination in pp Collisions at $\sqrt{s}=7$ TeV Using the ATLAS Detector at the LHC. *Eur.Phys.J.*, C71:1630, 2011.
- [8] ATLAS Collaboration. Measurement of Wgamma and Zgamma Productions and Searches for Technicolor in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector Supporting Note v1.2. Technical report, CERN, Geneva, Jul 2012.
- [9] The ATLAS Collaboration. Atlas homepage. <http://atlas.ch/>.
- [10] ATLAS Collaboration. *Expected performance of the ATLAS experiment: detector, trigger and physics*. CERN, Geneva, 2009.
- [11] ATLAS Collaboration. Expected photon performance in the ATLAS experiment. Technical Report ATL-PHYS-PUB-2011-007, CERN, Geneva, Apr 2011.
- [12] H Abreu et al. Measurement of isolated di-photon cross section in pp collision at $\sqrt{s} = 7$ TeV with the ATLAS detector. Technical Report ATL-COM-PHYS-2011-301, CERN, Geneva, Mar 2011. This is the supporting note for the 2010 SM diphoton cross section paper, approved by ATLAS and submitted for publication.
- [13] ATLAS Collaboration. Purity Estimates for the Inclusive Isolated Photons. Technical Report ATL-COM-PHYS-2010-804, CERN, Geneva, Febr 2011.

- [14] B Resende. Muon identification algorithms in ATLAS. Technical Report ATL-PHYS-PROC-2009-113, CERN, Geneva, Sep 2009.
- [15] ATLAS Collaboration. Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at $\sqrt{s}=7$ TeV. Technical Report ATLAS-CONF-2012-101, CERN, Geneva, Jul 2012.
- [16] G. Aad et al. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys.Rev.*, D83:052005, 2011.
- [17] The ROOT team. The ROOT Users Guide 5.26, Dec 2009.
- [18] H Wang and N Berger. Statistical studies for Higgs to gamma gamma. Technical Report ATL-COM-PHYS-2011-783, CERN, Geneva, Jun 2011.

Erklärung:

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

Heidelberg, den (Datum)

Danksagung

An dieser Stelle möchte ich mich herzlich bei allen Mitgliedern der ATLAS/ILC Gruppe für die freundliche Aufnahme während meiner Zeit als Mitarbeiter in der Gruppe bedanken.

Vielen Dank an Professor Hans-Christian Schultz-Coulon, der mir die Möglichkeit gab, an einer der größten Kollaborationen weltweit aktiv teilzunehmen und schon im Rahmen des Bachelorstudiums „richtig“ zu forschen.

Besonders viel habe ich meinem Betreuer Veit Scharf zu verdanken, der mich zunächst in die $W + \gamma + \gamma$ Analyse einführte und anschließend unzählige Stunden damit verbrachte, mit mir über offene Fragen zu diskutieren und bei Problemen jeder Art zu helfen. Ohne die vielen Ratschläge, Hilfen und das bereitgestellte Analyse-Framework wäre diese Arbeit nicht möglich gewesen. Hierfür herzlichen Dank. Für die weitere Arbeit an der $W + \gamma + \gamma$ Analyse wünsche ich gutes Gelingen und viel Erfolg im (hoffentlich) baldigen Endspurt des Promotionstudiums.

Außerdem möchte ich Dr. Rainer Stamen danken, der sich immer wieder viel Zeit genommen hat, während meine Arbeit in Gesprächen den aktuellen Stand meiner Arbeit zu besprechen und immer wieder Impulse gab, so dass ich einem roten Faden bis zum Ende der Arbeit folgen konnte.

Des weiteren vielen Dank an Thorsten Dietzsch, der sich bei Fragen immer viel Zeit für eine ausführliche Diskussion genommen hat und dessen kritische Herangehensweise an offene Fragen mich immer wieder inspiriert hat.

Abschließend möchte mich an dieser Stelle auch bei meinem Eltern und meiner Familie bedanken, ohne die mein Studium niemals möglich wäre.