Title
Measurement of the Upsilon(1S) Production Cross Section in Proton-Proton Collisions at Center of Mass Energy 7 TeV

Permalink
https://escholarship.org/uc/item/30c9p2f8

Author
Scherzer, Maxwell Ian

Publication Date
2011

Measurement of the $\Upsilon(1S)$ Production Cross Section in Proton-Proton Collisions at Center of Mass Energy 7 TeV

by

Maxwell Ian Scherzer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Beate Heinemann, Chair
Professor Stephen Derenzo
Professor Robert Jacobsen

Fall 2011

Measurement of the $\Upsilon(1S)$ Production Cross Section in Proton-Proton Collisions at
Center of Mass Energy 7 TeV

Abstract

Measurement of the $\Upsilon(1S)$ Production Cross Section in Proton-Proton Collisions at Center of Mass Energy 7 TeV

by

Maxwell Ian Scherzer

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Beate Heinemann, Chair

A measurement of the cross section for $\Upsilon(1S)$ production in proton-proton collisions at center of mass energy 7 TeV is presented. The measurement uses the $\Upsilon(1S) \rightarrow \mu^+\mu^-$ decay mode and covers up to a transverse momentum of 26 GeV in two bins of rapidity; $|y^{\Upsilon(1S)}| < 1.2$ and $1.2 < |y^{\Upsilon(1S)}| < 2.4$. The results are based on an integrated luminosity of 1.13 pb^{-1}, collected with the ATLAS detector at the Large Hadron Collider. Templates derived from data are used to model the continuum background. In order to eliminate theoretical uncertainties due to the unknown polarization, the measurement is restricted to the case where both muons have a transverse momentum of $p_T^\mu > 4$ GeV and a pseudorapidity $|\eta^\mu| < 2.5$. The results are compared to two theoretical predictions: they agree to within a factor of two with the PYTHIA implementation of NRQCD while disagreeing by up to a factor of ten with the color singlet model at next-to-leading order.

Experimental high energy physics is an incredible human undertaking. It's also pretty hard. My understanding of it was primarily shaped during conversation, debate, and collaboration with

Tobias Golling, Beate Heinemann, Ted Kolberg and Seth Zenz

I dedicate this thesis to them in appreciation.

Contents

List of Figures

List of Tables

Acknowledgments

Joe Kroll, for - as nearly as I can remember the exact phrasing - “You should really consider majoring in physics.”, “I don’t know why you wouldn’t apply to Berkeley too.” and “You guys are getting Beate in a few months. Make sure you work with her.”

Bob Jacobsen, my first advisor on ATLAS, for giving me a third chance to join a research group when very few others were prepared to do so. I’ll always be grateful for that. For the wide ranging and very interesting discussions and advice.

Tobi Golling, for taking me under his wing and helping me get started on my first real ATLAS project when the pixel production lab began to wind down. Athena, job options, local coordinates, biased residuals, unbiased residuals, cluster width, cluster size - the number of nouns in a single corner of high energy physics can be completely baffling at first; for helping me to get started and beyond for his friendship and support.

Marjorie Shapiro, den mother to all the Berkeley ATLAS cubs, for teaching me - especially when first starting out on alignment - to *begin at the beginning* and to never ignore the simple questions; for her benevolent ‘Marj meddling’ during my uncertain start on ATLAS and, more generally, for just caring; for taking considerable time to critique the three most important talks of my grad career; for her wise and unerroring advice.

Andreas Korn, for freely sharing his knowledge of tracking, fitting, muon reconstruction and B-physics with me. This expertise is considerable, and was kindly offered to me on many, many occasions in which I came into his office confused.

Beate Heinemann, first and foremost for taking me on as her student: in my eight years of grad school this was undoubtably my biggest stroke of luck. For showing me how to brush aside the details and cut right to the heart of a problem - often doing so within minutes of first being confronted with it. These demonstrations were a near constant source of education and inspiration! For her positive attitude and work ethic: I do not consider hard work to be fun, yet while working with her I repeatedly found myself loosing sight of this. For suggesting this thesis topic: it was perfectly suited for me. I don’t know how grad school would have turned out without her guidance and encouragement and I’m glad that I’ll never have to.

Mauro Donega, for being a hell of an office mate during my long years at CERN; for always listening to me think through a problem aloud; for teaching me how to be a good working physicist by example.

Ian Hinchcliffe, primarily for funding my regular travel to CERN during the year that this research was carried out. This was pretty crucial! (And for the lively spirit that he brings to group meetings; it does keep them interesting.)

Shih-Chieh Hsu, for being there in the trenches with me during the beginning of data taking, especially during the very stressful month leading upto ICHEP. During this period we were seperated by nine time zones. Yet there seemed to be no time of day during which he was unavailable to talk or instant message about the latest plots. For always being happy to plug a leak; for being an all around great collaborator.

Laurent Chevalier, for teaching me everything I know about ATLAS's most insanely complicated subsystem; for always being up for a coffee on the balcony; for his gentle reminder 'how is the thesis going?' whenever he knew that it wasn't.

Yi Yang - my fellow upsilon grad student - for keeping me on my toes; for his patience during the times in which none of our plots made any sense.

Ana Ovcharova and David Yu, for reading and helpfully commenting on sections of this thesis.

Harish Agarwal, Brad Hagan, Ziggy Kermish, Jeff Moffitt and Andres Morey, for your friendship, especially during the first two years of grad school; an extremely difficult period. Though I know some of you would not willingly be associated with it, I do feel as if a part of this degree is yours.

Florian Hirsch, Ted Kolberg and Rachel Yohay, for the great times, the great food and the great beer; for being my Euro-family.

Lauren Tompkins and Seth Zenz: what to say? Without the two of you physics just won't be quite the same.

My family, for trips to the Franklin Institute, for my science encyclopedias, for the model rocket kits, for going with the whole physics thing even when it was completely unclear where it was leading; for your constant encouragement and help along the way; for your love during the last 31 years. I never could have done it without you.

Measure what is measurable, and make measurable what is not so.
-*Galileo Galilei*

Chapter 1

Introduction

This dissertation studies the forces that act between the constituents of hadrons by measuring the $\Upsilon(1S)$ production cross section in high energy proton-proton collisions.

First, what is the $\Upsilon(1S)$? From a long distance point of view it can be characterized as one of the heavy (flavorless) vector mesons; hadrons with the same quantum numbers as the photon. These are fascinating physical systems that played a significant role in the development of elementary particle physics. Because they couple to e^+e^- and $\mu^+\mu^-$ pairs, they are both copiously produced in electron-positron collisions and, crucially for this research, easily observed in hadron collisions. This is true for all the vector mesons, including the ω, the ρ and ϕ. The heavy vector mesons have the additional property that they are much more massive than the characteristic scale of the strong force, Λ_{QCD}. The significance of this will become apparent shortly.

It is now understood - with the study of the heavy vector mesons playing no small role in this understanding - that all hadrons are bound states of more elementary particles called quarks; Dirac fermions, and thus fundamentally quite similar to the electron. The hadronic analogs of ortho-positronium - the s-wave bound states of quark-antiquark pairs in the triplet spin state - are precisely the vector mesons expounded on above.

Among this already special category, the $\Upsilon(1S)$ is somewhat further distinguished. It is, roughly speaking, the lightest of the heaviest of the vector mesons: the most massive quark, the top quark, does not live long enough to hadronize. The heaviest quarkonia[1] is therefore bottomonia; bottom-antibottom quark ($b\bar{b}$) bound states. (The Υ family is then just a synonym for vector bottomonia.) The lightest vector bottomonium is the $\Upsilon(1S)$.

As the mass of the bottom quark is more than order of magnitude greater than Λ_{QCD}, (approximately 4.5 and 0.2 GeV, respectively) bottomononium is a nonrelativistic system. Massive and easily observed experimentally: together these properties make the $\Upsilon(1S)$ a system *par excellence* with which to study the production of quark

---

[1] The somewhat colloquial expression quarkonium follows from analogy with positronium.

bound states.

The elementary forces that act between quarks are understood, and the description of these forces, Quantum Chromodynamics (QCD), is well established experimentally [1, 2, 3, 4] and [5]. According to QCD, the forces between quarks are described by a Yang Mills theory with a non-Abelian gauge group. These theories generalize electricity and magnetism in such a way that the gluon - the generalized photon - is itself charged.

Knowing the fundamental principles behind quark interactions does not, however, make the proton-proton → Υ cross section straight forward to obtain. There are essentially three reasons for this. First, even in the case of positronium, it is necessary to solve the Schroedinger equation in order to obtain the bound state wave function. In the case of QED this amounts to solving the r^{-1} potential but in QCD the exact 2-body potential is not known. Second, the formation of bound states with constituents nearly at rest is an inherently long wavelength / low momentum process. QCD becomes strongly coupled at low momentum transfer and thus perturbation theory becomes inadequate. Finally, QCD is simply an inherently difficult theory to calculate with: even small numbers of diagrams can quickly lead to hundreds or thousands of individual terms.

Several approaches have been proposed to cope with this challenge. Though they all begin from QCD, they differ by making different approximations and taking different limits. In the case of Υ production, these differences can lead to predictions that span more than an order of magnitude.

This dissertation will test some of these ideas by presenting a measurement of the $\Upsilon(1S)$ production cross section in proton-proton collisions at the CERN Large Hadron Collider using the ATLAS detector. ATLAS was not built to study quarkonia. Rather, it's primary purpose is to elucidate the mechanism of electroweak symmetry breaking. More generally, it was designed to search for new phenomena at scales from hundreds of GeV to the low TeV range. However, the very nature of these searches necessitates an instrument capable of efficient and accurate detection of a broad range of final states over energies and momenta spanning several orders of magnitude.

Chapter 2 presents an overview of heavy quarkonium and QCD. Chapter 3 describes the ATLAS detector and indicates why, given the central aims of its physics program, it was designed such at it was. Chapter 4 is essentially an overture; while omitting many important details it defines the measured quantities and outlines the measurement strategy. The analysis cleanly breaks down into two parts, the first of which is the subject of Chapter 5; the measurement of detector efficiencies. The second part of the analysis centers on fits of reconstructed mass spectra and is the subject of Chapter 6. Systematic uncertainties, both those associated with efficiencies and those associated with the mass fits, are discussed in Chapter 7. Results and comparison with two theoretical predictions are presented in Chapter 8.

Chapter 2

Theoretical and Historical Background

The $\Upsilon(1S)$ was discovered in the summer of 1977 [6]. However by that time, much of the physics of heavy quarkonium was already understood from studying the charmonium system, the first state of which had been discovered almost three years previous [7, 8].

Meanwhile during the same period that the charmonium system was first being explored, an excess of anamolous events recorded with the SLAC-LBL detector was interpreted as pair production of a third species of charged lepton with mass between 1.6 and 2 GeV [9]. This interpretation was controversial for several years. However if correct, it would be suggestive of a third generation of quarks as well.

Additional evidence for a third quark generation - though of a somewhat more indirect nature - was provided by an experiment involving the neutral kaon system [10]. These experiments demonstrated for the first time processes that are not invariant under CP i.e. processes that do not occur at the same rate upon replacing every particle with its antiparticle (charge conjugation, C) of opposite helicity (parity, P). In 1973 - the experiment itself had been carried out a decade earlier - it was pointed out by Kobayashi and Maskawa [11] that CP violation could naturally be accomodated with, as they put it, a six-quark model.

Thus while the actual observation of the $\Upsilon(1S)$ was significant, it naturally sits within a chain of events beginning some years earlier.

2.1 "The New Degree of Hadronic Freedom... is Charm"

The fall of 1974 witnessed a discovery so significant that the period has since become known as 'The November Revolution.' It has been recounted in detail many times[1] and will not be overly dwelt on here. Briefly, the events centered on the observation of a resonance with a mass of 3.1 GeV, the J/ψ. This was considerably more massive than any particle that had previously been discovered yet the width indicated that it was extremely long lived compared with other particles that decayed strongly.

Subsequent experimental and theoretical advances established that the J/ψ is the 1S triplet bound state of a pair of heavy quarks with charge $\frac{2}{3}e$. Because the detailed properties of charmonium[2] provide some of the most simple, yet compelling, reasons for believing in the quark model, some of them are reviewed here. They can be grouped into (at least) three broad lines of evidence.

First, the two body bound state model can be used to derive relationships between the decay rates of charmonium to light hadrons and the bound state wave function:

$$\Gamma(J/\psi \rightarrow hadrons) \simeq \Gamma(J/\psi \rightarrow ggg) = \frac{160}{81}(\pi^2 - 9)\frac{\alpha_s^3}{M_{J/\psi}^2}|\phi(0)|^2 \tag{2.1}$$

(The bound state wave function has been denoted by ϕ rather than use the symbol ψ twice.) This equation follows from analogy with the rate of ortho-positronium decay to three photons. The wave function at zero separation, $|\phi(0)|^2$, can be determined from the partial decay width to e^+e^-. A similar relationship exists for the 1S singlet state, the η_c, in analogy with the para-positronium decay to two photons.

$$\Gamma(\eta_c \rightarrow hadrons) \simeq \Gamma(\eta_c \rightarrow gg) = \frac{32\pi}{3}\frac{\alpha_s^2}{M_{\eta_c}^2}|\phi(0)|^2 \tag{2.2}$$

These are both approximately satisfied.

The above argument involves only the s-states. But if the J/ψ is really charmonium there should also be p-wave states that are related to the s-states by single photon transitions. The 2^3P_0, 2^3P_1, 2^3P_2 have all been identified and are called χ_{0c}, χ_{1c} and χ_{2c}. They are evident in e^+e^- collisions at center of mass energy equal to the mass of the 2S triplet charmonium, the ψ', in which they appear as resonances in the $J/\psi + \gamma$ invariant mass distribution. These are summarized in Fig. 2.1; all indicated transitions have been observed in cascade decays of the form $\psi' \rightarrow \chi\gamma \rightarrow J/\psi\gamma\gamma$. The $\psi' \rightarrow \chi\gamma$ decays have also been seen using events in which the χ decays hadronically.

As expected - they lack the correct quantum numbers needed to couple to virtual photons - these states do not show up directly as hadronic resonances in e^+e^- collisions.

A final (not exhaustive, but final for this summary) notable piece of the puzzle is that if the states at ~3 GeV are charm-anticharm bound states, there should also

---

[1] See for instance [12, 13]. The latter is the source of the above quotation as well as most of the discussion in this section.

[2] From the vantage point of 2011 the same can be said of bottomonium but it was during the study of charmonium that these ideas first coagulated.

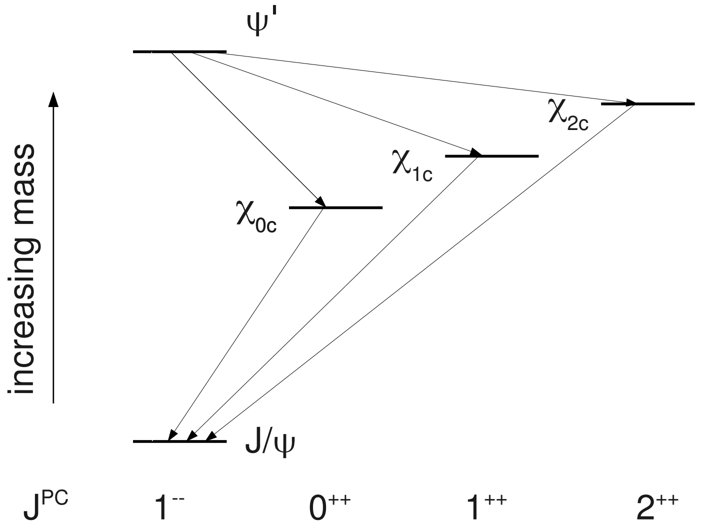

Figure 2.1: Grotrian diagram showing radiative transitions between ${}^3S_1$ charmonium states (the J/ψ and ψ') and several 2^3P_J states. The listing is not exhaustive; additional states and transitions are known. Adapted from [14].

exist 'charmed kaons' at a mass of approximately 1.5 GeV. That is, there should exist particles made up of one charm quark and one light quark possessing an overt charm quantum number and thus stable against both strong and electromagnetic decays. These were observed [15] through their decays $D^0 \rightarrow K^-\pi^+$ and $D^0 \rightarrow K^-\pi^+\pi^+\pi^-$ in 1976.

2.2 The Upsilon Family and Bottom Quark

Several years later the story was closely repeated at a mass scale approximately three times higher. To quote [16] 'Much of the theoretical and experimental development of the physics of the fifth quark was presaged by the earlier history of the fourth quark.' The J/ψ was co-discovered by studying the mass spectrum of e^+e^- pairs produced in proton collisions on a fixed target using a double armed spectrometer. Similarly, the Υ states were first seen by a double arm spectrometer studying $\mu^+\mu^-$ pairs produced in fixed target proton collisions.

As in the study of charmonium, general 4π detectors (see Ch. 3.1) at electron-positron machines made it possible to perform detailed measurements of masses and transition rates between the new states. Due to the increased mass of the bottom quark relative to the charm quark, two bottom quarks sit lower in the 2-body potential compared to the corresponding charmonium state. This means there is more room for additional states before the energy of the bound state exceeds the threshold required for disassociation through the production of a pair of light quarks. Where the charmonium system has two s-wave triplet states below this threshold, the J/ψ and the ψ', the bottomonium system has three, the $\Upsilon(1S)$, the $\Upsilon(2S)$ and the $\Upsilon(2S)$[3]. Mesons analogous to the kaons and D mesons containing one light quark and one B quark were also identified and have the expected properties [17].

2.3 QCD

2.3.1 Yang-Mills and the QCD Lagrangian

The laws describing the elementary interactions of quarks were unraveled during approximately the same time period. These laws take the form of a particular quantum field theory called Quantum Chromodynamics (QCD). Remarkably the discovery of the QCD Langrangian - what, with only small abuse, can be called the QCD force law - marked not the final step in identifying the correct theory, but in some sense *one of the first steps.* This is because the full phenomenological predictions of this field theory are highly non-trivial to work out starting from just the Lagrangian.

However, this first step, taken by Yang and Mills [18] was an enormous advance. What it did was essentially subsume two more 'primitive' ideas into a theory having the properties of both; the local gauge invariance of Quantum Electrodynamics (QED), and the isospin invariance seen in the physics of pions and nucleons. The local gauge invariance of QED refers to the invariance of the Lagrangian of the theory

$$\mathcal{L} = \bar{\psi} i \gamma^{\mu} (\partial_{\mu} - ieA_{\mu}) \psi - m\bar{\psi}\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.3}$$

under the transformations $\psi(x) \to e^{i\alpha}\psi(x)$, $A_{\mu} \to A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha$ where, crucially, α need not be a constant, $\alpha = \alpha(x)$.

Isospin invariance is Heisenberg's idea that protons and neutrons should be viewed as different S_z eigenstates of the same spin 1/2 particle. The rotations which turn one into the other take place in an abstract internal space, so the spin operators in this space are called $\overrightarrow{I}$ and I_z. The pions are likewise viewed as different states of the same spin 1 particle.

By itself these identifications have no consequences. The non-trivial part is the supposition that the Hamiltonian governing pion-nucleon scattering is invariant under rotations in the internal space i.e. pion-nucleon scattering is iso-spherically symmetric. Thus, whatever the 'true' theory is, it should be invariant under $N \to D^{(1/2)}N$ and $\pi \to D^{(1)}\pi$, where N is the nucleon spinor, π is the pion vector and $D^{(1/2)}$ and $D^{(1)}$ are rotation matrices representing the same rotation.

---

[3] Likewise, in addition to the 2P states visible in $\Upsilon(2S) \to \Upsilon(1S)\gamma\gamma$ cascade decays, there are a set of 3P states between the 2S and the 3S.

What Yang and Mills did was to exhibit a theory of two Dirac fermions, which they took to be the nucleons, in which isospin invariance is a local symmetry. The Lagrangian is given by

$$\mathcal{L} = \bar{\psi} i\gamma^\mu (\partial_\mu - igA^k_\mu \frac{\sigma^k}{2})\psi - m\bar{\psi}\psi - \frac{1}{4}F^k_{\mu\nu}F^{k\mu\nu} \tag{2.4}$$

where ψ is a column vector of Dirac fermions, $(\psi_{proton}(x), \psi_{neutron}(x))^T$, A^k_μ indicates three vector fields indexed by k and σ^k are the 3 Pauli matrices familiar from elementary quantum mechanics. The field strength tensor is generalized to $F^k_{\mu\nu} \equiv \partial_\mu A^k_\nu - \partial_\nu A^k_\mu + g\epsilon^{klm}A^l_\mu A^m_\nu$.

This is invariant under the simultaneous transformations $\psi \rightarrow e^{i\alpha^k\sigma^k/2}\psi$, $A^k_\mu \rightarrow A^k_\mu + \frac{1}{g}\partial_\mu\alpha^k + \epsilon^{klm}A^l_\mu\alpha^m$, where, as in the case of QED, the three α^ks need not be a constant but can vary with point to point.

The absolutely remarkable aspect of this work is that the physics - the fermion-fermion interactions - was derived by starting from the notion that a geometric principle should be satisfied!

The QCD Lagrangian is obtained immediately from Eq. 2.4 upon replacing the group of isospin rotations, SU(2), with SU(3). The 3 Pauli matrices are replaced by a set of matrices that generate SU(3) - conventionally the Gell-Mann matrices - and ψ becomes a triplet of Dirac fermions describing quarks, one for each color. The introduction of this new quantum number is necessary even irrespective of dynamical considerations; as pointed out in [19] it otherwise leads to inconsistencies between the Pauli exclusion principle and the quark model of baryons.

2.3.2 QCD as a Theory of Quark Interactions

The road between Yang-Mills and the acceptance of QCD as a theory of quark interactions was long and tortuous. Indeed it took more than a decade until it was understood how to systematically quantize these systems [20]. The climactic theoretical step was the discovery [21, 22] that QCD is asymptotically free; at increasingly high energy the behavior of the theory approaches that of a free field theory in which the particles do not interact.

For QCD, their result is

$$\mu\frac{d\alpha_s}{d\mu} = \frac{\alpha_s^2}{\pi} \times (\frac{-11}{2} + \frac{n_f}{3}) \tag{2.5}$$

where α_s ('s' for strong) is analogous to the fine structure constant; it measures the strength of the interaction. The number of quark flavors is n_f and is at any known energy scale effectively 6 or less.

The above equation is in some sense an intermediate result and not a relationship between physically meaningful numbers because μ is a 'crutch' [23]; an auxiliary variable introduced to parametrize divergent quantities.Given a convention for defining μ however, this behavior can be related to experimental observables. Fig. 2.2 shows a determination of $\alpha_s(\mu)$ from inclusive jet rates in e^+p collisions as measured by the H1 collaboration [24]. An increase in the independent variable corresponds to an increase in the momentum exchange between the positron and the proton or more energetic jets. In either case the coupling monotonically decreases. This, in a nutshell, is asymptotic freedom; a free theory corresponds to a coupling of zero.

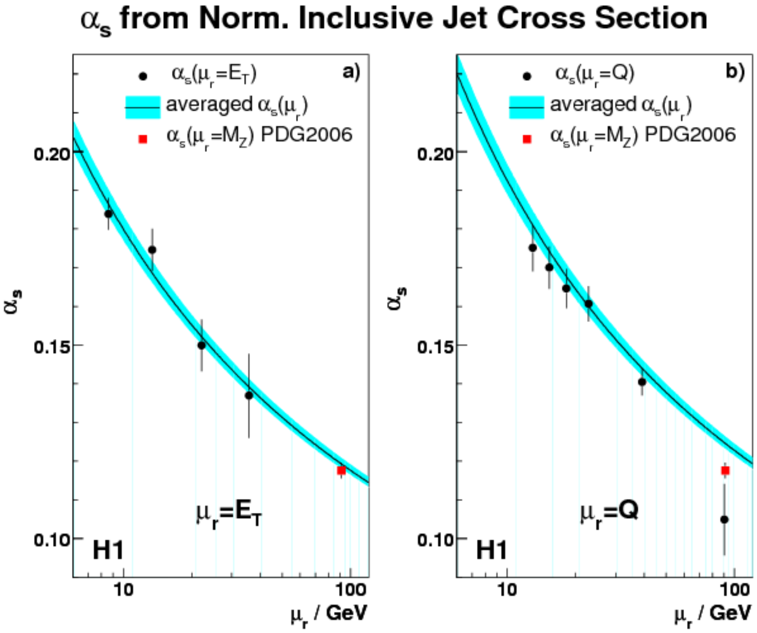

Figure 2.2: Results from the H1 Collaboration showing the strong coupling constant as a function of the renormalization scale. From the original caption: "The error bars denote the total experimental uncertainty for each data point. The solid curve shows the result of evolving $\alpha_s(M_Z)$ averaged from all Q^2 and E_T regions, with the band denoting the total experimental uncertainty." The left and right plots differ in how the renormalization scales - the 'crutch' described in the text - are defined.

This result at once explained a number of confusing experimental results at the time, among the most pressing of which (at least to one of the discovers) were experiments studying high energy electron - nucleon scattering [26]. These indicated that the constituents of the nucleons were only weakly bound together, an observation difficult to reconcile with the absence of freely observed quarks. Asymptotic freedom showed that known models could naturally account for both results.

2.3.3 Hadronic Initial States and PDFs

Because the coupling constant is small at high energies[4] QCD can be used to calculate quark and gluon scattering cross sections in a perturbative series in α_s in the same way that QED can be used to calculate processes involving electrons and photons as a series in α_{EM}.

Some processes, e.g. the inclusive production of hadrons in e^+e^- collisions, can then be described from first principles alone. However when the initial state includes one or more hadrons, additional input data is required. Intuitively, the required data is the 'quark beam' and 'gluon beam' equivalent of one hadron. For instance, given a 7 TeV beam of protons of a given instantaneous luminosity, what is the instantaneous luminosity of gluons with energy between (say) 70 GeV and 70+δE GeV?

This data is collectively called Parton[5] Distribution Functions (PDFs) [28, 29]. The conventional notation for the PDF of a hadron h is $f_{i/h}(x, Q^2)$. The quark species (or gluon) is indexed by i. The fraction of the hadron's energy carried by the quark or gluon is x. The second argument, Q^2, is the momentum exchange of the process. (That the PDF depends on this at all means, intuitively, that the hadron looks different depending on the spatial resolution of the probe.)

Many processes can be used to constrain PDFs but in practice the most useful is electron or positron - nucleon scattering [30]. (To achieve high center of mass energies the nucleon must be a proton so it can be accelerated.) The reason this process is so useful is that, excluding higher order QCD corrections, electron-quark scattering is an electroweak process which can be calculated very accurately. Thus there is little ambiguity in relating the observed cross sections to the unknown PDFs.

2.4 NRQCD and Quarkonium Production

Calculating quarkonium production cross sections from first principles, i.e. from QCD, is very challenging. Unlike a high energy scattering process (quark - antiquark annihilation to two gluons, for example) in which the sole energy scale is given by the $q\bar{q}$ center of mass energy, quarkonium production is a process involving many scales. For bottomonia, studies of the two body potential suggest that the velocity of the constituent quarks relative to the meson center of mass, v, is ≈ 0.1. Thus the quark rest mass M, the inverse size of the meson Mv, and the kinetic energy of the constituent quarks Mv^2 span two orders of magnitude.

---

[4] A fairly natural place to quote it is at the mass of the Z^0 boson where, depending on precisely what techniques and data are used, the value [27] is $\alpha_s(M_Z) = 0.1184 \pm 0.0007$

[5] The terminology stems from a time when it was not completely accepted that the elementary constituents in hadronic scattering processes ('partons') were synonymous with quarks. The latter had been originally postulated just to explain the spectrum of observed hadrons. Today they would probably just be called Quark Distribution Functions.

Nonrelativistic QCD (NRQCD) [31] is an effective field theory that systematically separates out processes with p≈M from longer wavelength modes. The kinetic energy terms for the gluons and the light quarks are unchanged in the Lagrangian. For the heavy quarks, the 4-component Dirac spinor is split into two 2-component Pauli spinors; the 'large' and 'small' components. (This is identical to the formalism used to describe the hydrogen atom starting from the Dirac equation.) The heavy quark kinetic energy is then decomposed as $\mathcal{L}_{heavy} + \delta\mathcal{L}$ where

$$\mathcal{L}_{heavy} = \psi^\dagger \left(iD_t + \frac{\overrightarrow{D}^2}{2M} \right) \psi + \chi^\dagger \left(iD_t + \frac{\overrightarrow{D}^2}{2M} \right) \chi \tag{2.6}$$

D_t and $\overrightarrow{D}$ are the time and space components of the covariant derivative, $D_\mu \equiv \partial_\mu - igA_\mu$. Terms such as $\psi^\dagger g A_0 \psi$ describe heavy quark - gluon interactions, but there are no terms bilinear in $\psi^\dagger\chi$ or $\chi^\dagger\psi$: the creation and annihilation of a $Q\bar{Q}$ pair is an inherently relativistic process that has been integrated out by definition.

The remaining part of the Lagrangian, $\delta\mathcal{L}$, reproduces the relativistic effects of full QCD as an infinite series of additional terms. For instance via the optical theorem, the process of quarkonium annihilation to light hadrons (LH) can be accommodated by 4-fermion operators:

$$\Gamma(\mathcal{Q} \to LH) = 2\mathrm{Im}\langle \mathcal{Q} | \delta\mathcal{L}_{4-fermion} | \mathcal{Q} \rangle \tag{2.7}$$

There are an infinite number of 4-fermion operators of increasing dimension. Because $[\psi] = \frac{3}{2}$ the lowest possible dimension is 6. In this case the color singlet ${}^1S_1$ operator is proportional to $\psi^\dagger\chi \cdot \chi^\dagger\psi$ and the color singlet ${}^3S_1$ to $\psi^\dagger\overrightarrow{\sigma}\chi \cdot \chi^\dagger\overrightarrow{\sigma}\psi$.

These operators, like all operators in NRQCD, can be characterized not just by their power in g or α_s but by their power in v as well e.g. a term with the 'gluon magnetic field' $B^i = \frac{1}{2}\epsilon^{ijk}F^{jk}$ is one higher power in v than a similar term with $E^i = F^{0i}$ because by assumption the system under study is non-relativistic.

The cross section for quarkonium production can then be written as the product of the cross section to produce a heavy quark-antiquark pair in a state n and the matrix element for such a pair to hadronize into a quarkonium particle $\mathcal{Q}$. Integrating over the quark and gluon PDFs as well, the cross section for quarkonium production in proton-proton collisions can be written as

$$\begin{aligned} \sigma(pp \to \mathcal{Q} + X) &= \sum_{i,j,n} \int dx_1 dx_2 f_{i/p} f_{j/p} \\ &\times \hat{\sigma}\left[ij \to (Q\bar{Q})_n + X\right] |\mathcal{M}^n_{\mathcal{Q}}|^2 \end{aligned} \tag{2.8}$$

The hadronization process is described by $|\mathcal{M}^n_{\mathcal{Q}}|^2$ and is not calculable perturbatively, either in QCD or NRQCD. The calculation of the short distance part - the creation of the $(Q\bar{Q})_n$ state - *is* calculable and is the motivation for the formalism described above.

Chapter 3

The ATLAS detector

3.1 Common Features of 4π Detectors

Although ATLAS [32] is a one-of-a-kind scientific instrument, it shares many features in common with other 4π detectors[1]. The most obvious feature of all such detectors is their cylindrical symmetry around the beamline. As a particle propagates out from the interaction point, it crosses successive layers of instrumentation in the form of nested cylinders of increasing radii. The individual layers can in most cases be coarsely grouped into three sub-detectors: an inner tracker, a calorimeter, and a system of muon detectors.

The modern inner tracker is the evolutionary descendent of the bubble chamber. It determines the momentum of charged particles by measuring the radius of curvature of the particle in a known magnetic field. The trajectory of the particle - and thus the curvature - is deduced by making multiple position measurements, in most cases by a regular array of gaseous drift tubes or silicon wafers. The magnetic field is generated by a solenoid, usually located between the tracker and the calorimeter.

After the tracker, the next layer out is the calorimeter. Calorimeters and trackers can in many ways be viewed as orthogonal approaches when it comes to the identification and kinematic measurement of elementary particles. Calorimeter resolution becomes better at higher energies/momenta; tracker resolution becomes worse. With trackers, more material results in more multiple scattering and thus a poorer ability to measure the original trajectory. With calorimeters, the more radiation and interaction lengths, the better: the basis of the measurement is the deposition of all or most of the energy of the particle in the active medium.

The outermost layer in the generic 4π detector is a series of chambers for the detection of muons. Among stable charged particles, muons are unique in the depth

---

[1] '4π' indicates that the detector covers most of the 4π solid angle surrounding the interaction point. The terminology is historical and stems from the time when many experiments made measurements over only small fractions of the full solid angle. Today, most (though not all) detectors at the forefront of experimental high energy particle physics are 4π detectors.

which they can penetrate matter. This is because of their unique properties: like the electron, they do not participate in the strong interaction. They are, however, considerably more massive than electrons and thus loose less energy per unit length as they scatter off of atomic electrons. The result of this is that the identification of muons can be accomplished in a relatively straightforward manner by the use of instrumentation placed behind the calorimeter (relative to the interaction point). The instrumentation technology itself varies widely; in ATLAS alone there are four separate detector technologies in the muon system.

In the current generation of general purpose detectors - the two general purpose detectors at the LHC - the notion of muon chambers at radii outside of the calorimetry has been taken one step farther. Instead of being solely devoted to particle identification, the muon systems of ATLAS and CMS are embedded in magnetic fields, like the generic inner tracker discussed above. Thus each experiment can perform two separate momentum measurements on muons; one using the tracker and one using the dedicated muon system. (Though at the outset it should be noted that this capability is essentially[2] not used in the present analysis.)

Fig. 3.1 shows diagrams of two 4π detectors (neither of which are ATLAS). They are chosen from nearly ‘opposite ends of the spectrum’. The SLAC-LBL magnetic detector was an early 4π detector dating back to the early 1970s, and it studied e^+e^- collisions. The Compact Muon Solenoid (CMS) is an LHC general purpose detector - ATLAS’s ‘sister experiment’ - and is thus at the forefront of design for hadron collider detectors.

In the SLAC-LBL detector, the spark chambers and the shower counters are, in modern parlance, the tracker and the calorimeter. With these identifications, the common features are apparent. Both detectors conform to the basic design discussed above, with an inner tracker surrounding the interaction point followed by a calorimeter, and finally a muon system.

---

[2] The muon system momentum measurement is *implicitly* used in that it enters into one of the algorithms used to identify muons for the analysis. However all of the kinematic information entering in the analysis of selected events uses only the inner tracker.

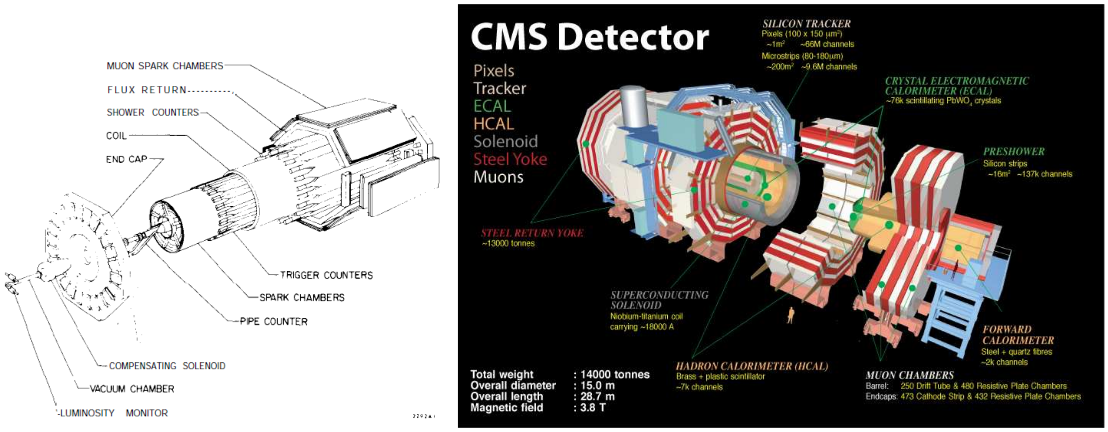

Figure 3.1: The left diagram [34] is of the SLAC-LBL magnetic detector, an experiment at the SLAC SPEAR storage ring. The right diagram [35] is of the Compact Muon Solenoid, an experiment at the CERN Large Hadron Collider.

3.2 Distinguished Features of ATLAS and the Challenges of the LHC

Fig. 3.2 shows a diagram of the ATLAS detector[3]. The pixel detector, semiconductor tracker, and transition radiation tracker together comprise the Inner Detector; the inner tracker. The enormous scale of ATLAS is evident by the two adults shown standing on the shielding (the cone surrounding the beam) on the left side of the figure.

While the general layout is seen to conform to the discussion of the previous section, ATLAS and CMS have unique capabilities associated with the enormous energy and instantaneous luminosity of the LHC [36]. The closest point of contact with which to make comparisons is the Tevatron, a proton-antiproton collider at FNAL (Fermilab), and its two detectors, CDF and D0. The center of mass energy of the Tevatron is 1.96 TeV. CDF and D0 are thus optimized for the measurement of electrons and photons, muons, hadronic jets, etcetera upto energies of several hundred GeV. The design energy of the LHC is 14 TeV (though it is currently operating at 7 TeV). To fully exploit the opportunities afforded by such a collider ATLAS and CMS must make accurate measurements of muons in the TeV range, a considerable challenge. (The same is true of electrons, photons and jets, but only the issue of

---

[3] This is a good place to dispense with some definitions. The z coordinate axis is essentially the beamline. The x and y axes are orthogonal to the z axis and each other. (Their orientation within the x-y plane will not be relevant.) The azimuthal angle, $\phi \equiv \tan^{-1}(\mathrm{y}/\mathrm{x})$, and the polar angle, $\theta \equiv \tan^{-1}(\sqrt{\mathrm{x}^2+\mathrm{y}^2}/\mathrm{z})$, are standard definitions. Much less standard is a reparametrization of the polar angle motivated by the physics of hadron collisions: the pseudorapidity, η, is defined as $-\ln(\tan(\theta/2))$. Note that $\eta = 0$ corresponds to a right angle to the z axis.

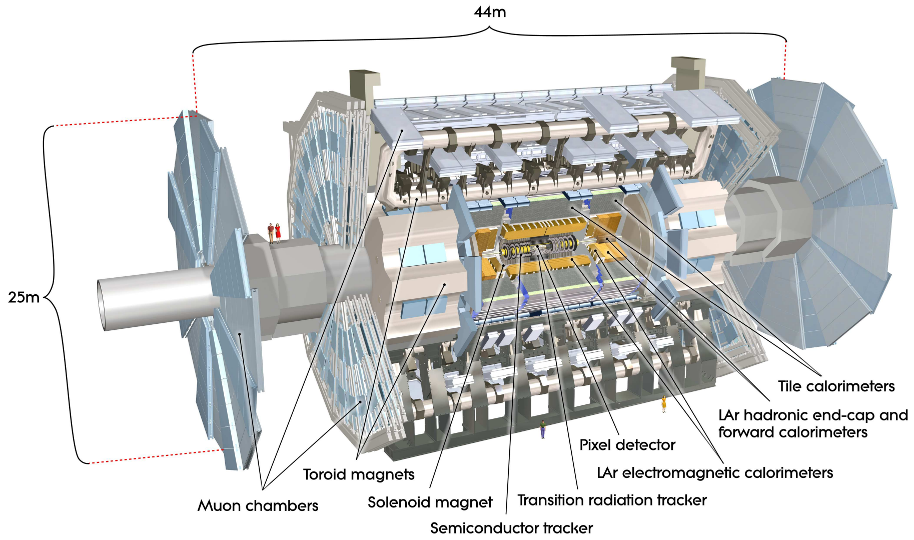

Figure 3.2: A cutaway of ATLAS showing the various sub-detectors. On standard A4-sized paper, the pixel detector extends over less than 2 millimeters.

muon resolution will be expounded on in what follows.)

Before discussing the solution, it will be worth elaborating on the problem a bit more: momentum measurements made with trackers work, as noted, by measuring the curvature[4] of a particle in a magnetic field. This strategy 'runs out' at high momentum because, given the finite curvature resolution of the instrument, tracks of increasingly high momentum suffer increasingly large fractional mismeasurements by the *same* curvature mismeasurement. The limiting example of this effect is that it becomes impossible to distinguish the particle's trajectory from that of a straight line and thus impossible to reliably determine its charge. For fixed single hit precision, the curvature resolution scales inversely as BL^2, where B is the magnetic field strength and L is the lever arm of the detector; larger magnetic fields and larger lever arms allow for particles of increasingly high momentum to be resolved from that of a straight line.

ATLAS's (extremely ambitious) solution to this problem is the system of air core toroids at the heart of the Muon System (MS). This sub-detector tracks muons over a volume of only moderate magnetic field strength, ∼0.5 Tesla, but with a very large lever arm of approximately 5 meters. Compared with the 2 Tesla field strength and 1 meter lever arm of the Inner Detector, this leads an increase in the product BL^2 by a factor of 6.25.

The above considerations center on the energy scale of the ATLAS physics goals.

---

[4] The usage of 'curvature' here is the same as in the differential geometry of curves: in the simplest example, a circle with radius R has curvature R^{-1}.

No less challenging is the problem of coping with the *rate* of interactions. Comparison with the Tevatron is again instructive: as of the summer of 2011, peak instantaneous luminosity at the LHC is almost a factor of 6 higher than at the Tevatron [37, 38] ($2 \times 10^{-33} cm^{-2} s^{-1} / 3.5 \times 10^{-32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$). The design luminosity of the LHC, $10^{-34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, is higher still. Even with a bunch frequency of 40 MHz, this translates to an average of approximately 20 proton-proton collisions per event; the average event will have approximately 1000 tracks in it and events with thousands of tracks will be commonplace.

Coping with these multiplicities is one of the prime motivations for the pixel detector. Although the inner trackers of previous hadron collider experiments have used silicon detector technologies, these were in all cases silicon strips; ATLAS and CMS are the first hadron collider experiments to use pixels. (The two are distinguished by the dimension of the individual sensing elements along the direction parallel to the beam. 'Strips' refers to elements that are centimeters or tens of centimeters long while 'pixels' refers to elements that are on the scale of hundreds of micrometers.) The advantage of using pixels is that the fine spacial granularity resolves ambiguities in the tracking pattern recognition. Another way to phrase the issue is that when a large number of tracks are present, the result of using coarser detector elements is not only poorer resolution: sometimes two particles go through the same strip (or straw) and in these cases there will be an ambiguity in the assignment of hits to tracks. Using pixels, as supposed to strips, greatly reduces the rate of this phenomenon.

To summarize, ATLAS has a basic layout common among many collider detectors. However, it has a number of novel features; the motivation for these novelties lies in the higher energy scale and greatly increased collision rate of the LHC.

3.3 Inner Detector

Fig. 3.3 shows a cutaway of the ATLAS Inner Detector (ID). Its purpose is to measure the momentum vector and the impact parameters[5] of charged tracks traversing its volume; primary tracks originating from the interaction point as well as tracks from the conversion of photons and secondary vertices. It is composed of three nested sub-detectors; the pixel detector is innermost followed by the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Sub-detectors at successively larger radii have increasingly poorer resolution but are more cost affordable. In the $R-\phi$ plane of the detector the intrinsic resolutions of the three types of instrumentation are approximately 10 μm, 17 μm and 130 μm.

Coverage of the ID extends out to $|\eta| = 2.5$, corresponding to an angle of approximately 9.4deg to the beampipe and almost 99% of the solid angle. The inner radii of the first few layers and the strength of the magnetic field determine a lower limit to

---

[5] Roughly, how close the track approaches the interaction point. The precise definition is given shortly.

the transverse momentum range that the ID is sensitive to. Tracks with p_T below this range have radii too small to cross a sufficient number of layers to be identified. This value depends on the exact criteria that is required of the track, i.e. the minimum number of required hits, but in [39] charge multiplicity measurements are performed down to 100 MeV.

However of all the figures that can be used to characterize the ATLAS ID, the two of greatest consequence to this analysis are the multiple scattering term to the momentum resolution and the asymptotic curvature resolution, which have been measured [40] using cosmic ray data as 1.6% and 0.53 TeV^{-1}, respectively. The meaning of these figures is that particles with true $p_T = \overline{p_T}$ have a fractional resolution, σ_{p_T}/p_T, of 1.6% $\oplus$ ($\overline{p_T} \times 0.53$ TeV^{-1}). This immediately allows for an estimate of the two track mass resolution of the ID at the mass of the $\Upsilon(1S)$. The mass of a two particle system with labels + and - is given by $m^2 = 2p_+p_-(1 - cos(\theta))$. When the two particles are massless with back-to-back momentum of equal magnitude p this gives $\sigma_m = \sqrt{2}\sigma_p$. Setting $p = p_T$ to further simplify and combining equations gives an $\Upsilon(1S)$ mass resolution of 109 MeV. In the case of $\Upsilon(1S)$ events at small rapidity this is remarkably consistent with what is observed in data!

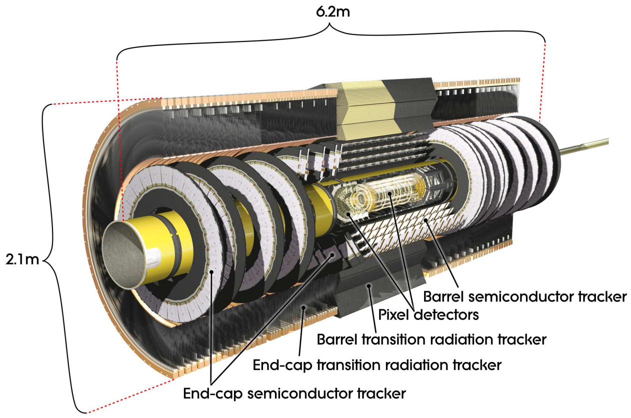

Figure 3.3: A cutaway of the ATLAS Inner Detector. The yellow tubing encloses the *pixel package*; the pixel detector as well as the optical data link electronics. The beam pipe is evident in the anterior of the drawing.

3.3.1 Pixel Detector

A single pixel module is shown in Fig. 3.4. The sensor is the active material: particles propagating through it deposit energy as they ionize atomic electrons. Collection of this charge forms the basis of a 'hit'; the registering of a particle through a single sensing element. The 16 FEs per module are custom integrated circuits, each of which reads out 2,880 pixels. The primary function of the FEs is to amplify and discriminate the charge collected from the sensor. The discriminator voltage and the feedback current on the amplifier are programmable. What this means in practice is that both the minimum charge threshold required to detect a hit and the mapping between digital counts and the charge of the hit are adjustable. Communication between the FEs and the readout drivers in the counting room is performed by the module control chip (MCC); it distributes clock and triggers to the FEs and assembles the hits from the 16 FEs into a common data structure that is then sent off detector.

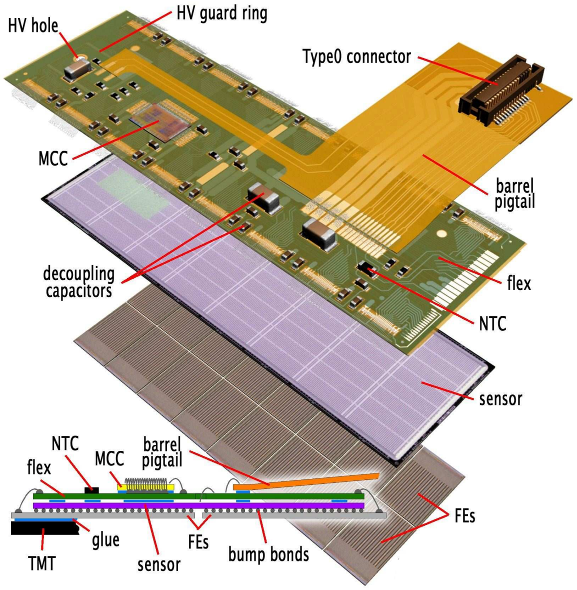

Figure 3.4: A telescoped schematic of a single pixel module showing the three constituent layers. Unlike most silicon strip modules, the front end electronics (FEs) extend over the same area as the sensor.

The pixel detector is then contructed from 1,744 individual modules. Approximately 80% of these are arranged in three concentric cylinders. The remaining 20% are arranged into six discs, three on each side of the interaction point. The innermost cylindrical layer is at a radius of 50.5 mm; this is the very first layer of instrumentation in ATLAS.

As note above, the pixel detector is crucial to making tracking feasible when the LHC runs at high luminosity. Although this analysis uses data taken only at low luminosities, the pixel detector still plays a primary role because of the precision with which it can measure track impact parameters.

Upto a sign, the definition of the transverse impact parameter of a track, d_0, is the distance of closest approach of the track to the line that intersects the primary vertex and is parallel to the z axis[6] This is illustrated in Fig. 3.5. The longitudinal

---

[6] There are multiple possibilities for defining $\mathrm{sgn}(d_0)$. In ATLAS the definition is that

impact parameter, z_0, is defined as the Δz between the point of closest of approach of the track to the aforementioned line and the primary vertex. (These quantities are sometimes referred to as d_0 and z_0 with respect to the primary vertex, since analogous definitions are possible for other reference points such as the beam spot and the global coordinate system origin. However in this analysis the point of reference is always the primary vertex.)

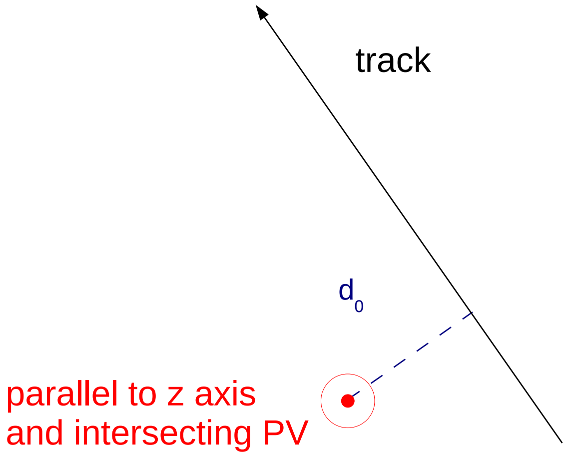

Figure 3.5: Sketch illustrating the definition of the transverse impact parameter, d_0. The z axis is orthogonal to the plane of page. For simplicity the track has been drawn with vanishing curvature but will in general project to a circle.

The motivation for using these quantities is discussed in Sec. 5.5. Here it is only noted that without the pixel detector, the impact parameter resolution would be *significantly* degraded. Any quantitative figure of exactly just how degraded depends on the alignment of the detector and the details of the track sample, but studies using cosmic ray data taken in 2009 are instructive. Cosmic rays are unique in leaving hits in opposing sides of detector. By performing separate track fits to the hits in the upper and lower hemicylinders and comparing the resulting track parameters, cosmic ray data allows for a simple method to measure track resolutions. The result of one such study [41] is shown in Fig. 3.6. This shows the d_0 resolution of tracks with and without pixel hits as a function of track p_T. The high p_T asymptotic resolution of the two categories is determined to be 29 μm and 108 μm respectively.

$\mathrm{sgn}(d_0)=\mathrm{sgn}(l_z^{PV})$ where l_z^{PV} is the z component of the angular momentum of the track with respect to the primary vertex. Thus $d_0 > 0$ in Fig. 3.5. This convention is not relevant in anything that follows.

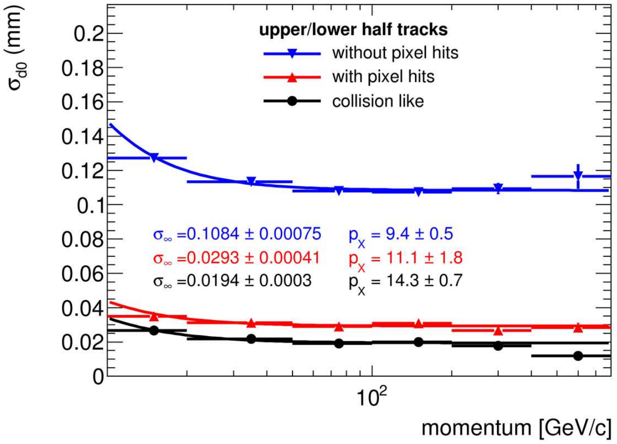

Figure 3.6: The d_0 resolution measured in-situ with cosmic rays. At low p_T the degraded resolution due to multiple scattering is evident. Tracks classified as 'with pixel hits' and 'collision like' differ in their precise selection. In either case an asymptotic high p_T resolution of several tens of micrometers is obtained.

3.3.2 Semiconductor Tracker

Although lacking the fine $R-\phi$ segmentation and small layer radii - and thus the impact parameter resolution - of the pixels, the SCT is in many regards the backbone of the ID. The technology of silicon strips strike a balance between hit precision and affordability which allows for over 63 square meters of sensor coverage[7] extending over a lever arm of 230 millimeters with an $R-\phi$ hit precision of 17 μm.

Constructed from 4,088 individual modules, the SCT has 4 layers in the barrel and 9 in each end-cap. Each layer is double-sided: a second sensor rotated by 40 microradians relative to the first allows for a hit precision in the $R-z$ plane of 580 μm using strips of length approximately 10 centimeters. (There are four different variants of the SCT module with varying strip length.)

3.4 Calorimeter

Although the calorimeter is as central to the ATLAS physics program as the ID or the Muon System, its role in this analysis can be fully characterized by a single number: 3 GeV. That, approximately, is how much energy a muon deposits in the calorimeter at small pseudorapidities. At larger angles from the normal to the beampipe a muon traverses more material and thus deposits more energy. However because the amount of material - assuming cylindrical symmetry - and p/p_T both scale as $\cos^{-1}(\theta)$, the 'p_T tax' is nearly a constant of 3 GeV in the barrel. (In the

---

[7] The analogous figure for the pixels is 1.7 square meters.

end-caps it is considerably less.) An illustration of this is shown in Fig. 3.7: a software model of the detector is used to map the change in p_T of a muon as it crosses the calorimeter. Repeating this for many muons produces a scatter plot that traces the p_T dependence of this effect. The value of the plateau is seen to be at approximately 2.8 GeV. This is the limiting constraint that sets the minimum p_T below which muons cannot be identified with ATLAS.

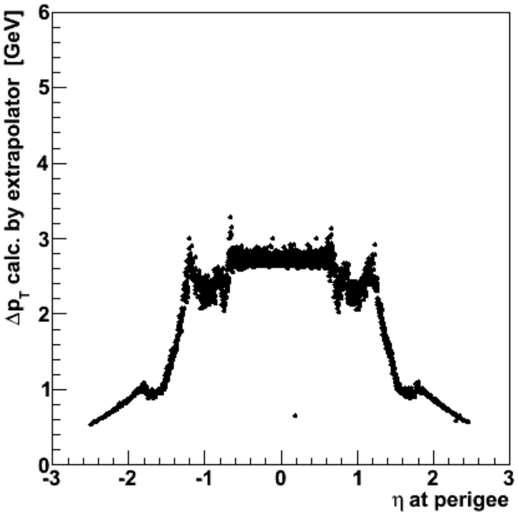

Figure 3.7: Scatter plot showing the change in p_T of muons after traversing the calorimeter. The plot uses a simulation of the detector and makes no use of data.

3.5 Muon System

3.5.1 Trigger Chambers and Triggering on Muons

ATLAS does not record every event. In fact, though LHC beam crossings occur at a nominal rate of 40 MHz, only a few hundred events per second can be written to disk. Events that satisfy one (or more) of the conditions indicating that they warrant closer study are selected by a system of hardware and software collectively called the trigger. The importance of a properly functioning trigger to an experiment's physics program *cannot be overemphasized.*

What sort of event properties indicate potentially interesting event? The ATLAS trigger menu - the list of conditions that have been deemed sufficient to justify keeping an event - has several hundred entries. However most of them first require at least one of four basic conditions: an electromagnetic object (meaning an electron or photon),

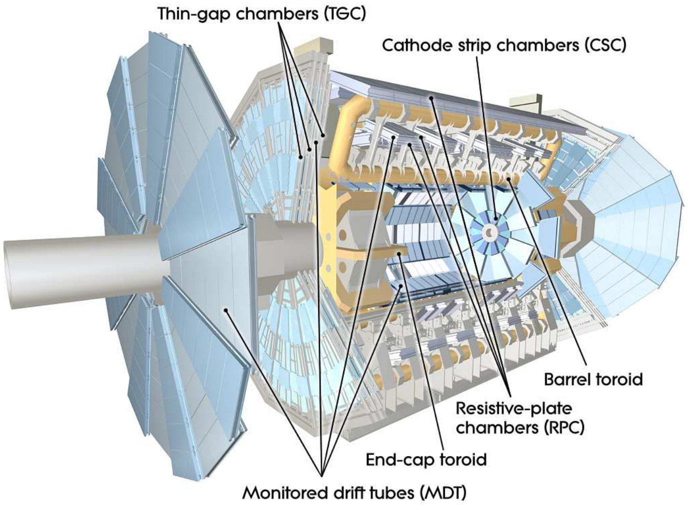

Figure 3.8: A cutaway of the ATLAS Muon System. The MDT stations constitute most of the area of the precision instrumentation. The RPC and TGC layers are used for triggering in the barrel and end-cap respectively. The CSC stations are used at high η only.

a jet of hadrons, an imbalance of energy in the transverse plane of the detector, or, of particular importance to this work, a muon.

The hardware component of the muon trigger system is a series of resistive plate chambers (RPC) and thin gap chambers (TGC) located within the volume of the air core toroid and shown in Fig. 3.8. These technologies were selected for their signal speed at the cost of sacrificing precision. The RPC chambers, for instance, have a precision of only 10 mm but a signal rise time of 1.5 ns. The first step of a muon trigger is a coincidence of several RPC channels that satisfy programmable logic running on custom-built electronics. If there is such a coincidence, the bunch crossing that is read out from the rest of the sub-detectors is determined by the timing of the RPC (or TGC) hits; this necessitates sensors that are fast compared to the 25 ns bunch period.

All of the active components of the Muon System, including the monitored drift tubes (MDT) and the cathode strip chambers (CSC) discussed in the next section, are organized into mechanical components called stations. The length and width of a station (in the barrel, the extent along z and ϕ) range from 1-6 meters and 1-3 meters, respectively, but their height is only tens of centimeters [42]. (They are essentially big panels.) The arrangement of stations in the Muon System Barrel is shown in the diagram in the left of Fig. 3.9.

The diagram in the right of Fig. 3.9 shows a detailed view of three sectors. (A sector refers to a collection of stations positioned at equal ϕ.) The colored panels on one side of the outer stations and either side of the middle stations are the RPC's. Most muon triggers require a coincidence on all three layers. However some of the hardware triggers require hits on only the two middle station layers and can therefore be satisfied by muons with very low momentum. The trigger condition with the lowest threshold is called L1_MU0. The L1 refers to the fact that it is a condition satisfied at level 1 - the hardware level - while the MU0 denotes a single muon with a threshold of '0' GeV. (There is of course a minimum momentum threshold for this trigger, but this is due to the fact that muons with momentum below approximately 3 GeV do not penetrate the calorimeter rather than an explicit cut.)

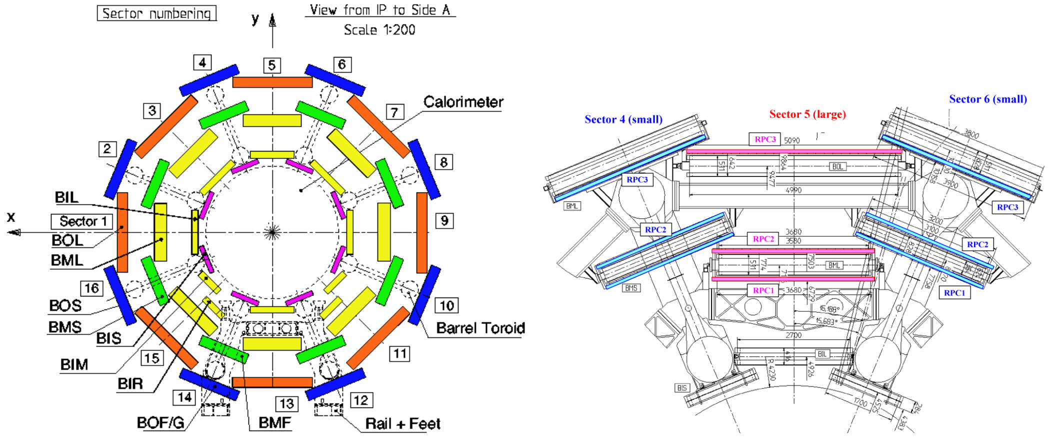

Figure 3.9: The arrangement of the barrel muon stations in the transverse plane (left). The long axis of the toroid magnets are normal to the plane of the page and are indicated by the dotted circles. A close up three barrel stations (right) indicates the placement of the RPC trigger layers.

Excluding runs of the very lowest luminosities, the event is subjected to further examination before being written to disk. This is the task of the high level trigger (HLT). The HLT can be broken down into the level 2 trigger (L2) and the event filter (EF). Both are software based but there is an important difference between them which follows from the different operating requirements of the two levels. The input event rate of the L2 trigger is 75 kHz while the EF input rate, which is same as the L2 accept rate, is only 3.5 kHz. In the case of the EF, there is enough time available per event that an adaptation of the offline algorithms described in [43] can be used to determine if a muon was likely to be present. The L2 decision must be made with much faster dedicated algorithms. In particular, the momentum evaluation is made with look up tables [44].

The complete trigger criteria used to record the data for this analysis can now be stated: the event must satisfy L1_MU0 and have at least one muon with $p_T >$ 4 GeV identified by the event filter. This trigger decision is called EF_mu4. (The instantaneous luminosity was low enough in these runs that the EF could handle the full L1 accept rate i.e. the L2 was not used.)

Two additional pieces of terminology will be used in Chapter 5: a level 1 region-of-interest (RoI) essentially means the collection of trigger hits associated with a level

1 trigger decision. For the present purpose the only important property of a RoI is that it has a direction. Similarly the simplified representation of a muon forming the basis of an EF trigger is called an 'EF feature'.

3.5.2 Momentum Resolution and Precision Detectors

In addition to identifying and triggering on muons, the MS is also a spectrometer i.e. an instrument that makes momentum measurements. The design goal is a momentum resolution of 10% for 1 TeV muons. Given the lever arm and the magnetic field of the MS, this translates into a requirement of a 50 μm sagitta resolution[8]. This is achieved by the MDT and CSC chambers. They compliment the fast response time of the trigger chambers with a precision of 35 μm and 40 μm, respectively.

The MDT chambers, which account for almost 99% of the area of the precision chambers, are drift tubes with a radius 15 mm. Because the MS magnetic field is toroidal, the bending plane is not the $R-\phi$ plane but rather the $R-z$ plane. The MDT tubes are consequently oriented with their axis in the circumferential direction - along ϕ - rather than along z as is commonly the case in inner tracker wire chambers. In the barrel, measurements along ϕ are provided only by the trigger layers.

At angles very small relative to the beampipe ($2.0 < |\eta| < 2.7$) the counting rate at the inner station is too high for the use of MDT's. Instead, the precision instrumentation in this region consists of CSC's. In CSC's the active medium is a gas between a wire and a cathode held at high voltage with respect to each other. However the read out electronics are not coupled to the wires; they're coupled to the cathodes. Alternating CSC layers have cathode strips along alternating axes, so the CSC's provide measurements in both η and ϕ.

3.6 Common Reconstruction Methods

Though every ATLAS measurement and search is ultimately unique, there are a large number of algorithms which can be shared among widely varying analyses. For instance, the software that identifies electron candidates from signals in the calorimeter can be shared between electroweak measurements and searches for supersymmetry. These methods, foundational in the sense that they comprise the first steps in the interpretation of the raw detector signals, are *reconstruction*[9] algorithms.

---

[8] The sagitta of an arc is the maximum perpendicular distance of the arc from the line segment connecting the endpoints. The sagitta and arc length together specify its curvature and because in the context of a tracking detector the latter is simply the lever arm, the sagitta and the curvature are equivalent data.

[9] The term is used quite broadly and applies to any process in which a composite object is built from a group of simpler things. For instance a tracklet indicating a charged particle (but lacking a momentum measurement) can be reconstructed from 3 pixel hits. A pair of top quarks can be reconstructed from missing transverse energy, an identified lepton, and several hadronic jets;

The reconstruction methods used in this analysis can be grouped into 3 operations. Inner Detector tracking comprises both track finding and fitting. The former is the process of associating together multiple hits likely due to a common charged particle. The latter is the determination of the charged particle's trajectory - its momentum vector and impact parameters - from these Inner Detector hits that have been associated together. Primary vertex reconstruction is the association of multiple tracks that intersect at a common point (to within error) in order to measure the position of a proton-proton interaction. Lastly, information in both the ID and the MS is used to identify muons, and, in some cases, combined to performed a momentum measurement that refines on that performed purely using the ID[10].

3.6.1 Track Finding and Fitting

A space point refers to either a pixel hit or a pair of hits on opposing sides of a single SCT module. The first step of track finding [45] is the association of three silicon space points into a *seed*. The reason a seed is composed from three points rather than two or four is that a helix can be determined by three points, and to a good approximation the path of a charged particle in the Inner Detector is a helix.

The track parameters that characterize the path of the particle in the bending plane - p_T, d_0, and ϕ - are determined by projecting the seed points and the helix along the direction of the helical axis; the result is a circle, illustrated in Fig. 3.10. The remaining two helical parameters, η and z_0, are estimated by averaging the η position of the three space points and then finding the intersection of a line with the given η with the nominal interaction point.

After the track parameters are estimated, seeds having an impact parameter above a maximum value or a momentum below a minimum value are discarded. These values are configurable and depend on the merits of the physics and the time that can be allocated, given limited computer resources, to reconstruction. Additional hits are then searched for within a *road* - a narrow three dimensional volume corresponding to the path of the putative particle. If more hits are found, the track parameter estimates, and hence the road, is refined using a Kalman filter.

At this point, tracks that do not have a minimum *track score* are discarded. The track score takes into account the number of holes - expected hits that are absent - and the track χ^2. Track ambiguities, such as multiple tracks with similar hit contents, are merged.

Many track fitting methods have been implemented in the ATLAS reconstruction framework. The current default method, and the only one used used for extraction of the $\Upsilon(1S)$ cross section, is a global χ^2 algorithm [46]. This chooses the track parameters by minimizing the χ^2 function, defined as

---

objects which themselves have been reconstructed from data using almost every sub-detector in the experiment.

[10] This refined measurement is never used in this analysis.

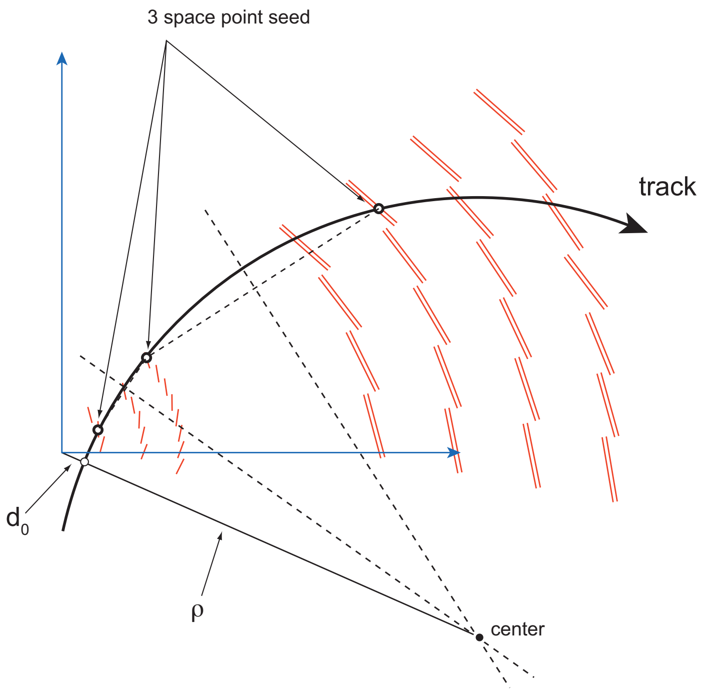

Figure 3.10: Illustration of initial track parameter estimate in the bending plane from three seed spacepoints. Three points in the plane determine a circle, which can be specified by its radius and the circle center (two coordinates). The track parameters p_T, ϕ and d_0 can then be calculated using basic algebra.

$$\chi^2 = \sum_i \frac{r_i^2}{\sigma_i} \tag{3.1}$$

where the sum runs over all space points. The residual of the i^{th} measurement, r_i, is the distance between the actual space point and the point at which the track intersects the plane parallel to the wafer and located at mid sensor depth. The error on the residual, σ_i, is assigned on a hit-to-hit basis but the approximate average values are the same as the instrumentation resolutions listed above.

3.6.2 Primary Vertex Reconstruction

The first step in the reconstruction of primary vertices is the identification of vertex seeds. This is done by looking for clusters of tracks in the distribution of longitudinal impact parameter, z_0. (This quantity refers to z_0 with respect to the nominal coordinate origin of the detector, not with respect to the primary vertex, which is as yet unknown.) Fig. 3.11 shows an example distribution in simulated events in which a Higgs boson is produced and decays to two photons. The red arrow indicates the position of the proton-proton interaction at which the Higgs particle is produced. Other proton-proton interactions are visible in the several other peaks.

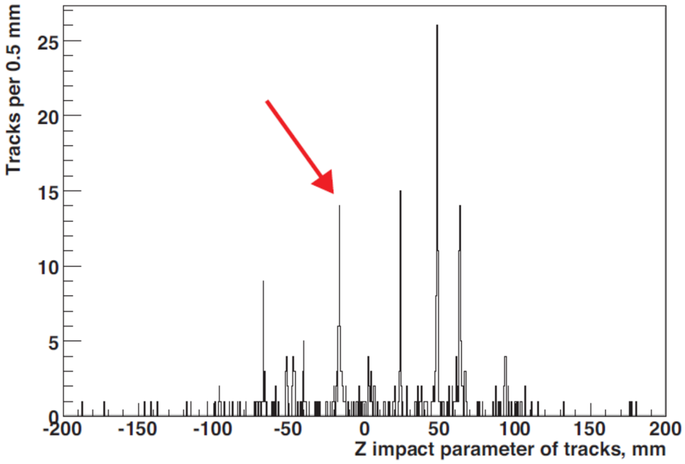

Figure 3.11: Histogram illustrating the identification of primary vertex seeds. Multiple primary vertices are evident in the several peaks. From [47].

Additional tracks are then assigned to the seed using a 'down-weighting deterministic annealing' scheme [48]. What this means is that an auxiliary 'temperature' parameter is introduced; the track-vertex χ^2 and the temperature determine a new variable, the track-vertex weight. In early, high temperature, iterations, only tracks with low χ^2 can result in low weights and be added to the seed. In later iterations with lower temperature, all tracks within a cut-off of $\chi^2 < 49$ are attached. This process is then repeated with tracks not previously associated until no additional vertex can be found.

The primary vertex resolution can be evaluated in data [49]. The idea is to split a single vertex into two sets of tracks and reconstruct the two sets as independent vertices. The assignment must be random and care must be taken not to introduce inputs that would correlate the measurements e.g. a beamspot constraint. The distribution of the differences in the two fitted positions can then be related to the vertex resolution. For events used in this analysis, $\sqrt{\sum_{tracks} p_T^2}$ is greater than or equal to $\sqrt{32}$ GeV; events with this level activity have a vertex resolution in the transverse plane of approximately 45 μm.

3.6.3 Muon Reconstruction

Muon reconstruction [50] proceeds in three steps. The first step is track finding and fitting based only on information from the Muon System. This is complicated by the fact that there are large distances (2-6 meters) between the station layers. Because of this, individual tube hits in a single station are first grouped together in straight line segments. Segments are then associated together to form tracks. Inner barrel stations as well as most of the inner end-cap stations (the CSC's excluded) do not make measurements along ϕ. The segment association thus begins in either the outer or middle stations. Once a track candidate has been found, a full refit is performed using the drift time information in each individual tube hit.

Next, tracks reconstructed from the Muon System are paired to and combined with Inner Detector tracks. Given a track from the Muon System, a crude ID track selection is performed based on the direction (η and ϕ) of the muon track. The two sets of track parameters, P, are then combined by *averaging* them, weighted by their error matrices. Defining the inverse of a track covariance matrix C by $\mathcal{C}$ for notational simplicity, the averaging equation is simply

$$\mathcal{C}_{combined} P_{combined} = \mathcal{C}_{ID} P_{ID} + \mathcal{C}_{MS} P_{MS} \tag{3.2}$$

$$\mathcal{C}_{combined} \equiv \mathcal{C}_{ID} + \mathcal{C}_{MS} \tag{3.3}$$

For the average to make sense, the parameterization of both tracks must be identical i.e. P_{MS} describes the muon track after extrapolating it back through the calorimeter to the interaction point.

In general, multiple ID tracks may pass the $\eta - \phi$ selection and thus a single MS track may result in many combinations. The best match is ascertained by evaluating the combination χ^2:

$$\chi^2_{combination} = (\Delta P)^T \mathcal{C}_{combination} (\Delta P) \tag{3.4}$$

Only the combination with the lowest $\chi^2_{combination}$ is retained. The resulting object is called a 'combined muon.'

The last step is dedicated to recovering muons with p_T too low to reach the middle and outer stations, see Fig. 3.12. ID tracks that extrapolate close to an inner station segment are associated with that segment to form a 'tagged muon'. As it is largely insensitive to the precise geometry of the toroidal field and energy loss in calorimeter, this reconstruction algorithm is much simpler than the methods described above. ID tracks and MS segments that have been previously associated to a combined muon are not candidates to become a tagged muon so by construction there is no overlap between these methods.

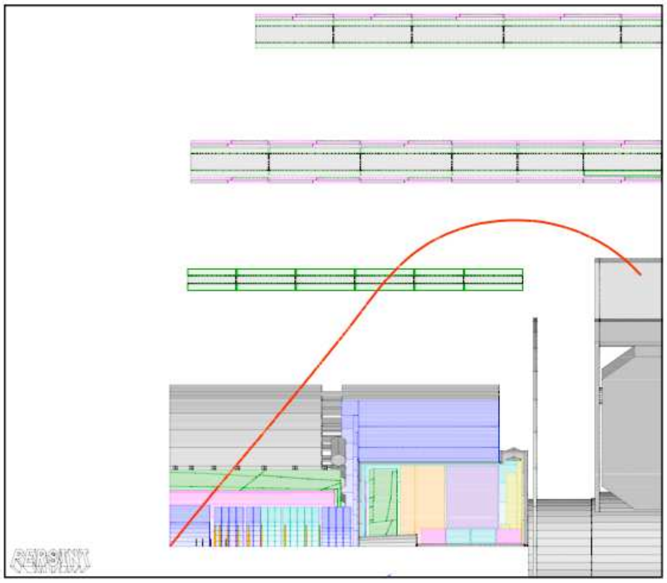

Figure 3.12: Event display (of a simulated event) illustrating a low p_T muon that can be recovered with the tagging algorithm. The view is in the R-z plane; the toroidal field is normal to the plane of the page. From [51]

3.7 Detector Simulation

This analysis is largely data driven. Simulated events are used in six cases, summarized below. In only one case does a result from 'first principles' simulation enter the analysis, and then only to ascertain a subdominant systematic. In the other five cases, the detailed predictions of simulation are largely factored out by supplementing the relevant quantities with the results from data. (The simulated events can be thought of as placeholders.)

All samples are produced by generating[11] events in PYTHIA 6 [52] or an ATLAS-specific modification thereof termed PYTHIA B [53]. The GEANT4 [54] toolkit is then used to model the interaction of stable particles with both active and inert detector material. The output of GEANT is essentially a list of energy depositions in simulated detector elements. Digitization is the final step and consists of translating this list into the ATLAS event data format. Simulated events are then reconstructed in precisely the same manner as the data. For a detailed discussion of the simulation infrastructure see [55].

Nota bene: because both event generation and the detector simulation make extensive use of randomly generated numbers, simulated events are often called Monte Carlo (MC) samples.

3.7.1 Summary of the Uses of MC

As the use of MC *can* be the most tenuous link in an analysis, it is worthwhile to collect in one place all the ways in which it is used.

(1.) Muon reconstruction efficiencies use maps derived from simulation. The actual efficiency that enters into the event weight is of the form $\mathrm{f}^{\mathrm{T\&P/MC}} \times \epsilon_{MC}$ where $\mathrm{f}^{\mathrm{T\&P/MC}}$ is a data driven scale factor determined using tag and probe. (Sec. 5.4)

(2.) Impact parameter efficiencies for the signal are determined by calculating the fraction of the resolution function less than the cut value. The resolution is determined from smeared MC, where the degree of smearing is a parameter in fits to data. (Sec. 5.5)

(3.) A cross check on the resolution function is determined from scaled or stretched MC i.e. a function $r^{\mathrm{scaled}}(d_0) \equiv r^{MC}(\mathrm{scale} \times d_0)$. The size of the scale is determined from the data/MC ratio of the high p_T resolution. (Appendix A)

(4.) Closure tests used to check the efficiency weight formalism use $\Upsilon(1S)$ MC. The results of this test are also used to assess the bin migration systematic uncertainty. This uncertainty is subdominant compared with both the statistical error and systematic uncertainties associated with the fits. (Sec. 5.7.1)

(5.) Signal PDFs used for the $N(\Upsilon(1S))$ extraction are interpolated histograms made from smearing $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ MC. The smearing parameters are determined from studying the momentum resolution in data. (Sec. 6.1)

(6.) Alternative background PDFs use $b\bar{b}$ and $c\bar{c}$ MC. The shape of these PDFs are extremely broad and slowly varying with dimuon mass; they depend only very weakly on the modeling of the detector resolution function. (Sec. 6.2)

These six cases are part of the extraction of the central values and the estimation of uncertainties; the analysis proper. MC is also used to validate the method used to measure the reconstruction and trigger efficiencies.

---

[11] In experimental HEP the word 'simulated' is customarily reserved for the simulation of detector response. The simulation of the collision process per se is called 'generation.'

Chapter 4

Definition and Strategy of the Measurement

A general problem in high energy physics concerns the treatment of events that are outside of the detector's range of sensitivity. In the current analysis, this motivates a very specific definition of the quantities that are being measured; these subjects are the topics of Secs. 4.1 and 4.2. Sec. 4.3 then surveys the 'macro issues' of the analysis and outlines the analysis strategy. The event selection is listed in Sec. 4.4.

4.1 Acceptance Corrections: A General Dilemma

Every detector has limitations in the phase space over which it can make efficient measurements of stable particles; particles at small angle to the beam, or possessing energy or momentum below some minimum threshold, will not be able to be reconstructed. Because of this it is sometimes not possible to measure a simply defined quantity without making large theoretical extrapolations. As an example consider the measurement of $dN_{ch}/d\eta$, the charged particle multiplicity as a function of pseudorapidity, with the ATLAS detector reported in [56]. The Inner Detector was *not* optimized for the high efficiency tracking of charged particles with transverse momentum below a few hundred of MeV. Yet the fraction of charged particles with p_T less than 500 MeV is assuredly not insignificant. How then to measure $dN_{ch}/d\eta$?

One possibility is to extrapolate the p_T spectrum below the minimum value at which the tracking efficiency is judged to be well understood. This has the advantage of preserving the simplicity of the final quantity - to wit, $dN_{ch}/d\eta$. The disadvantage in this approach, however, is that the final experimental result has been convoluted with a theoretical model - to wit, whatever set of assumptions were made in order to carry out the p_T extrapolation.

This dilemma is quite generally unavoidable. Although specialized (and computationally intensive) low p_T algorithms can indeed extend the sensitivity of the tracking down to 100 MeV, the Inner Detector, like any scientific instrument, has physical limitations to its dynamic range that cannot be circumvented with more powerful computing. In this case, the radius of the first layer of pixels and the strength of the solenoidal magnet field (50.5 millimeters and 2 Tesla, respectively) mean that particles with transverse momentum below 30 MeV do not leave even a single detector hit; they are completely outside of the instrument's range of sensitivity.

A less obvious approach, and the one that was taken in [56], is to 'change the question.' That is, instead of measuring $dN_{ch}/d\eta$, the somewhat more specific quantity $dN_{ch}^{p_T>500\ MeV}/d\eta$ is instead reported. This choice is representative of a general idea in modern experimental particle physics: to forgo a simple result with (possibly large) theoretical uncertainties in favor of a result of more limited meaning but independent of any model or method of extrapolation.

4.2 The Υ Polarization and Acceptance Uncertainties

An analogous issue, though perhaps more subtle, arises in the present cross section measurement. In this case the major limiting constraints are both the $p_T^\mu > 4$ GeV trigger threshold and the rapid variation of the muon reconstruction efficiency below, at least in the barrel region, the same value of 4 GeV. In light of these constraints, this analysis only uses reconstructed muons with $p_T^\mu > 4$ GeV and $|\eta^\mu| < 2.5$. (The η requirement has nothing to do with the Muon System and is determined solely by the ID coverage.)

Like the charged multiplicity example above, many $\Upsilon(1S) \rightarrow \mu^+\mu^-$ events will be outside of the detector acceptance. That is, many events will result in one or two daughter muons that do not satisfy the p_T^μ and η^μ requirements. In order to report the production cross section with no qualifications, the fraction of events failing these requirements would have to be accounted for. The crucial issue is that this fraction depends on the polarization.

Fig. 4.1 shows this fraction as a function of p_T^Υ and y^Υ for a variety of polarizations assumptions. For values of $p_T^\Upsilon \approx M(\Upsilon(1S))/2$ most two body decays will result in one of the decay products nearly at rest in the laboratory frame. As $M(\Upsilon(1S)) =$ 9.46 GeV, this effect is most prominent at 4-5 GeV. This is evident in the horizontal bands in all 5 polarization assumptions, which indicate a minimum acceptance as a function of p_T^Υ that is nearly independent of y^Υ. After this minimum the acceptance rises monotonically with p_T^Υ.

However it is the implied uncertainty on the acceptance, rather than the acceptance for a specific polarization, that is the most relevant feature of Fig. 4.1. The uncertainty can be quantified as the ratio of the maximum to minimum acceptance of the assumptions in Fig. 4.1 and is shown in Fig 4.2. (The ratio is also a function of p_T^Υ and y^Υ: for different values of $\Upsilon(1S)$ kinematics the maximum and minimum

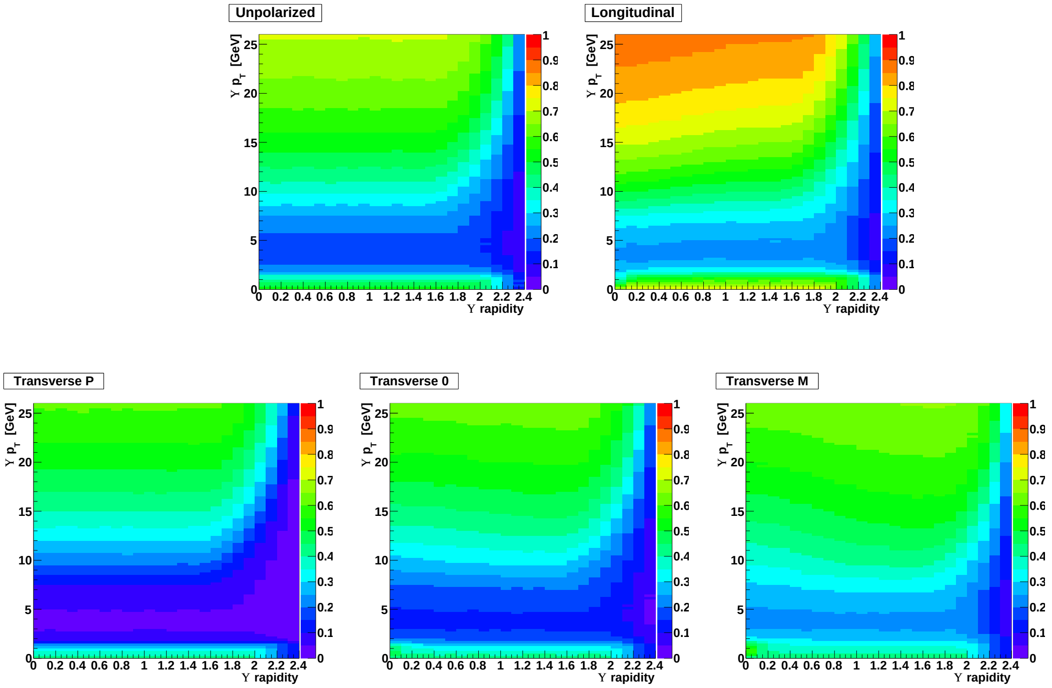

Figure 4.1: Acceptance of the muon selection cuts ($p_T^\mu > 4$ GeV, $|\eta^\mu| < 2.5$) for $\Upsilon(1S)$ mesons as a function of the p_T and y of the $\Upsilon(1S)$. Shown are the acceptances for no polarization (top left), longitudinal polarization (top right), and different transverse polarizations $\Upsilon(1S)$ mesons (bottom).

may correspond to different polarizations.) This uncertainty exceeds a factor of 4 below $p_T^{\Upsilon} = 8$ GeV and exceeds a factor of 1.5 over most of the displayed range. This is precisely the dilemma of large theoretical uncertainties alluded to in the above discussion.

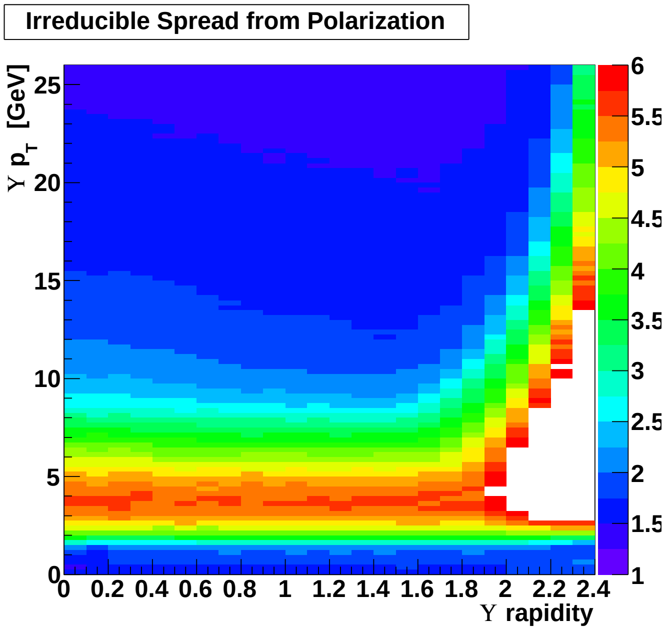

Figure 4.2: Uncertainty on the acceptance of the muon selection cuts for $\Upsilon(1S)$ mesons as a function of p_T^{Υ} and y^{Υ}.

Previous measurements of Υ production cross sections at hadron colliders [57, 58] have calculated the detector acceptance assuming no polarization or [59] have reported several cross sections based on the same data corresponding to multiple polarization hypotheses. It will not surprise the reader that the aim of this analysis will be to instead report the cross section only for $\Upsilon(1S) \to \mu^+\mu^-$ events producing two muons within the kinematic region that the detector is directly sensitive to ($p_T^{\mu} > 4$ GeV and $|\eta^{\mu}| < 2.5$).

Differential cross sections and event yields should *always* be understood to have this qualification. Properly, they would then be written as $\sigma(\Upsilon(1S)|_{p_T^{\mu}>4 \text{ GeV},\ |\eta^{\mu}|<2.5})$ and $N(\Upsilon(1S)|_{p_T^{\mu}>4 \text{ GeV},\ |\eta^{\mu}|<2.5})$ or something similar. However because in this work acceptance corrections are never made, they will just be written as $\sigma(\Upsilon(1S))$ and $N(\Upsilon(1S))$.

4.3 Overview of the Measurement Strategy

Having defined *what* is to be measured, this section will survey *how* the measurement is to be done.

The differential cross section is given by

$$\frac{d^2\sigma}{dp_T dy} = \frac{N(\Upsilon(1S))/\int \mathcal{L}\mathrm{d}t}{\Delta p_T \Delta y} \tag{4.1}$$

where Δp_T and Δy are the widths of a given bin in dimuon transverse momentum and rapidity. Determination of the luminosity is an 'infrastructure' analysis standardized across the experiment [60, 61]. Thus the sole focus will be the measurement of $N(\Upsilon(1S))$. The two major issues are accounting for missed events and separating out signal and background components.

In all but the most straightforward of analyses there are usually multiple viable strategies; in the present case a significant decision concerns the order in which the two issues noted above are addressed. In many cross section analyses the ordering is first to extract the number of reconstructed events due to the signal process and then to correct for detector inefficiencies. This method can be summarized as

$$N_{raw}^{sample} \xrightarrow{signal\ extraction} N_{raw}^{signal} \xrightarrow{\div\epsilon} N^{signal} \tag{4.2}$$

The signal extraction is often performed by fitting a distribution of one or more variables describing the event kinematics. The total event efficiency, ϵ, is in general the product of several factors. (Specifically, in this analysis the event efficiency is factorized into three terms: the trigger efficiency, the reconstruction efficiency, and the quality selection efficiency.)

While an $\Upsilon(1S)$ cross section measurement could very well be approached in this manner, the present analysis does not do so. Instead, the procedure is to first reweight the observed dimuon mass distribution by the inverse of the event efficiency. The corrected signal yield is then extracted by performing a fit of the weighted mass spectrum. This method can be summarized as

$$N_{raw}^{\mu+\mu-} \xrightarrow{\div\epsilon} N_{corrected}^{\mu+\mu-} \xrightarrow{mass\ fit} N(\Upsilon(1S)) \tag{4.3}$$

which, with more specific notation, is just the reverse of the procedure indicated in Eq. 4.2.

The second ordering is arguably conceptually simpler than the first. In the first ordering, the quantity resulting from the signal extraction could be termed $N_{raw}(\Upsilon(1S))$. The appropriate efficiency would then be an average over the square of the muon efficiency map i.e. $\epsilon_\Upsilon = \overline{\epsilon_\mu^2}$. But ϵ_μ varies strongly depending on the momentum of the muon while a given dimuon mass distribution includes events spanning a large range of muon momentum. Another way of saying this is that one needs to know how to do the average over $\vec{p}_\mu$.

A common method of doing so uses Monte Carlo that has been reweighted such that the relevant kinematic distributions (and, if need be, the correlations) are the same as those observed in data. For example in a two body decay such as $\Upsilon(1S) \to \mu^+\mu^-$ the relevant variables are p_T^Υ, η^Υ, $\theta^\star$ and $\phi^\star$. ($\theta^\star$ and $\phi^\star$ are the polar and azimuthal angle of the positive muon momentum vector in the production frame, a coordinate system in which the Υ is at rest.)

Fig. 4.3 illustrates[1] the method. First, a fit to the raw data is performed, resulting in 143 for $N_{raw}(\Upsilon(1S))$. Then using Monte Carlo, ϵ_Υ^{-1} is determined to be 6447/5157. The product of the two gives a signal yield of 178.

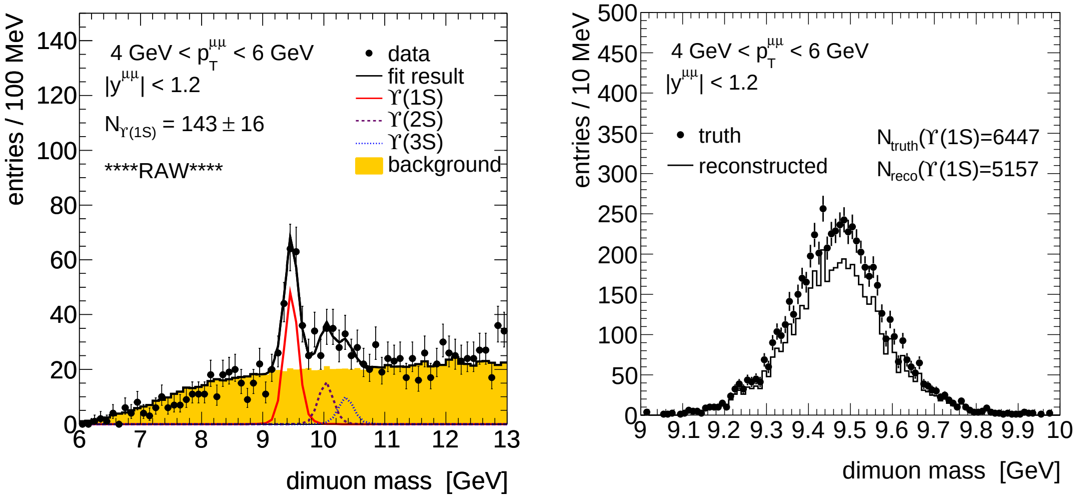

Figure 4.3: Illustration of the method by which the signal is first separated from background and then an efficiency correction deduced from MC is applied. On the left is a fit to a raw dimuon mass spectrum; the signal yield is 143. On the right is a plot of pure signal MC - notice the absence of any background - in which 5147 out of 6447 events have been reconstructed. The product of the raw yield and the inverse MC efficiency determines $N(\Upsilon(1S))$.

While this method has been successfully used to perform many measurements, it is somewhat more complicated than applying the efficiency correction before the signal extraction. In this case, the total event efficiency is schematically given by the unique expression $\epsilon_+(\overrightarrow{p_+})\epsilon_-(\overrightarrow{p_-})$; the kinematics of the data *uniquely* determine the efficiency. No average is performed: fitting the reweighted distribution gives $N(\Upsilon(1S))$ at once. This is shown in Fig. 4.4, which corresponds to the same bin as that in Fig. 4.3.

The only drawback to this approach is a purely technical complication: when fitting a weighted distribution, the calculation of the covariance matrix must be generalized from the unweighted case to ensure that the statistical error scales with the

---

[1] The figure is intended as an illustration only; no reweighting has been applied to the Monte Carlo.

actual statistics of the sample, rather than the sum of the weights.

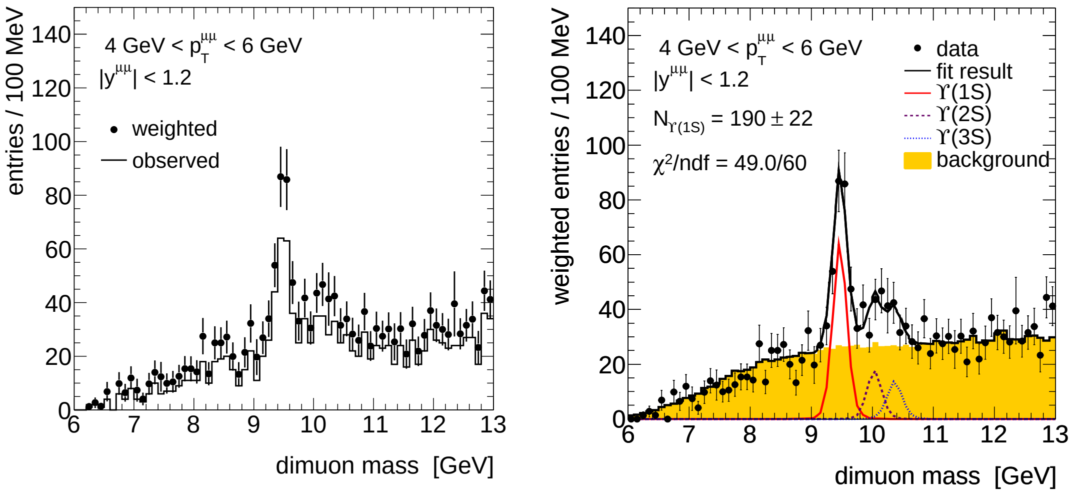

Figure 4.4: Illustration of the method by which the signal is separated from background after correcting for inefficiencies on an event-by-event basis, the method that is actually used. On the left is a dimuon mass spectrum before and after the application of the efficiency weights. On the right is a fit to the weighted spectrum.

Ultimately, however, the differences between these two methods are small. In Fig. 4.5 the results of two cross checks in which the detector inefficiencies are accounted for after the fits are shown. In the first cross check, the event-by-event weights are still determined from the kinematics observed in data, but they are not applied before the fit. Instead, the average is taken over events inside the range $|M(\mu^+\mu^-) - M(\Upsilon(1S))| < 600$ MeV. This average weight is then applied to the raw yield after the fit. The second cross check is a poor man's version of the Monte Carlo method discussed above. In this case the weight is the fraction $N_{truth}(\Upsilon(1S))/N_{reconstructed}(\Upsilon(1S))$ observed in $\Upsilon(1S)$ Monte Carlo. Done properly, this method would also have reweighted the Monte Carlo sample itself both in order to account for differences between observed and simulated detector efficiencies and between the observed and generated $\Upsilon(1S)$ and muon spectra. Neither of these steps have been taken and for that reason it should be viewed merely as a cross check.

The highest p_T^{Υ} bin in the central rapidity region is seen to be higher in both cross checks by about 8%, which is less than $\delta_{statistical}/3$ for this bin. In the other bins the agreement is at approximately the 4% level.

Given the statistical precision that is possible with this dataset, the conclusion is that either choice in the 'ordering' of the analysis is equally viable.

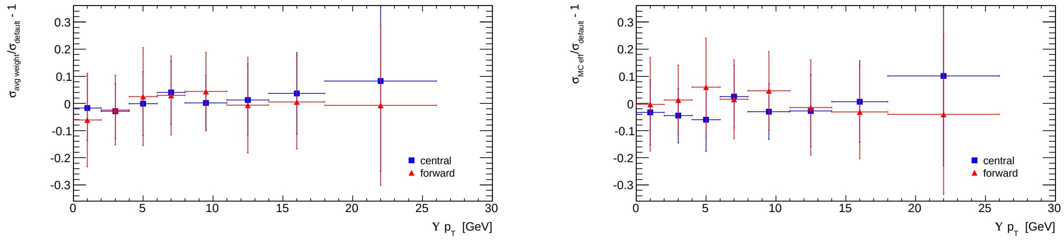

Figure 4.5: On the left is the result of applying the average weight determined from data to the results of the raw fit. On the right is the result of using $\Upsilon(1S)$ Monte Carlo to determine the weight. In both cases the quantity plotted is $\sigma_{cross\ check}/\sigma_{default} - 1$. Error bars are the statistical errors from the default method.

4.4 Event Selection and Motivation

To a first approximation, the event selection is simply the requirement of two muons of opposite electric charge (OS, for opposite sign). At a greater level of detail, the selection has a number of features, some of which will be briefly discussed after listing the precise requirements.

The event selection:

- EF_mu4: the event filter criteria of 1 muon with a p_T greater than 4 GeV
- a primary vertex with $\geq$3 tracks

The muon selection:

- #pixel hits $\geq$1, #SCT hits $\geq$6
- $p_T > 4$ GeV, $|\eta| < 2.5$
- $|d_0| > 150\ \mu$m, $|z_0| > 1.5$ mm

The muon pair selection:

- OS
- at least one must satisfy $\Delta R^{lvl1\ RoI} < 0.4$ and $\Delta R^{EF\_mu4\ feature} < 0.3$
- at least one must be combined

The pixel and SCT requirements are standard track quality cuts in ATLAS. The p_T and η cuts are motivated by both the trigger threshold and ID acceptance, as discussed above, while the impact parameter cuts are for the purpose of background suppression, see Sec. 5.5. Both the momentum vector and the impact parameter are those resulting from the ID track; track kinematics resulting from fits using information from the Muon System are never directly used.

Finally, the requirement that one muon must be combined deserves explanation: conceivably the requirement could have been loosened to include events with two tagged muons. However as the entire point of the tagging algorithm is to reconstruct a muon that does not propagate to the trigger stations, events triggered by the muon system but containing only tagged muons are somewhat pathological; hence the ‘at least one combined’ requirement.

Chapter 5

Efficiencies and Efficiency Weights

The aim of the present analysis is to count $\Upsilon(1S) \rightarrow \mu^+\mu^-$ events. A crucial prerequisite to doing so is to understand how to count muons i.e. to understand how well the detector apparatus and the trigger and reconstruction algorithms record events with muons and measure muon kinematics.

In particular, the three primary operations in the event and object selection can be identified: triggering, muon identification, and impact parameter selection i.e. the efficiency for a prompt muon to have reconstructed $d_0 < 150\ \mu m$. This list is not exhaustive: the event selection also requires the presence of a primary vertex and the measurement of the muon impact parameter also implicitely requires the identification of the track. However, these three operations constitute the largest source of inefficiencies. Furthermore, the associated studies needed to determine the efficiency of these operations are the most non-trivial. Because of this these three issues will be the focus.

5.1 The Basic Idea of Tag and Probe

Muon trigger and reconstruction efficiencies can be directly deduced from data using events selected on the mass of a dimuon resonance. The basic idea - which is not restricted specifically to muon performance - is to perform two event selections in which one is strictly tighter (more stringent) than the other. By comparing the number of events in the resonance peak satisfying the tight selection with the number of events satisfying the loose selection, the efficiency for a true muon to pass the additional cuts (requirements) can be ascertained. The dimuon resonance, in other words, serves as ‘a muon gun’; a source of muons that can be identified using kinematic information, i.e. mass spectra, rather than the Muon System.

The above discussion is somewhat general. In the case of reconstruction efficiencies, the loose selection is one muon identified in the MS and one ID track. The tight selection is two muons identified in the MS. The latter goes like the square of the reconstruction efficiency; the dimuon selection, roughly speaking, uses the Muon System and muon reconstruction twice. The yield from the muon-track selection uses these tools only once; these two numbers can thus be related to the efficiency. This idea is sketched in Fig. 5.1.

In the case of the trigger efficiencies, both selections require two reconstructed muons, but they differ in how many must be associated with a trigger object. The loose selection requires only one trigger match while the tight selection requires both be trigger matched.

For both studies, the resonance used is the J/ψ. Use of the $\Upsilon(1S)$ would result in less statistical power due to the lower cross section and considerable complications due to the $\Upsilon(2S)$ and $\Upsilon(3S)$, which can only be partially resolved from the $\Upsilon(1S)$. The J/ψ is both copiously produced and well separated from any other dimuon resonance.

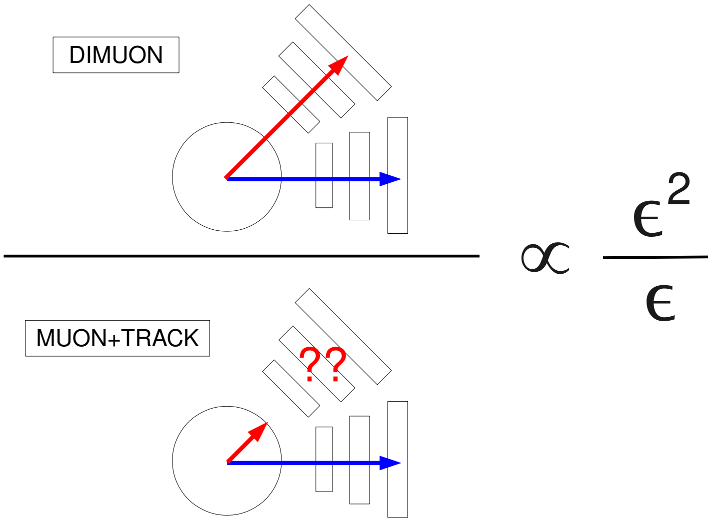

Figure 5.1: Sketch illustrating the use of tag and probe to measure the muon reconstruction efficiency. The number of events in the muon-track peak is the denominator. The number of events in the dimuon peak is the numerator.

The precise procedure, to be carefully defined below, is *slightly* different than that sketched above in order to simplify the treatment of errors. The above procedure would yield samples that are highly correlated for high efficiencies. The actual procedure calculates the efficiency using yields from mutually exclusive samples. The 'pass' or 'matched' sample satisfies the tight requirements while the 'fail' or 'unmatched' sample satisfies the loose requirements and fails the tight requirement.

$$\epsilon = \frac{PASS}{PASS + FAIL} \tag{5.1}$$

In either case the method just described is referred to as 'tag and probe'. The tag refers to the muon passing the tighter requirement; the reconstructed muon or the muon associated with the trigger object. The track that is not necessarily identified in the MS, or the reconstructed muon that is not necessarily associated with a trigger object is the probe; it is unbiased with respect to the operation under study.

5.2 Muon Reconstruction Efficiency

5.2.1 Muon Reconstruction Efficiencies from Simulation

Before turning to the data driven method the prediction from simulation is shown. This will give some idea of what should be expected and will motivate the binning as well.

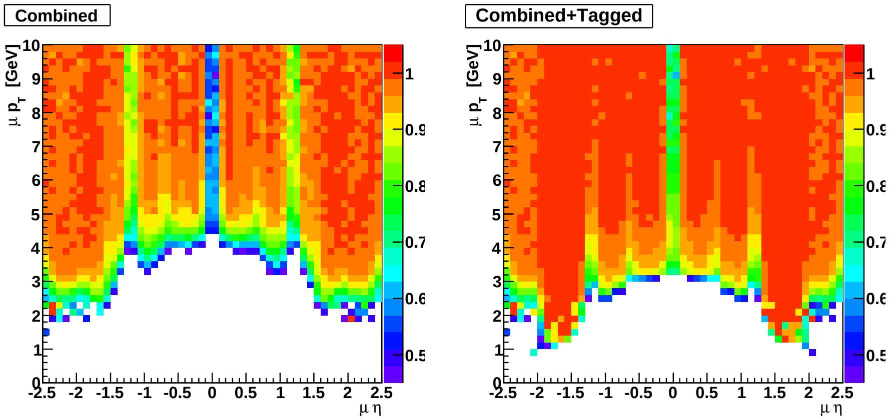

Figure 5.2: The muon reconstruction efficiency as a function of p_T and η predicted by simulation for the combined (left) and combined or tagged (right) muons.

Fig. 5.2 shows the reconstruction efficiency for combined muons and combined or tagged muons as a function of p_T and η as predicted by simulation. The most prominent features are the areas of lower efficiency at fixed η i.e. the vertical stripes; these result from the crack in central η in the large sectors and the barrel / end-cap transition region. This motivates the η binning:

1. $|\eta| < 0.1$: the central crack; for some values of ϕ there are gaps in the MDT and RPC coverage due to the services

2. $0.1 < |\eta| < 1.1$: the barrel; the detector in this region is nearly hermetic

3. $1.1 < |\eta| < 1.3$: the transition region between the barrel and the end-cap; both the magnetic field and the instrumentation in this η range are very non-uniform

4. $1.3 < |\eta| < 2.5$: the end-caps; like the barrel this is nearly hermetic

A tag and probe study based on a muon+track selection cannot provide the absolute reconstruction efficiency. Rather, it provides the conditional efficiency to reconstruct a muon in the muon system *given that it is reconstructed in the Inner Detector.* As the Inner Detector efficiency exceeds 99% the difference between these numbers is very small. However to be consistent this is accounted for in the production of the simulation maps.

A second notable feature of the simulation maps shown above is the method of associating reconstructed muons to generator level muons. This method does not make use of either 3-vector. Specifically, there is no ΔR cut[1]. Instead the association is done using a truth probability algorithm. These take as input the truth particle(s) associated to simulation hits. The significance of this is that because neither the truth nor reconstructed kinematics enter into the association decision, there is no dependence on the tails of the detector resolution function in the maps.

5.2.2 Event Selection

The event selection closely follows the cross section analysis event selection specified in Sec. 4.4 including both the ID quality and impact parameter criteria. The only requirement on the properties of the pair is that the muon and track must be opposite sign (OS). In particular there is no 'combinatoric' cut; all distinct muon-track pairs in a given event satisfying the above requirements are used in the remaining steps of the procedure.

A given muon-track pair probes the muon reconstruction efficiency according to the 3-vector of the track. The final maps are binned in this variable and thus in this case it makes sense to relax the track p_T cut down to whatever minimum range the maps are to be produced with (3 GeV and 2 GeV for combined and tagged muons, respectively).

The tag must be trigger matched. The is significant and deserves special mention. To see that some requirement along these lines is necessary consider muon-track pairs that fail this requirement. That is, consider pairs in which there is no trigger object near the muon. For simplicity of argument exclude combinatoric background i.e. exclude events in which the track is not a true muon. How were these events triggered? Ignoring fake triggers (again, for simplicity) the conclusion is that the trigger hits must belong to *the track.* But this at once spoils the central idea of having an event yield in which the muon system has been used only once. To be precise, the use of such a sample would give reconstruction efficiencies with respect to an Inner Detector track *and a trigger object.* As these are not the quantities of interest, these events are vetoed. The tag must also be combined and have $p_T > 4$

---

[1] The ΔR between two vectors is defined as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$

GeV, to avoid the same pathologies associated with tagged muons discussed at the end of Sec. 4.4.

5.2.3 Classifying the Tag and Probe Pair

The track is then labeled 'matched' or 'unmatched' according to whether or not it is the Inner Detector component of a second muon in the event. A potentially confusing point is that while the muon in the muon-track pair is always required to be combined, this in no way precludes the study of the tagged muon reconstruction efficiency; the definition of 'matched' is simply adjusted accordingly.

In order to determine the efficiency as a function of p_T and η, the muon-track pair is also sorted according to the kinematics of the track. There are thus $2N_{p_T bins}N_{\eta bins}$ classifications of tag and probe pairs. As each pair has a unique invariant mass there are consequently $2N_{p_T bins}N_{\eta bins}$ mass distributions.

5.2.4 Background Subtraction

Fig. 5.3 shows an example of the two mass distributions corresponding to one bin in p_T and η. This example was chosen because it is fairly unremarkable; the corresponding efficiency is neither exceedingly high nor low. While the peak in the matched distribution is very clean, the unmatched distribution is background dominated. Two related issues then naturally arise: how is the peak yield to be quantitatively defined? And how robust is the yield definition?

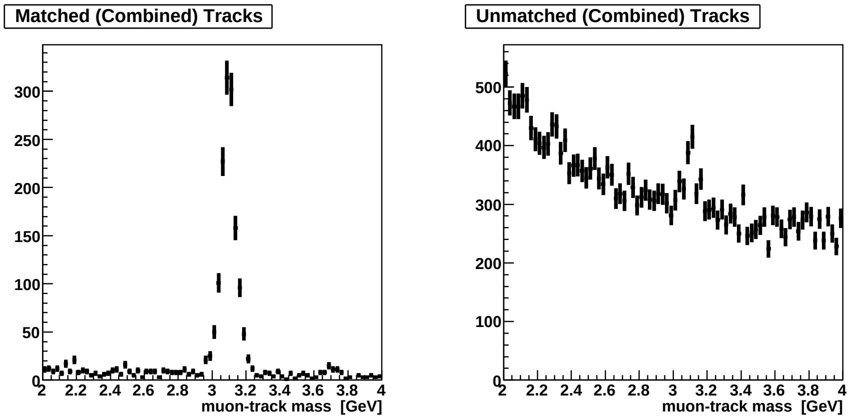

Figure 5.3: Muon-track invariant mass distribution for tracks in the barrel ($0.1 < \eta < 1.1$) with 4 GeV $< p_T <$ 5 GeV. Pairs in which the track is matched to a combined muon are shown at left. The right shows the mass of pairs in which the track is unmatched.

Three methods, were explored:

- linear sideband subtraction: the background yield from 3.0-3.2 GeV is estimated by the averaging the yields between 2.7-2.9 GeV and 3.3-3.5 GeV.

- quadratic sideband subtraction: the yield in the range 2.4-2.6 GeV is assigned an ‘x coordinate’ of 2.5 GeV. Likewise the yields between 2.7-2.9 GeV and 3.3-3.5 GeV are assigned to the average of the bounds of their respective integrals. Having constructed three points in the plane, there is now a unique parabola that intersects all three. The background yield from 3.0-3.2 GeV is then estimated by the value of this parabola at 3.1 GeV. In the limit that the background rate as a function of mass is linear, this method reproduces the result of the linear side band subtraction method. The basic idea, as well as the decision to use two low sidebands - as supposed to two high sidebands and one low sideband - is sketched in Fig. 5.4.

- quadratic fit: a 6 parameter fit is made to the full mass distribution. Three of the parameters describe a Gaussian that is used to model the ‘signal’ i.e. the J/ψ peak. The other three parameters describe a second degree polynomial. In this method there is no explicit background subtraction: the peak yield is given immediately by the fit results in the form of the integral of the Gaussian.

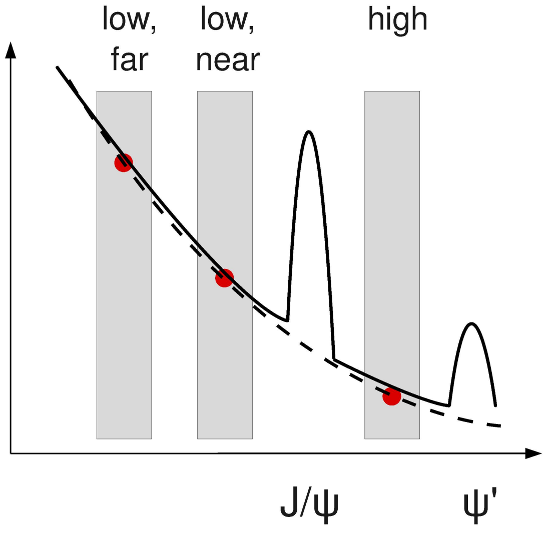

Figure 5.4: Sketch illustrating the method of quadratic sideband subtraction. The background model is a parabola determined by the yield in the three sideband regions. These are chosen in such a way as to avoid the ψ'.

5.2.5 Validation of the Method

There are two non-obvious questions related to the tag & probe procedure, irrespective of the efficacy of the background subtraction procedures. The first is in some sense combinatoric: given a $J/\psi \to \mu^+\mu^-$ decay in which both muons are reconstructed, there will be two muon-track pairs classified as matched while in the case that one is reconstructed and one is not there will be a single pair classified as unmatched. (Of course, in the case that neither muon is reconstructed there will be zero pairs in either category.) Is it in fact true that the matched yield divided by the sum of matched and unmatched gives the efficiency? Put differently: is there no double counting implicit in the proposed procedure? Defining N_2 (N_1) as the number of J/ψ events in which both (one) muons are reconstructed, the question[2] can be written as

$$\frac{2N_2}{2N_2 + N_1} \stackrel{?}{=} \epsilon \tag{5.2}$$

which is seen to be a valid identity upon noting that $N_2 = \epsilon^2$ and $N_1 = \epsilon(1-\epsilon) + (1-\epsilon)\epsilon$.

The second question concerns the RoI and EF feature requirement: is it sufficient to provide a sample of probes that are unbiased with respect to the Muon System?

This issue, (and, implicitly, the combinatoric question as well) are addressed by performing the method using a J/ψ Monte Carlo sample. The results are compared to the efficiency in simulation determined using truth information and found to be in good agreement, at least for muons with p_T above 4 GeV where the efficiency is not rapidly varying. Fig. 5.5 shows the test results for both the combined efficiency and the combined + tagged efficiency in both the barrel and end-cap regions.

Shown also are the results of performing the method on an $\Upsilon(1S)$ sample. This further demonstrates the basic soundness of the method by showing the results to be independent of the mass of the dimuon resonance.

5.2.6 Reconstruction Efficiency Results

The results from the quadratic sideband subtraction (which should be thought of as default) and the other two methods are shown in Figs. 5.6- 5.9. Shown are the measured efficiencies for both combined muons and 'inclusive tagged' muons i.e. the logical or of combined and tagged muons. In the barrel and end-caps the combined efficiency plateaus at or above 95%, while in the central crack and transition region

---

[2] The question has been phrased in such a way as to lump combined muons and tagged muons together as 'reconstructed muons.' However since the two objects are treated differently during the muon-track pair selection, the question as stated is really just a 'warm-up' version of the full issue. Properly generalized the classification would not be in terms of N_2, N_1 and N_0 but rather N_{CB-CB}, N_{CB-TG}, $N_{CB-missed}$, N_{TG-TG}, $N_{TG-missed}$ and $N_{missed-missed}$. But this conveys the basic idea.

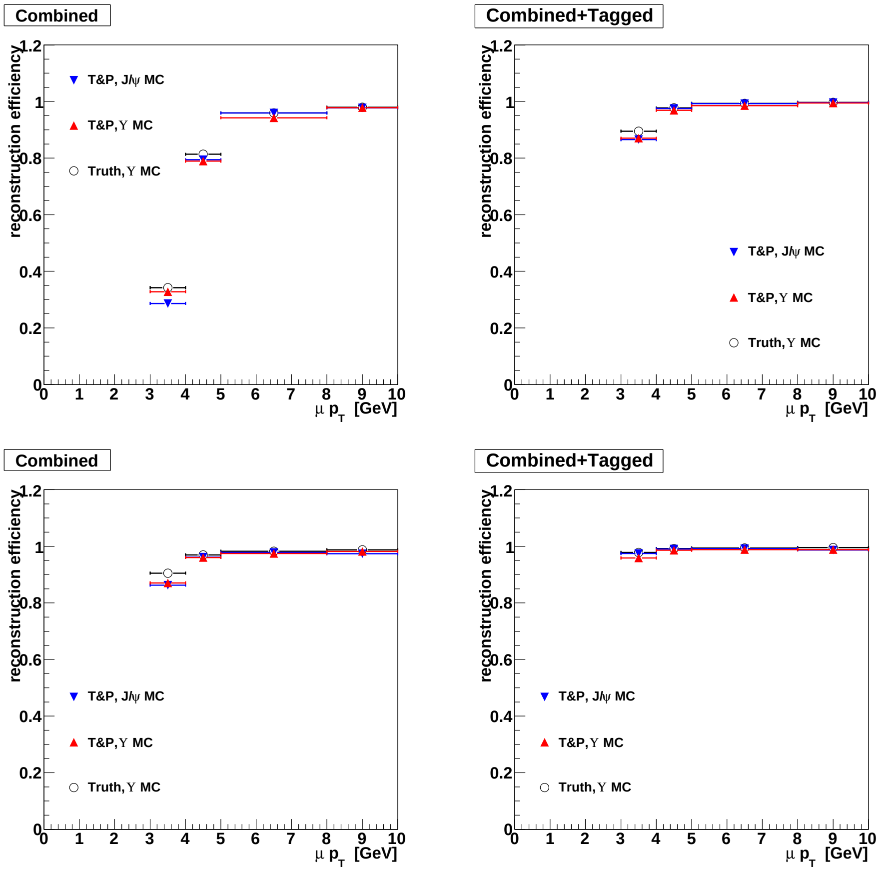

Figure 5.5: The top panels show the validation results for the barrel region; the bottom for the end-cap region.

the plateau is about 60% and 85%. In all regions but the end-caps, the use of tagged muons is seen to make a significant difference between 3 and 5 GeV e.g. in the 4-5 GeV barrel bin the use of tagged muons results in an efficiency approximately 15% higher. Above 4 GeV - the range relevant for the cross section analysis - the inclusive tagged efficiency is above 95% in the barrel and end-caps.

The prediction from simulation (using truth information) is also shown. Although the efficiency predictions from simulation are never used in the cross section analysis, it can be observed that the muon reconstruction efficiency is well modeled. A possible exception to this statement is the combined efficiency in the barrel which trends low by several percent. (However for each single bin, the discrepancy is only at approximately the 1σ level.)

The three background subtraction methods also largely agree within errors for each individual bin in p_T and η. This demonstrates that the yield definition is relatively robust. However the linear sideband subtraction consistently gives results a few percent higher than the quadratic sideband subtraction method. This implies that the linear sideband subtraction method overestimates the amount of background relative to the quadratic method, which in turn implies that the quadratic term in the parabola describing the background distribution enters with a positive coefficient i.e. that the mass distribution is ‘concave up.’ These observations agree with the example mass distribution shown in Fig. 5.3.

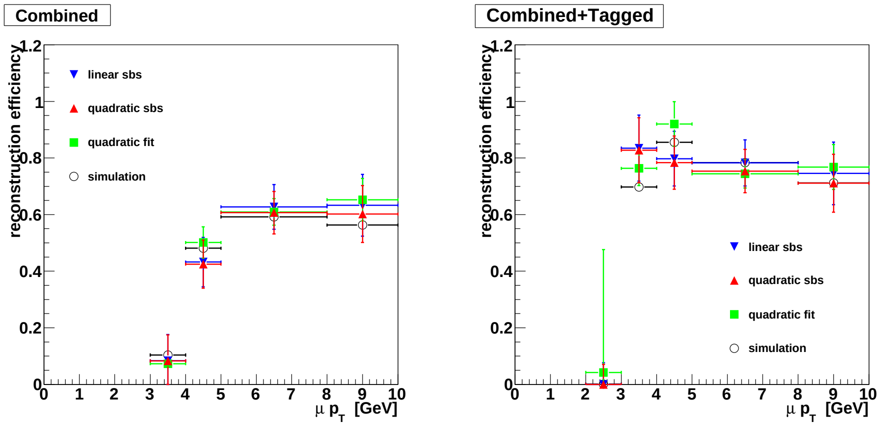

Figure 5.6: Shown are the measured muon reconstruction efficiencies for the central crack region ($|\eta| < 0.1$). The degraded efficiency due to the limited coverage is especially evident in the combined efficiency. The small angular extent of this region reults in limited statistics that are evident in the large error bars.

The results are summarized in Fig. 5.10, which shows the difference between the efficiencies observed in data and those predicted from simulation. The significance, the difference divided by the statistical error, is also shown. The mean of the significance

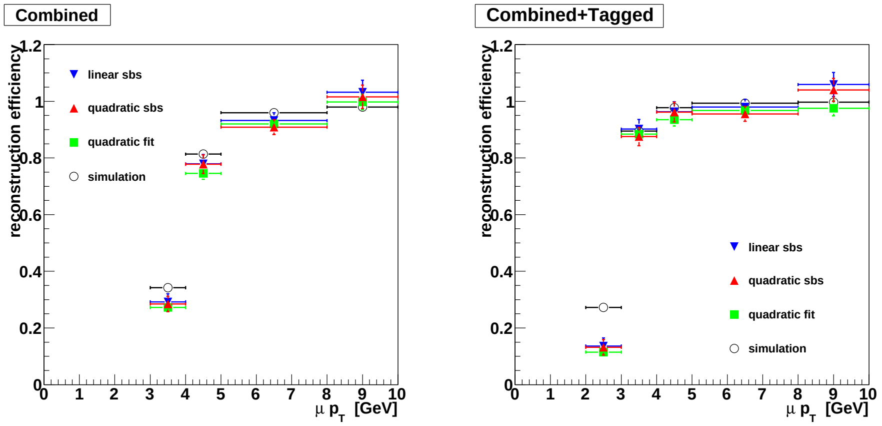

Figure 5.7: Shown are the measured muon reconstruction efficiencies for the barrel ($0.1 < |\eta| < 1.1$). The increase in efficiency afforded by tagged muons is especially evident in this region.

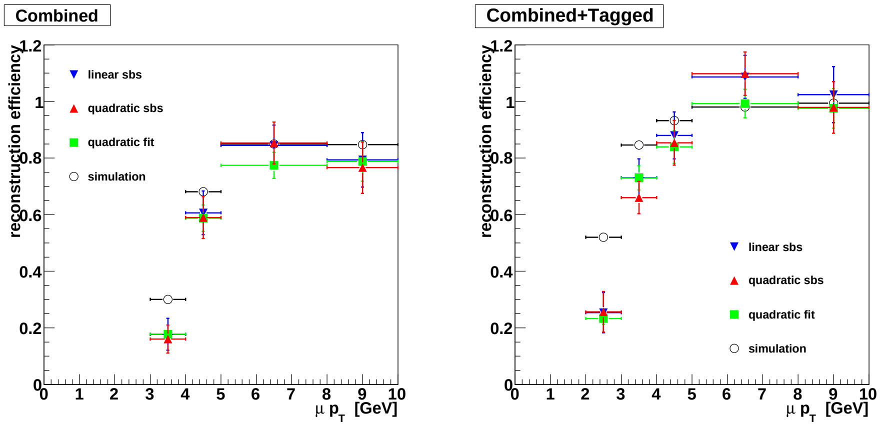

Figure 5.8: Shown are the measured muon reconstruction efficiencies for the barrel/end-cap transition region ($1.1 < |\eta| < 1.3$). Like the central crack, this region has both lower absolute efficiency than the barrel and limited statistics.

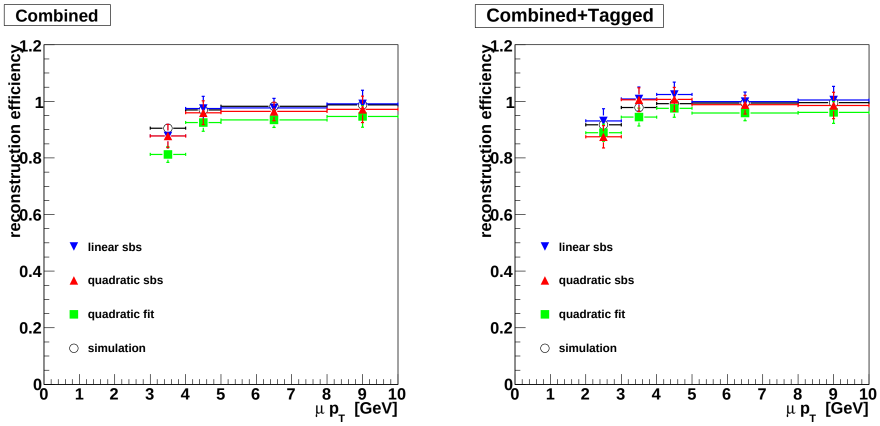

Figure 5.9: Shown are the measured muon reconstruction efficiencies for the end-caps ($1.3 < |\eta| < 2.5$). In this region, the efficiency is excellent above 4 GeV.

distribution for the combined efficiency in the barrel and end-cap is -0.74 ± 0.33 while the same quantity for the combined or tagged efficiency is -0.22 ± 0.22.

This concludes the measurement of the muon reconstruction efficiency. The next section turns to the determination of trigger efficiencies. This study will also makes use of the idea of tag and probe.

5.3 Trigger Efficiency

The most obvious difference between measuring reconstruction efficiencies and trigger efficiencies involves redefinition of the matched/unmatched criterion from ‘an Inner Detector track that is identified as a muon’ to ‘a combined (tagged) muon that is associated with a trigger object.’

Closer inspection reveals two additional differences between the trigger tag and probe study relative to the situation in the reconstruction study; one simplification and one complication. The simplification is the near absence of background. In the reconstruction study, the intermediate objects are muon-track invariant mass distributions. In the present case, the probe is always required to be an identified muon, even in the unmatched sample. Ignoring fakes, the amount of background in the trigger study is thus related to the rate of $pp \to \mu^+\mu^- + X$ events, rather than the rate of events with a single muon in it as is the case for the reconstruction study. This is clearly illustrated by contrasting the unmatched distribution in Fig. 5.3 with the analogous distribution in Fig. 5.11.

The complication stems from the poor angular resolution of the trigger chambers. The tail of lvl1 RoI resolution function is such that in $J/\psi \to \mu^+\mu^-$ events with a narrow opening angle between the muons, there is a potential ambiguity in the

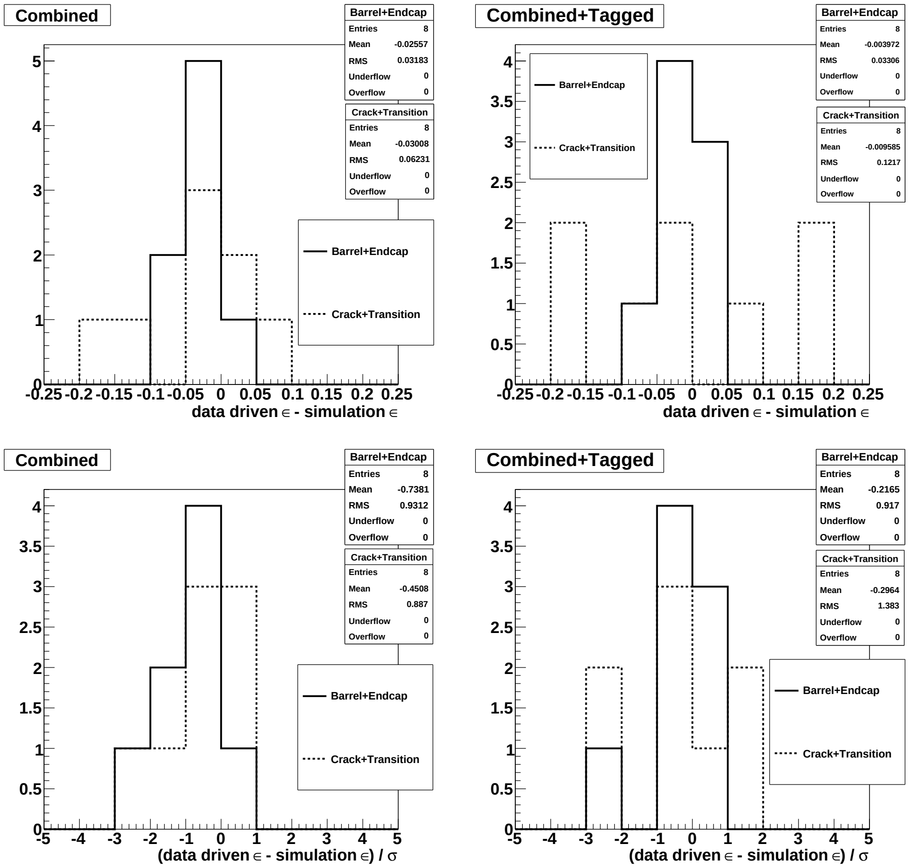

Figure 5.10: Difference in efficiency between data and MC for combined (top left) and combined or tagged (top right) muons. The barrel and end-cap regions (solid histogram) are shown separately from the crack and transition regions (dashed histogram). The bottom plots show the significance, i.e. the difference divided by the statistical error on the data.

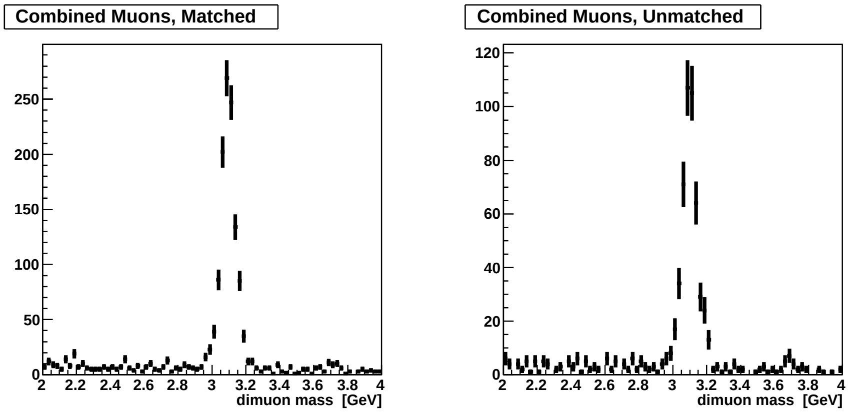

Figure 5.11: Dimuon invariant mass distribution when the probe is in the barrel ($0.1 < \eta < 1.1$) with 4 GeV $< p_T <$ 5 GeV. The case in which the probe is matched to a trigger object is shown at left. The right shows the unmatched case.

matching between reconstructed muons and trigger objects.

A simple qualitative estimate of the resolution function can be made by plotting the closest RoI to each reconstructed muon. This is shown in the first panel of Fig. 5.12.

The long tail extending out to values of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \approx 1$ can likely be ascribed to combinatoric background. That is, to additional muons in the event. However the RoI resolution function itself clearly extends beyond values of 0.3. When this is compared with the distribution of opening angles in J/ψ events, shown at right in Fig. 5.12, a clear overlap is observed.

Indeed, depending on the reconstruction-trigger matching algorithm, this effect can lead to a bias in the tag and probe results at high J/ψ p_T. In Fig. 5.13, the trigger efficiency is calculated from simulation in three ways. The control method uses an $\Upsilon(1S)$ sample. Because of the increased mass of the $\Upsilon(1S)$, the opening angle of the daughter muons is significantly wider and overlap effects due to the limited trigger resolution can be ignored. The other two methods are different variations of tag and probe using a J/ψ sample. The 'naive' method associates a reconstructed muon with a trigger if an RoI can be found within a ΔR of 0.4. A more refined method first proceeds identically to the naive method but then checks to see that all reconstruction-RoI associations are unique. If the associations are not unique - if two reconstructed muons have been matched to the same RoI - only the reconstructed muon closest to the RoI is retained as matched. The second reconstructed muon is reclassified as unmatched, regardless of how close it is to the RoI.

The naive method is clearly seen to lead to a bias that increases with p_T, while the method requiring unique association reproduces the efficiencies determined using the control method. In what follows only the method requiring uniqueness is used.

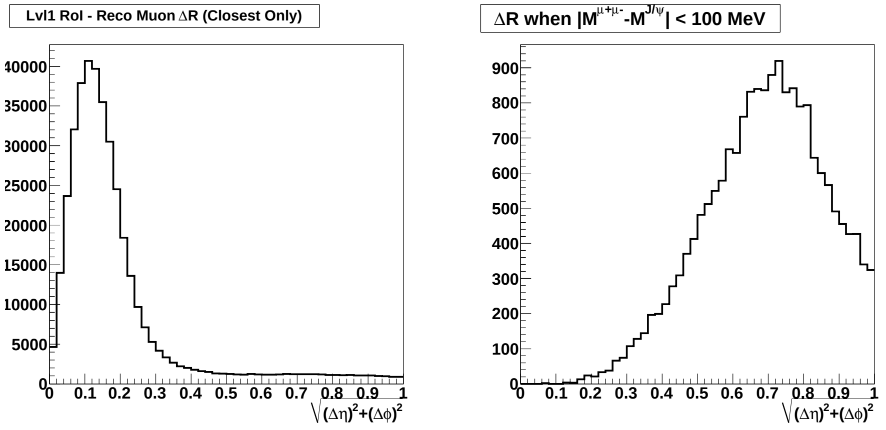

Figure 5.12: Shown at left is the ΔR between all reconstructed muons passing the selection and the closest lvl1 RoI. The peak at small values is a measure of the lvl1 RoI resolution. Shown at right is the ΔR between two reconstructed muons when the dimuon mass is within 100 MeV of the J/ψ mass.

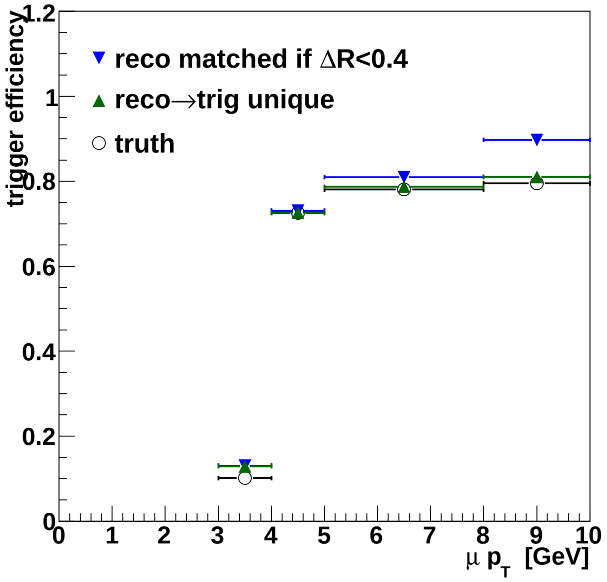

Figure 5.13: Trigger efficiency determined from simulation truth and two variations of tag and probe using a J/ψ MC sample. The 'inclusive matched' method overestimates the efficiency at high p_T. The 'unique matched' method requires that each RoI is associated with at most one reconstructed muon and reproduces the truth efficiency much more closely.

5.3.1 Event Selection

The event selection is nearly identical to the reconstruction tag and probe event selection. As before, probes are required to be combined, have $p_T > 4$ GeV and to be associated with a trigger. The tag and the probe must also have opposite electric charge. The only difference is that the probe must be a combined (tagged) muon when determining the trigger efficiency with respect to combined (tagged) muons.

Muons are then 'pre-labeled' as triggered or matched if a level 1 RoI and an EF_mu4 trigger feature can be found within ΔR values of 0.4 and 0.3, respectively. After pre-labeling each reconstructed muon, 2-to-1 reconstruction-trigger associations are vetoed as described above.

5.3.2 Trigger Efficiencies from Simulation

Like muon reconstruction efficiencies, trigger efficiencies are expressed as a product of finely grained maps made with simulation and coarser data driven scale factors. Fig. 5.14 shows the simulation maps. For combined muons, the high p_T asymptotic value of the efficiency in the end-cap is in the range 90%-100%. However in the barrel the efficiency plateaus at values between 75% and 85%. Like the reconstruction efficiency, the trigger efficiency in the central crack region and the barrel/end-cap transition region suffers from gaps in detector coverage.

The trigger efficiency with respect to tagged muons is poor at all values of p_T and η. This is completely consistent with the basic design of the tagging algorithm and the layout of the Muon System: the algorithm is designed to look for unmatched segments in the inner muon stations, and these stations lack trigger chambers altogether.

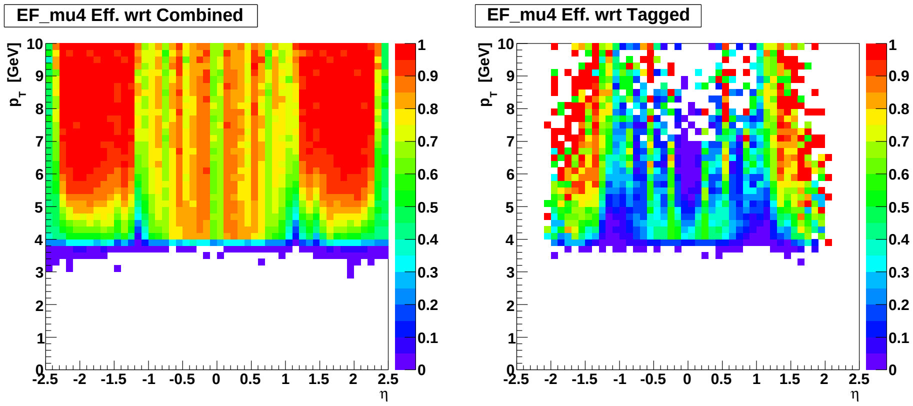

Figure 5.14: The efficiency for a combined (left) and tagged muon (right) to be associated with both a lvl1 RoI and an EF feature above 4 GeV as determined using simulation. In other words, the efficiency for a muon to satisfy the EF_mu4 criteria.

5.3.3 Trigger Efficiency Results

The results are shown in Figs. 5.15- 5.18. The agreement between simulation and data is extremely good. The most consistent disagreement is observed in the barrel/end-cap transition region, in which the simulation *underestimates* the efficiency by several percent. As this region is extremely challenging to correctly model - reduced symmetry of the instrumentation layout, inhomogeneous magnetic field strength, etcetera - any explanation of this disagreement would be purely speculative.

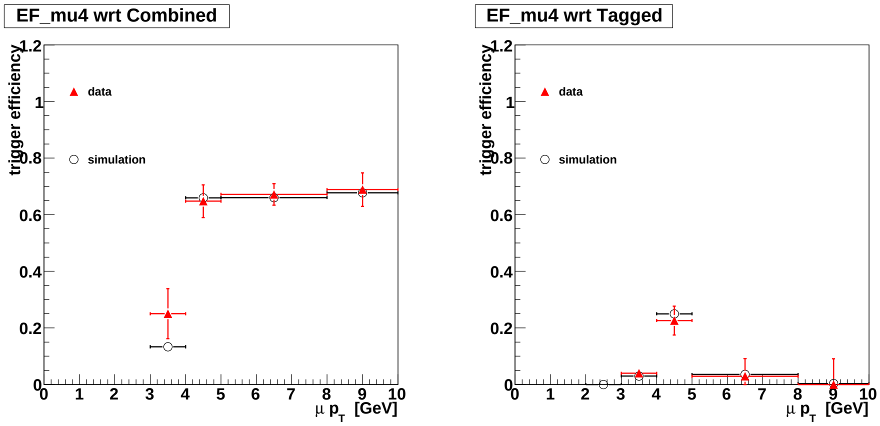

Figure 5.15: Shown are the measured muon trigger efficiencies for the central crack region ($|\eta| < 0.1$). Although the reduced trigger chamber coverage is evident in a high p_T asymptote of only ∼65%, the effect is seen to be successfully modeled.

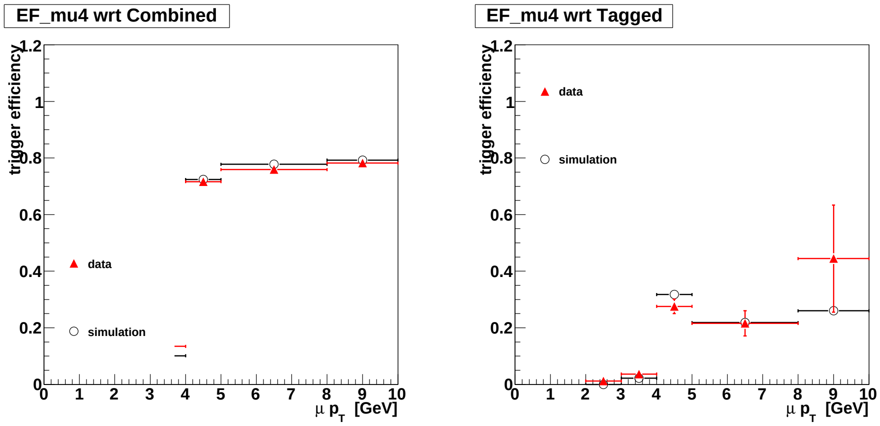

Figure 5.16: Shown are the measured muon trigger efficiencies for the barrel ($0.1 < |\eta| < 1.1$). A discrepancy between data and simulation, though evident, is within 2-3%.

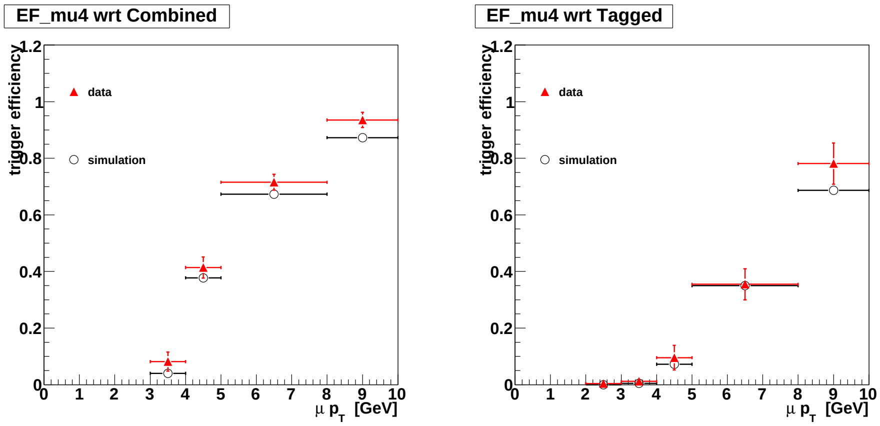

Figure 5.17: Results of the procedure for the barrel/end-cap transition region ($1.1 < |\eta| < 1.3$). A higher absolute efficiency is observed in data than in simulation.

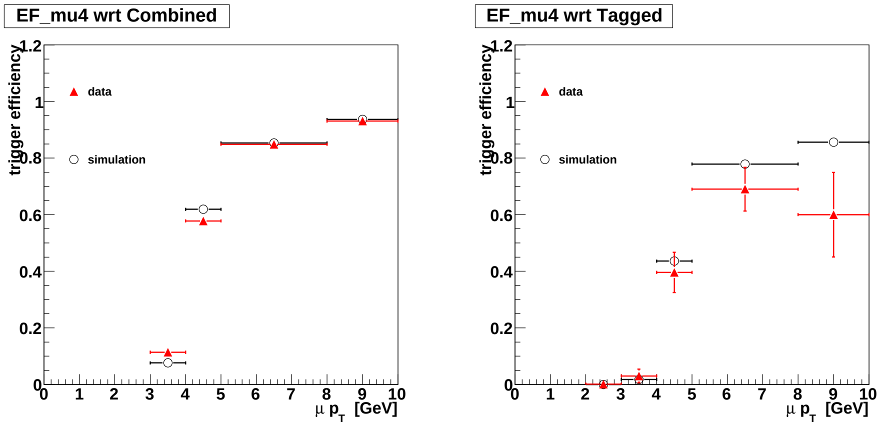

Figure 5.18: Results of the procedure for the end-caps ($1.3 < |\eta| < 2.5$). Above 4 GeV the agreement between data and MC is very good.

5.4 Reconstruction and Trigger Efficiencies Used in the Analysis

It should now be noted that the trigger and reconstruction efficiencies described above are not used in the following analysis; they should be viewed as cross checks. The efficiencies that do enter into the final cross section results are described in [62]. They differ in many regards including the precise input datasets and the use of background fits rather than sideband subtractions. However, the basic experimental technique is tag and probe.

Fig. 5.19 shows a qualitative comparison between the efficiency maps presented in Secs. 5.2.6 and 5.3.3 and those used for the final results. Discussions of the signal efficiency to pass the impact parameter (IP) cuts and the weighting formulae are the subjects of Secs. 5.5 and 5.7, respectively. Here it is noted that the comparisons of Fig. 5.19 use the same IP efficiencies and the same formulae. At central rapidity and low p_T^Υ as well as at high p_T^Υ for both rapidity examples, the agreement is seen to be excellent. A difference is visible at the level of several percent in the forward rapidity low p_T^Υ example. However this is well within the error of the *cross check efficiencies*: the errors on the efficiency maps range from 3% to 12%. When propagated to the $\Upsilon(1S)$ cross section - the method of doing so is discussed more fully in Sec. 7 - these lead to uncertainties of approximately 8%.

5.5 Impact Parameter Requirement and Efficiencies

The basic notion that suggests a cut on d_0 as a method to suppress background is sketched in Fig. 5.20. The signal, the $\Upsilon(1S)$, is prompt: its decay products are produced at the interaction point[3]. However a large background component comes from heavy flavor mesons decaying semimuonically. The lifetime of these particles is such that they only propagate hundreds of microns before decaying. However, because of the excellent spatial resolution of the Pixel detector the impact parameter of their daughter muons can be resolved.

The efficacy of such a cut is illustrated in Fig. 5.21, which shows two mass spectra for different choices of p_T^Υ and η^Υ. Irrespective of any quantitative considerations[4],

---

[3] Everything has a finite lifetime, but in the case of the $\Upsilon(1S)$ $c\hbar/\Gamma = \mathcal{O}(10^{-5})$ μm so its decay products are consistent with being prompt at a scale one one-millionth smaller than the detector resolution.

[4] It can be noted that $signal/\sqrt{background}$ is 11.0 (14.6) before (after) the cut for the spectrum shown at top left in Fig. 5.21. For the bottom left spectrum the analogous figures are 9.3 and 13.1. These statements are dependent on certain somewhat arbitrary choices of the width over which the background is integrated and are strictly a-posteri: the derivation draws on results in the following section and signal extraction techniques in Ch. 6.

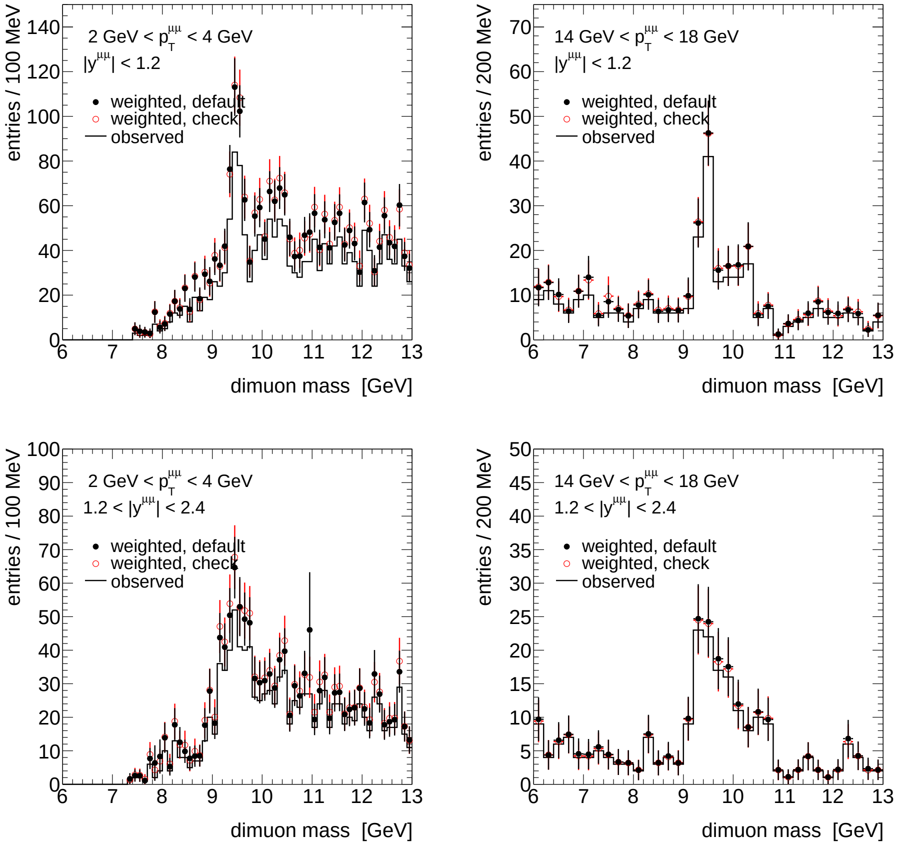

Figure 5.19: Four dimuon mass spectra shown before any weighting, with weights that are calculated based on the presented reconstruction and trigger efficiency maps, and with weights based on maps discussed in [62]. The latter are used in the following chapters and final results.

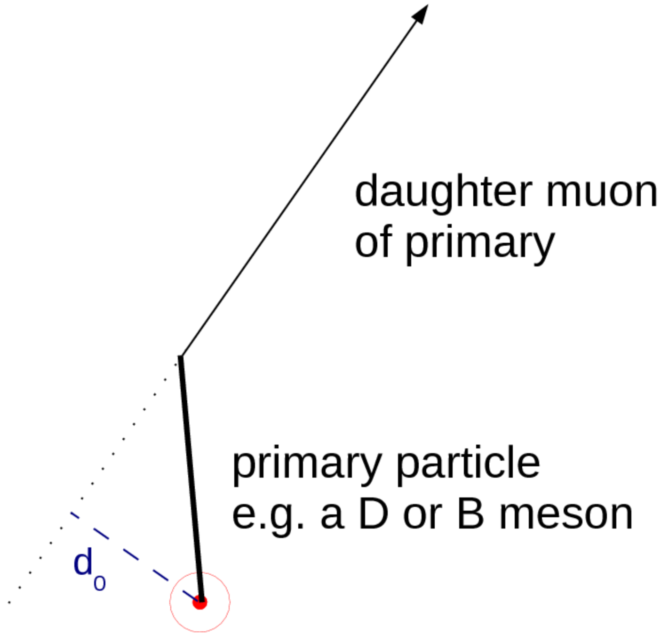

Figure 5.20: Sketch illustrating the production of a non-prompt muon via the semimuonic decay of a heavy flavor hadron. Such hadrons can propagate hundreds of micrometers or even millimeters. The process refers to many specific particles and decays e.g. $B^0 \rightarrow \mu^+ \nu_\mu D^-$ or $\Lambda_c^+ \rightarrow \mu^+ \nu_\mu \Lambda$.

the clarity of the peak is clearly enhanced by the cut.

The remainder of this section is devoted to measuring the efficiency for the signal to pass the impact parameter cut. Several methods were explored but in no case was a technique found that is as 'clean' and unambiguous as the methods used to determine the muon reconstruction and trigger efficiencies. Note that mirroring the tag and probe procedure is not a viable option; to do so would essentially amount to comparing the yield of J/ψ events with one muon passing the impact parameter cuts to the yield obtained after requiring both muons to pass the cut. The problem with such a procedure is that the ratio of these yields is determined not just by the detector performance, but by fairly complicated physics as well: a large fraction of J/ψs are produced not by direct QCD processes but as decay products of B mesons, and these will in general be displaced from the primary vertex by several hundred micrometers. In other words, muons from J/ψ events are not in general prompt.

The two methods that were ultimately used both make use of the same general strategy. First, the d_0 distribution of a certain muon selection in data is compared to the corresponding distribution in simulation. The results of this comparison are then used to derive smearing or scaling parameters in order to modify simulation to more closely match the distributions observed in data. Finally, the smeared or scaled simulation provides the impact parameter resolution function; this gives at once the

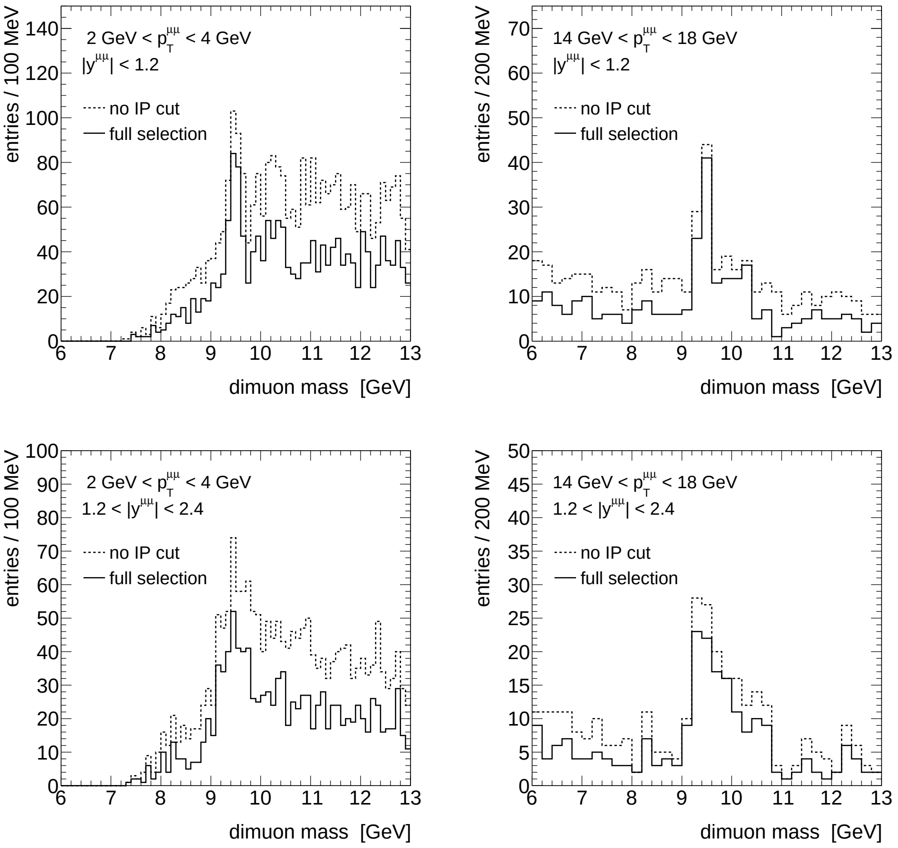

Figure 5.21: Shown are four mass spectra with the full event selection and with no cut on the muon impact parameters. The background is reduced by almost a factor of two while the change in the number of events in the peak at mass ∼9.5 GeV is negligible. These spectra correspond to those shown in Fig. 5.19.

cut efficiency for the signal since the muons from $\Upsilon(1S)$ events are all prompt.

The J/ψ *method* begins with an event selection that is identical to the cross section event selection, with the exception of the d_0 and $z_0 \sin(\theta)$ requirements which are dropped. In addition, the dimuon invariant mass must lie within 100 MeV of the J/ψ mass. Muons are then classified according to p_T^μ and η^μ. (To be clear we emphasize that this classification is done independently for the two muons in each event. For example if a given muon has $p_T^\mu = 4.5$ GeV and $\eta^\mu = 0.1$, it is classified accordingly irrespective of the kinematics of the second muon.)

Having constructed $N_{pTbins} \times N_{\eta bins}$ categories of muons, their d_0 distributions are now compared to Monte Carlo templates with Gaussian smearing ranging from 0 to 50 μm in 100 step sizes. Because J/ψs are not in general prompt, two templates are constructed for each kinematic bin and each degree of smearing: one template from a direct J/ψ sample and a second template from a $b\bar{b}$ sample. (Muons in this second sample come from both non-prompt J/ψ decay and the semimuonic decay of heavy flavor hadrons. Both processes result in muons that do not originate from the primary vertex and because of this these templates will be referred to as ‘displaced.’)

The smearing parameter that best describes the data is then determined by a procedure that can be termed a ‘2+1’ parameter fit. For each smearing parameter hypothesis, an unbinned extended maximum likelihood fit is performed in which the free parameters are the normalizations of the two templates; N_{prompt} and $N_{displaced}$. These constitute a family of 2-parameter fits indexed by the smearing parameter. The idea is illustrated in Fig. 5.22 which shows three examples chosen from the same 2-parameter family. A smearing parameter of 13 μm gives a fit that nicely describes the data while the other two examples, unsmeared Monte Carlo and 26 μm smearing, give fit results that are too narrow and wide, respectively, compared with the data.

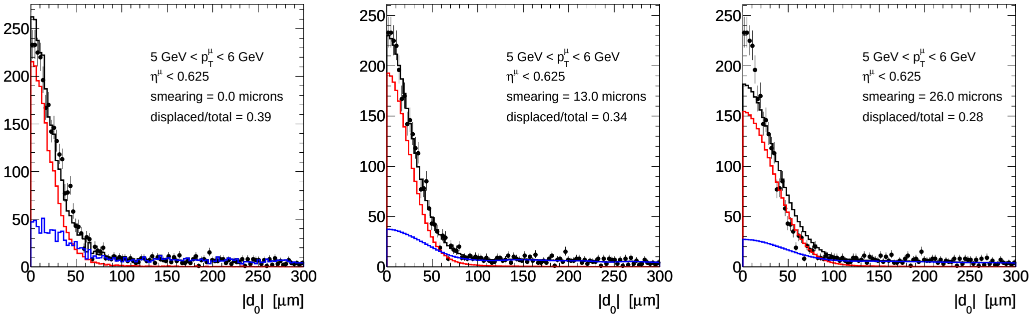

Figure 5.22: Three example fits in the $|\eta| < 0.625$, 5 GeV $< p_T <$ 6 GeV bin. The left panel is the fit to unsmeared Monte Carlo. The Monte Carlo templates in the middle panel have been smeared with a $\sigma = 13$ μm Gaussian; this fit results in the minimum NLL. The smearing parameter applied to the templates in the right panel is 26 μm.

This idea can be quantified by considering the minimum negative log likelihood (NLL) of the 2-parameter fits as a function of the smearing. For each p_T^μ and η^μ bin, this is shown in Fig. 5.23. The smearing parameter resulting in the minimum NLL is typically in the range of 10-20 μm and decreases for both increasing p_T and decreasing (more central) η. Having identified the best fit smearing value, the resolution function is then given by the smeared prompt template.

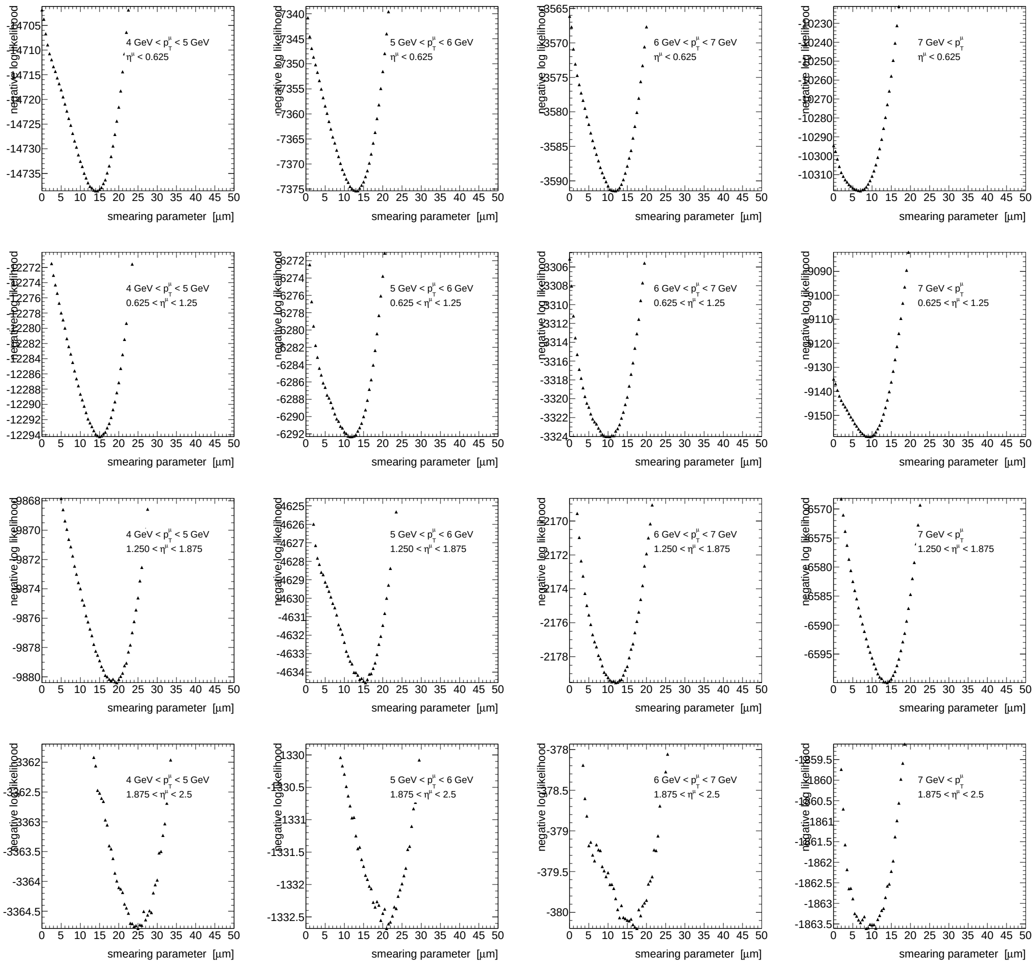

Figure 5.23: Each panel corresponds to a single bin in η^μ and p_T^μ and shows the NLL of the two template (promt and displaced) fit as a function of the smearing parameter value. η^μ and p_T^μ increase from top to bottom and left to right, respectively. The favored smearing value corresponds to the minimum of each plot. Note that the value of this smearing decreases with increasing p_T^μ.

The second method, the *Z method*, is complementary to the J/ψ method. The latter makes no assumptions about the p_T dependence of the cut efficiency, since the procedure is done independently for several bins in p_T^μ. This is possible given the p_T spectrum of muons from J/ψ decays. The drawback of the J/ψ method is that the fits are quite 'loose'. To wit they are, effectively, 3-parameter fits to a nondescript 1-dimensional distribution. The concern is that an excess of non Gaussian tails in the data distribution could be 'hidden' by assigning an incorrect prompt/displaced fraction.

Conversely, the p_T spectrum of muons in $Z^0 \rightarrow \mu^+\mu^-$ events is too hard to directly probe the IP resolution function for the low p_T values needed for this analysis. Muons from Z decays, however, do not contain a non-prompt component. Any study that begins by selecting on Zs will thus effectively swap the complication of having a sample with a nonprompt component for the complication of having to perform some sort of extrapolation down to low p_T. The details of the Z method are discussed in Appendix A.

The results of both methods are shown in Fig. 5.24. The efficiency measured with the J/ψ method - which is used for the cross section central values - exceeds 98% everywhere except for the 4-5 and 5-6 GeV bins in the highest η quartile, $1.875 < |\eta| < 2.5$. In the highest η^μ bin the results differ by upto 2.5% while in the three bins spanning $0 < |\eta^\mu| < 1.875$ the disagreement never exceeds 1%. This difference is propagated to the systematic uncertainty on the cross section, see Sec. 7. (It should be noted that in the extrapolation down to low p_T, the Z method allows for a 30% material uncertainty which is a worst case estimate.)

Although qualitative, an elementary check is included in which the impact parameter cut is in turn applied as described, reversed, and dropped altogether, see Fig. 5.25. Fits of the signal yield are the subject of Sec. 6; the only detail of relevance for this check is that the fit procedure is the same in all three cases. If one puts aside the large statistical errors in the reversed selection fit (to say nothing of systematic errors!) and takes the $\Upsilon(1S)$ yields at face value, this implies that the effective average efficiency is given by $\epsilon^2_{average} = 1 - 73/2250$ or $\epsilon_{average} \approx$98%.

5.6 Remaining Efficiencies

The efficiencies for three of the requirements in the above selection remain to be discussed: the efficiency to reconstruct a muon in the ID (including the Pixel and SCT quality requirements), the efficiency for a prompt muon to satisfy the $z_0 \sin(\theta)$ cut, and the efficiency for a signal event to have a reconstructed primary vertex with ≥ 3 tracks.

The ID reconstruction efficiency was also studied using $J/\psi \rightarrow \mu^+\mu^-$ events. The method is yet again a variant of the tag and probe idea. In this case a standalone muon and a combined muon with mass near that of the J/ψ is selected in order to provide a sample of probes that is unbiased with respect to the ID. The efficiency to reconstruct the track with 1 or more Pixel hits and 6 or more SCT hits is approximately 99.5% [63].

After dropping the primary vertex, d_0 and $z_0 \sin(\theta)$ requirements, there are 20,788 events. All of them have a primary vertex with 3 or more tracks and thus no correction is made for that part of the selection.

Similarly, no correction is made for the $z_0 \sin(\theta)$ requirement. In simulation the

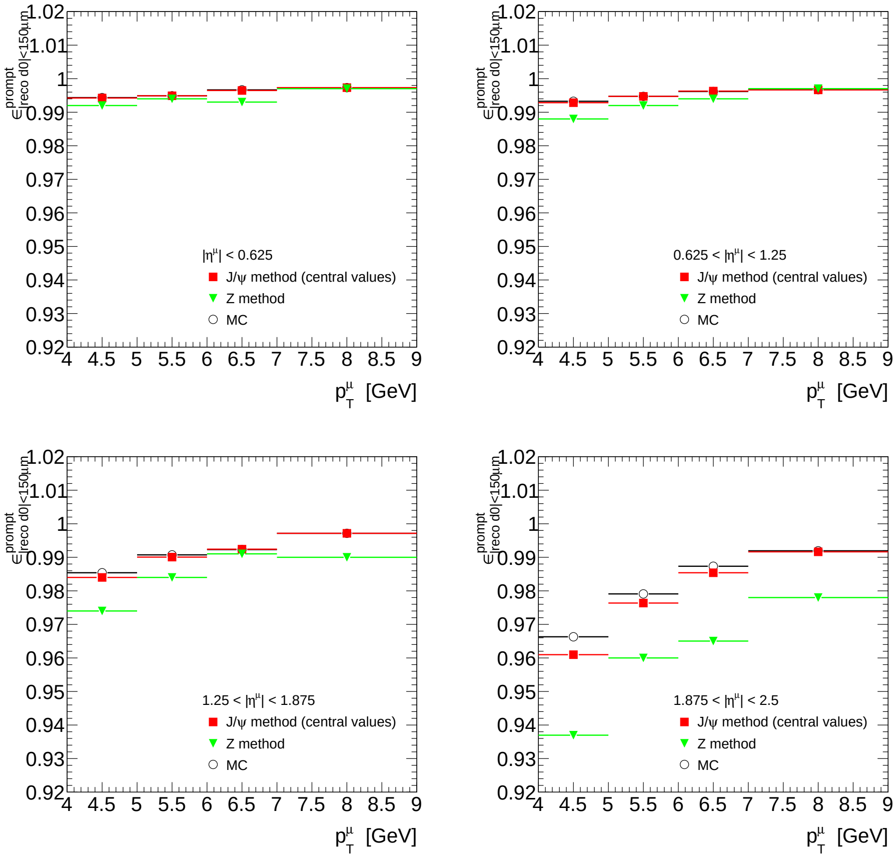

Figure 5.24: Efficiency of the impact parameter cuts as a function of p_T for four η ranges.

core of the $z_0 \sin(\theta)$ resolution is about 72 μm while in data it is 90 μm [64, 65]; both of these are more than an order of magnitude smaller than the cut value of 1.5 millimeters.

5.7 Efficiency Weights

5.7.1 Weight Formulae and Tests

The fraction of true dimuon events that are triggered, reconstructed, and selected can now be expressed in terms of the ingredients described above:

$$\epsilon_{\mu+\mu-} = \epsilon^{trig}_{\mu+\mu-}\epsilon^{MSreco}_{\mu+\mu-}\epsilon^{IDreco}_{\mu+\mu-} \tag{5.3}$$

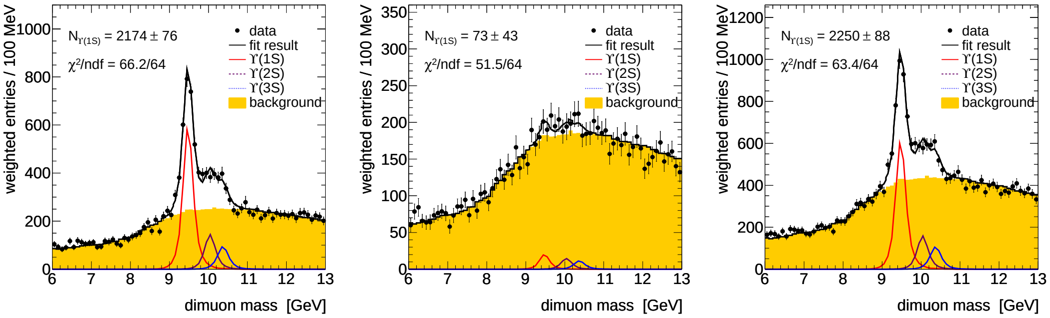

Figure 5.25: The default event selection is shown at left while in the middle the prompt cuts have been reversed: at least one of the two muons must satisfy either $|d_0| > 150\ \mu$m or $|z_0 \sin\theta| > 1.5$ mm. In the right plot the selection does not test on d_0 or $z_0 \sin\theta$ at all. In the reversed sample an excess of only small statistical significance is observed in the signal region.

where the individual terms are given by

$$\epsilon^{trig}_{\mu+\mu-} = 1 - (1 - \epsilon^{trig}_{+})(1 - \epsilon^{trig}_{-}) \tag{5.4}$$

$$\epsilon^{MSreco}_{\mu+\mu-} = \epsilon^{cb}_{+}\epsilon^{cb||tag}_{-} + \epsilon^{cb||tag}_{+}\epsilon^{cb}_{-} - \epsilon^{cb}_{+}\epsilon^{cb}_{-} \tag{5.5}$$

$$\epsilon^{IDreco}_{\mu+\mu-} = \epsilon^{track\&IP}_{+}\epsilon^{track\&IP}_{-} \tag{5.6}$$

The *event weight*, w, is then defined as $\epsilon^{-1}_{\mu+\mu-}$. The trigger, MS reco, and ID reco weights are similarly defined as $(\epsilon^{trig}_{\mu+\mu-})^{-1}$, etcetera.

The ϵ^{trig} terms appearing on the right hand side of Eq. 5.4 are functions of p^{μ}_{T}, η^{μ}, the charge of the muon *and the algorithm* with which the muon has been reconstructed; $\epsilon^{trig} = \epsilon^{trig\ wrt\ combined}$ in the case that the muon in question is combined, $\epsilon^{trig} = \epsilon^{trig\ wrt\ tagged}$ in the case that it is tagged.

The combined muon reconstruction efficiency, ϵ^{cb}, is a function of p^{μ}_{T}, η^{μ}, ϕ^{μ} and the charge of the muon. The ϕ dependence enters in through the data driven scale factors. (The tagged muon reconstruction efficiency is so well modeled by simulation that there are no scale factors and hence no ϕ-dependence.)

$\epsilon^{track\&IP}$ is given by the results of the J/ψ method for measuring the IP efficiency, shown in Fig. 5.24, times a flat factor of 99.5% for the track finding efficiency. It has no ϕ^{μ} or charge dependence.

The algebra of Eq. 5.4 follows from the use of a single muon trigger. Defining the inefficiency, $\iota \equiv 1 - \epsilon$, the relationship between dimuon and single muon quantities is thus given by $\iota_{\mu+\mu-} = \iota_{+}\iota_{-}$. (As supposed to $\epsilon_{\mu+\mu-} = \epsilon_{+}\epsilon_{-}$ in the case when both muons are required to fulfill a certain criterion.)

The algebra of Eq. 5.5 is best explained with a simple picture shown in Fig. 5.26. Given two types of muon reconstruction and two true muons, there are 9 mutually exclusive possibilities for how the dimuon system as a whole may be reconstructed.

The sum of the probability of all 9 outcomes is unity. The sum of the three outcomes meeting the event selection is given by the area of the (1,1), (1,2) and (2,1) cells in the diagram and gives the right hand side of Eq. 5.5.

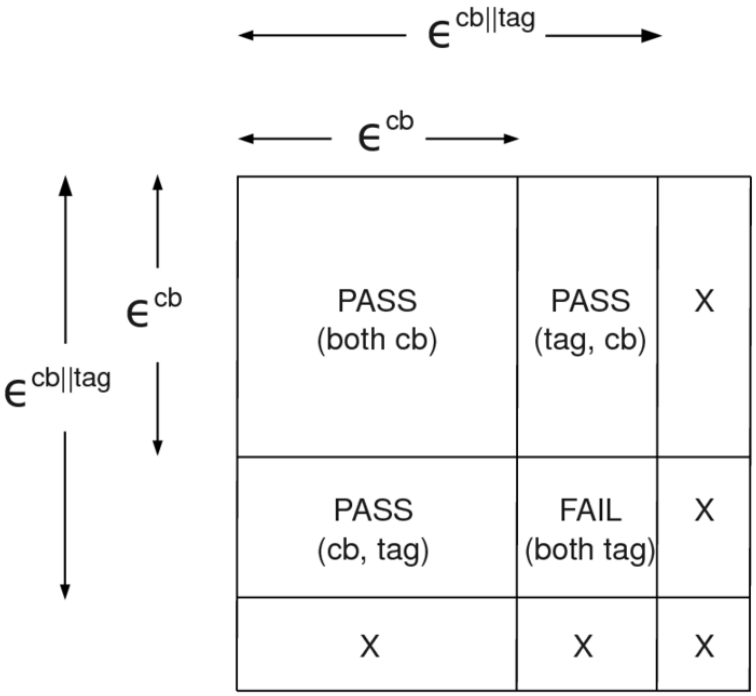

Figure 5.26: Diagram showing the 9 possibilities given an event with two muons and two 'grades' of reconstruction. Cells marked 'PASS' satisfy the event selection of having one or more combined muons. The single cell marked 'FAIL' has two tagged muons. The 5 remaining 'X' cells indicate possibilities in which one or more muon is not reconstructed at all.

In order to check the MS reco and trigger weight expressions, a closure test is performed using simulation. The results are shown in Fig. 5.27. The procedure is to select events in a $\Upsilon(1S)$ Monte Carlo sample and then reweight them according to Eqs. 5.4 and 5.5. The weighted distributions are then compared to the truth distributions in order to ascertain how well the latter are recovered. In 13 of the 16 bins, the deviation from perfect closure is less than 2% while the remaining 3 bins are within 1σ of 2%. The procedure closes to within 3% in all bins.

The test has two features that should be specifically noted. The first is a limitation: the test does not include an IP cut in the event selection, nor does it include terms in the event weight in order to account for inefficiencies incurred by such a cut. It therefore does not test the entire set of equations but takes for granted the validity of Eq. 5.6.

The second feature concerns the manner in which the truth and reconstruction histograms are filled. The truth histograms are filled with the truth Υ kinematics and only when the truth muons satisfy p_T >4 GeV and $|\eta|$ <2.5. The reconstruction histograms make no use of the truth kinematics but fill and test on only the reconstructed kinematics. Because of this, the effects of both muon migration across the fiducial boundary and dimuon migration across different p_T^{Υ} and y^{Υ} bins are also constrained by the closure test.

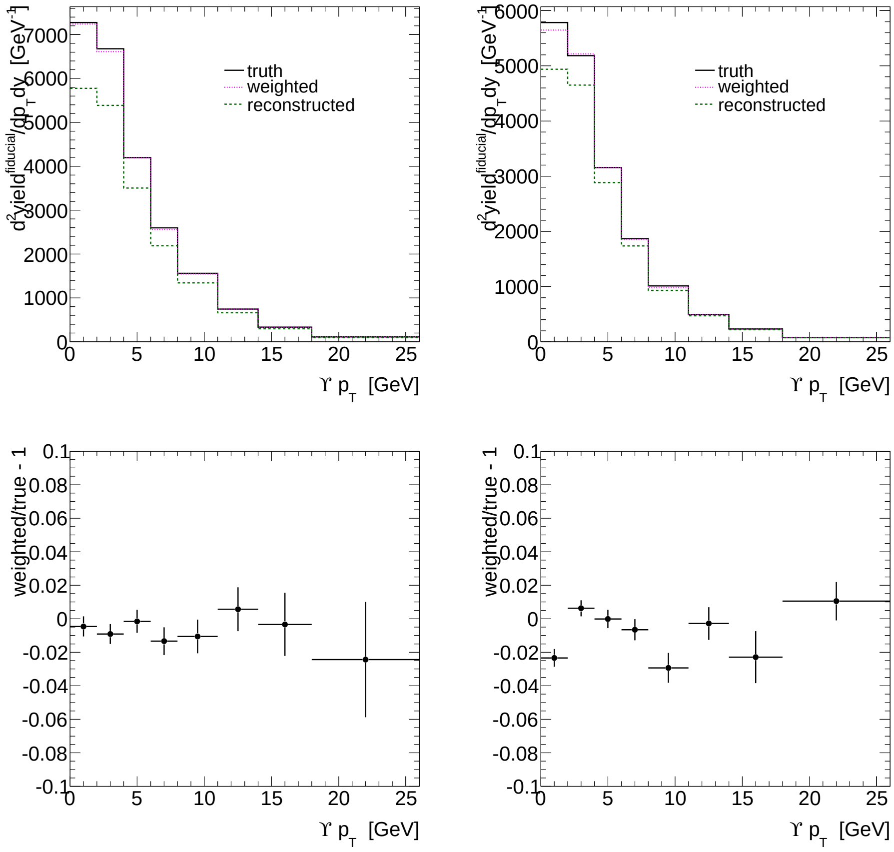

Figure 5.27: Shown are the truth, reconstructed and weighted p_T spectra of simulated $\Upsilon(1S)$ events (top) and the ratio (weighted-truth)/truth (bottom). Left panels are for $|y^{\Upsilon}| < 1.2$; right panels are for $1.2 < |y^{\Upsilon}| < 2.4$.

Because each weight is calculated based on the individual event kinematics (in both the closure test and the cross section extraction), the measurement strategy is independent of any assumptions on the polarization. To check closure for different polarizations, the above procedure is generalized to include a second weight.

This second weight has no relation to detector efficiencies but is instead inserted to mimic different polarizations using the same MC sample. It is calculated from truth kinematics only, and for a given event is strictly identical in both the truth and reconstruction histograms. The results of this generalization are shown in Fig. 5.28. A deviation slightly exceeding 3% is evident for one bin for the transverse- polarization. However the large majority of the 16×4 points in the polarized tests are within 2%.

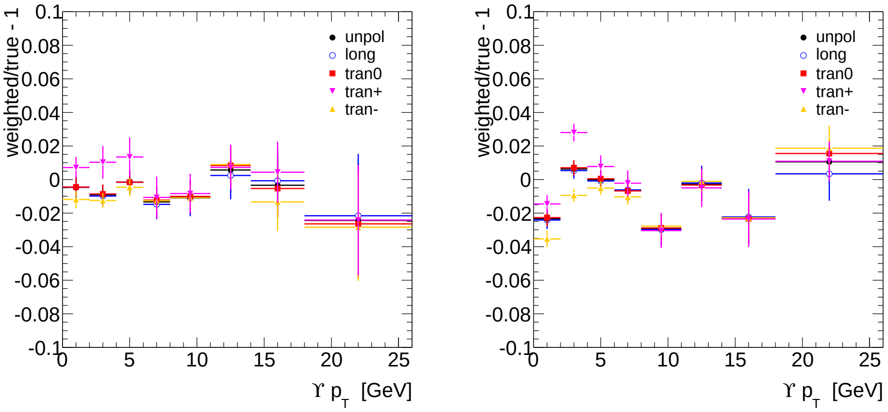

Figure 5.28: Shown is a generalization of Fig. 5.27 in which the same MC sample has been reweighted to mimic samples with different polarizations.

5.7.2 Average Weights in Data

The average weights applied to events passing the selection are shown in Figs. 5.29 and 5.30. The averages are taken over events with $|M(\mu^+\mu^-)-M(\Upsilon(1S))| < 600$ MeV and are performed separately for the 16 individual bins in $p_T^{\Upsilon(1S)}$ and $y^{\Upsilon(1S)}$.

The first set of plots shows the three factors in Eq. 5.3 separately. It is seen that the trigger weights are the largest of the three while the ID reconstruction weights are the smallest. Bins of increasing $p_T^{\mu^+\mu^-}$ are seen to contain events with smaller weights for all three factors. This corresponds to higher efficiencies. In the trigger and MS reconstruction weights, the central rapidity bins possess higher averages resulting from the lower efficiencies in the barrel compared with the end-caps. This is as one would expect from Figs. 5.2 and 5.14. The situation is reversed in the ID reconstruction efficiencies.

The second set shows the product of the three factors: the average event weight. These range from approximately 1.38 to 1.08. It also shows the composite weight, defined as the final yield of signal events, $N(\Upsilon(1S))$, divided by the event yield obtained before the application of any weights, $N_{raw}(\Upsilon(1S))$. (Both of these are the result of likelihood fits described in the following chapter.) The ratio of these two provide a second estimate of the size of the weights. The composite and average weights need not be strictly identical, as the former is sensitive to changes in the background shape incurred by the weighting and statistical fluctuations in the number of events (and size of the weights) outside of the range used for the average. However, they are seen to be closely correlated.

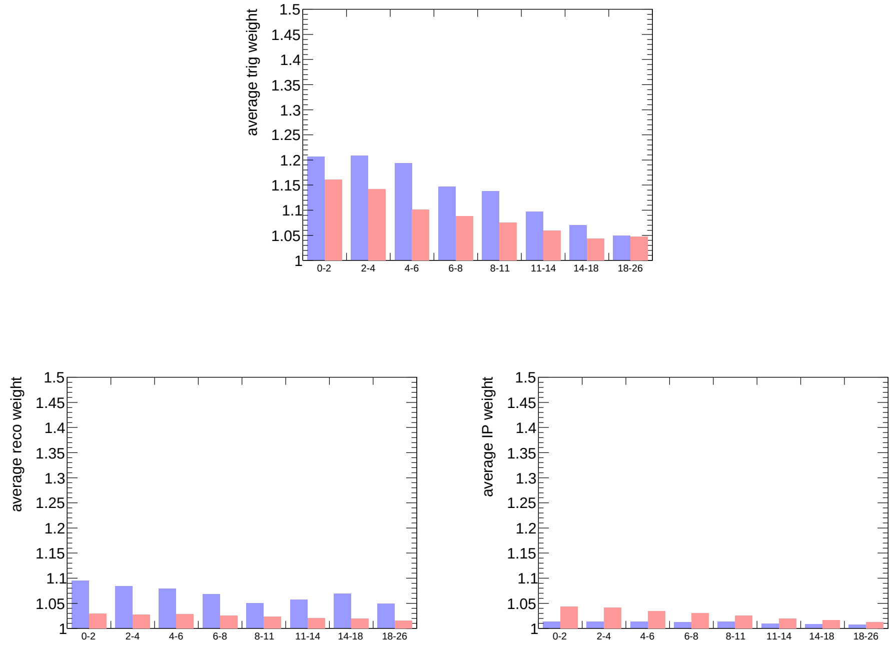

Figure 5.29: Average trigger, MS reconstruction and ID reconstruction weights for events with $|M(\mu^+\mu^-) - M(\Upsilon(1S))| < 600$ MeV shown at top, left and right respectively. Note that the range of the vertical axis does not extend below 1. The horizontal axis corresponds to the $p_T^{\mu^+\mu^-}$ *bin* (not $p_T^{\mu^+\mu^-}$).

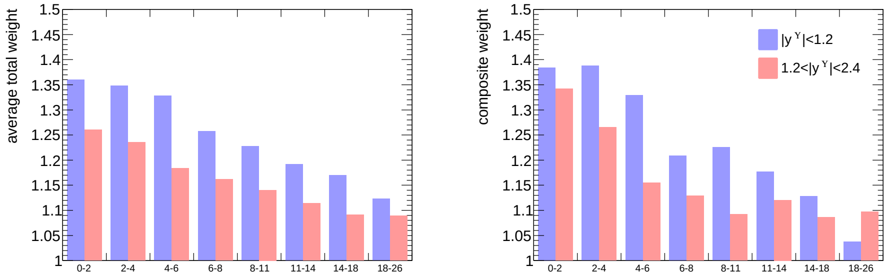

Figure 5.30: Average event weight for events with $|M(\mu^+\mu^-) - M(\Upsilon(1S))| < 600$ MeV shown (left) and 'composite' weights (right).

Chapter 6

Shape Modeling and Mass Fits

One way to summarize Ch. 5 is to say that it consists of a series of methods to decide 'how many events to put back in,' see Fig. 6.1.

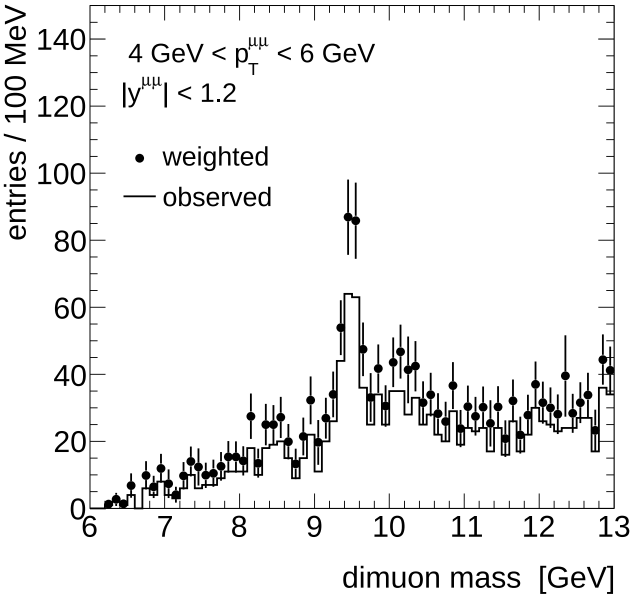

Figure 6.1: The solid histogram shows the reconstructed dimuon mass spectrum for events with 4 GeV< $p_T^{\mu^+\mu^-}$ <6 GeV and $|y^{\mu^+\mu^-}| < 1.2$. The dotted distribution is the spectrum after weighting.

The goal of this chapter will be to quantify how many events are in the peak centered at the $\Upsilon(1S)$ mass, ~9.5 GeV. The strategy is essentially to decompose the mass spectrum into component pieces, and then to reassemble the components such that the composite function approximates the spectrum in data as closely as possible.

The number of events in the peak is then given by the number of events in the 'peak component.'

This two step process is illustrated in Figs. 6.2 and 6.3. Fig. 6.2 shows the individual components, or *probability distribution functions* (PDFs). Their construction is described in Secs.6.1 and 6.2. A prominent feature of both classes of PDFs is that they have no free shape parameters. Thus the fit - the reassembly step discussed above - involves only the determination of the normalizations of each component.

Fig. 6.3 shows the resulting composite function - the fit result - obtained after maximizing the *extended likelihood* function. The definition and meaning of this function will be discussed in Sec. 6.3.

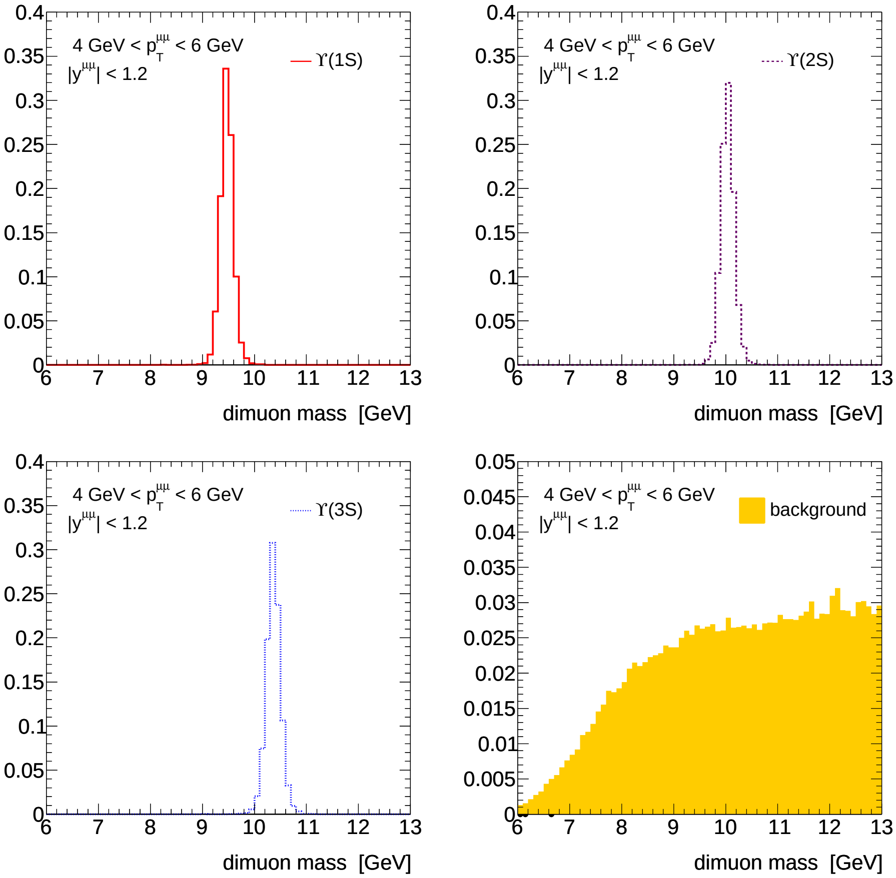

Figure 6.2: Shown are the four PDFs used to extract N(Υ(1S)) in the range 4 GeV< $p_T^{\Upsilon(1S)}$ <6 GeV and $|y^{\Upsilon(1S)}| < 1.2$, the same range as the distributions shown in Fig. 6.1. The three narrow peaks (the signal PDFs) are models of the detector resolution function at the mass of the Υ(1S), Υ(2S) and Υ(3S). The function extending over the full range of the horizontal axis (the background PDF) is a model of the mass spectrum produced by non-resonant background.

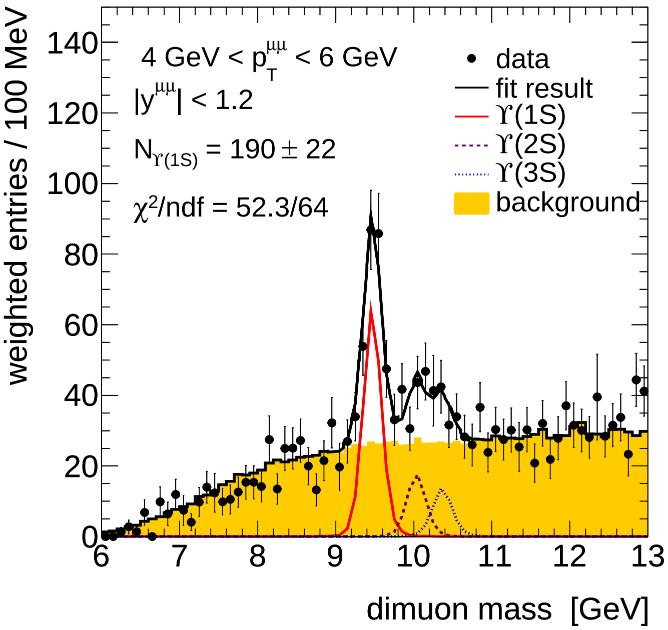

Figure 6.3: The black line is the fit result: the closest reproduction of the data distribution (filled circles) that can be achieved by varying the normalization of the 4 PDFs and summing them.

6.1 Signal Modeling

The signal PDFs are modeled using three Monte Carlo samples, one each for the respective resonance. Constructing the signal PDFs in this way, rather than just translating the 1S PDF by Δm_{2S-1S} (Δm_{3S-1S}) for the 2S (3S) PDF takes into account the dependence of mass resolution with mass.

This dependence can be estimated using the same two equations that were used to calculate the $\Upsilon(1S)$ mass resolution in Sec. 3.3. For tracks with momentum $\approx m(\Upsilon(1S))/2$, the error on the momentum is dominated by the multiple scattering term. In this limit, and again making the approximations that both muons are massless and back-to-back, the fractional error on the mass resolution σ_m/m scales as a constant $(1.6\%/\sqrt{2}.)$ Therefore, the ID mass resolution at $m_{\Upsilon(3S)}$ can be expected to be worse than at $m_{\Upsilon(1S)}$ by almost 10%.

The dimuon mass resolution observed in the study $Z^0 \rightarrow \mu^+\mu^-$ events is slightly worse than that predicted by simulation[1]. A smear-and-fit method [66] similar in idea to the J/ψ method discussed in Sec. 5.5 can be used to derive parameters to be applied to 'default' simulated samples to more closely model the momentum resolution. The

---

[1] Unlike the $\Upsilon(1S)$ peak, it is easy to draw conclusions about detector performance from the Z^0 peak because there are no nearby resonances and very little background.

smearing recipe is

$$p_T^{smeared} = p_T(1 + g_{(0,1)} f(\theta)\Delta \times p_T) \tag{6.1}$$

where $g_{(0,1)}$ is randomly generated from a Gaussian distribution with a mean of 0 and unit width. $f(\theta)$ is equal to $tan^{-2}(\theta)$ if $|\eta| > 1.9$ and is equal to 1 otherwise. Δ is an η dependent constant given by Table 6.1.

region	$\lvert\eta\rvert$ bounds	Δ $[TeV^{-1}]$
barrel	< 1.05	0.396±0.044
transition	$1.05 - 1.7$	0.900±0.091
end-caps	$1.05 - 2.0$	1.324±0.045
CSC/no TRT	$2.0 - 2.5$	0.129±0.004

Table 6.1: Table of constants used to smear the reconstructed p_T of simulated muons in order to emulate the momentum resolution observed in data.

Note that for values of $|g_{(0,1)}| \approx 1$, $p_T^{\mu} \approx$4 GeV and central η, the smearing term is approximately 0.4 TeV^{-1} × 0.004 TeV = 0.16%; quite a small[2] effect.

6.2 Background Modeling

While the signal PDFs are determined solely by detector response, the continuum background PDF is determined largely by kinematics.

As a limiting case, consider a dimuon system with $p_T^{\mu+\mu-} = 0$ GeV, $p_T^{\mu+}$ >4 GeV and $p_T^{\mu-}$ >4 GeV. Such a system must have a mass greater 8 GeV. Thus in the lowest $p_T^{\mu+\mu-}$ bins there is a kinematic cut off below which the number of events must be zero. The general point is that given the acceptance cuts (p_T^{μ} >4 GeV) and the mass of the $\Upsilon(1S)$ (9.460 GeV), the phase space accessible to background events is rapidly varying over the signal region.

In order to model the background shapes, three sets of templates are constructed; two sets from data and one set from Monte Carlo. The templates used for the extraction of the cross section central values are constructed from data by selecting on one combined muon and one Inner Detector track. The selection then proceeds identically to the event selection used for the cross section measurement. In particular the acceptance and IP cuts are applied, and the muon-track pair must have opposite electric charge. These templates will be referred to as the 'opposite sign (OS) μ+track templates'.

The second set of templates is made by repeating the OS μ+track selection, but with same sign (SS) muon+track pairs.

---

[2] A more extreme example would be $p_T^{\mu} \approx$10 GeV and $\eta \approx$2 giving a 17% smearing.

For the Monte Carlo templates, the selection is the same as the measurement: one combined muon and a second muon, opposite in sign from the first, that may be either combined or tagged. The sample is *open* $b\bar{b}$ and $c\bar{c}$: not quarkonium such as the ψ and Υ families, but states with open heavy flavor such as D and B mesons. Events with two final state muons are then produced when two heavy flavor particles each decay semimuonically.[3]

Fig. 6.4 shows the three sets of templates for the same 4 kinematic bins shown in Figs. 5.19 and 5.21. In all 4 bins the agreement between between the three different sets is seen to be extremely good. The two low p_T bins, as expected, both display a kinematic cutoff below which there are no events. Above this cut off, the background rate rapidly increases and then peaks. In the two high p_T bins, the background shape does not have these complex features and is very nearly linear.

6.3 Maximum Likelihood Fits

After constructing the signal and background shapes, the normalization for each component PDF must be determined. This is done by maximizing the extended maximum likelihood (ML) function. Before defining this function, the idea behind it is briefly reviewed.

Fig. 6.5 shows two fits to the same distribution. In both cases, the fit result is a sum of two Gaussians, each with unit width, with means of -2 and +2; the fit results only differ in the normalization of the two components.

The fit on the right is clearly preferable to that on the left, but how should the preference be quantified?

One idea is to assign a ‘penalty function’ to measure the deviation between the fit result and actual bin values. This is the idea behind the χ^2, defined as $\sum_{j \in bins} (h_j - f_j)^2/\sigma^2_{h_j}$ where h_j is the height of the jth bin and f_j is the value of the fit hypothesis at the center of the jth bin. (The χ^2 dependence on the fit parameters is implicit in this notation; $f_j = f(\nu_L, \nu_R; x_j) = \nu_L g_L(x_j) + \nu_R g_R(x_j)$ where ν_L is the normalization of the Gaussian on the left, etcetera.) The best fit is then defined as the hypothesis that minimizes the χ^2. Note that the data must be binned in order to define the χ^2.

Alternatively, one could note that the right panel in Fig. 6.5 has more ‘amplitude’ where there are more events. It is far more likely that the distribution could be obtained by sampling the hypothesis function on the right than on the left.

To quantify this notion, note that if the random generation of a single number x from a distribution f has a likelihood proportional to $f(x)$, the likelihood $\mathcal{L}$ for a sample with multiple points is the product of the function evaluated at every point

---

[3] A cascade decay of a single hadron such as $B \to \mu + D + X \to \mu\mu + K + X + X'$ can also result in two final state muons with OS. But as the $\mu^+\mu^-$ mass is necessarily lower than the mass of the mother, such events will be outside of the signal region.

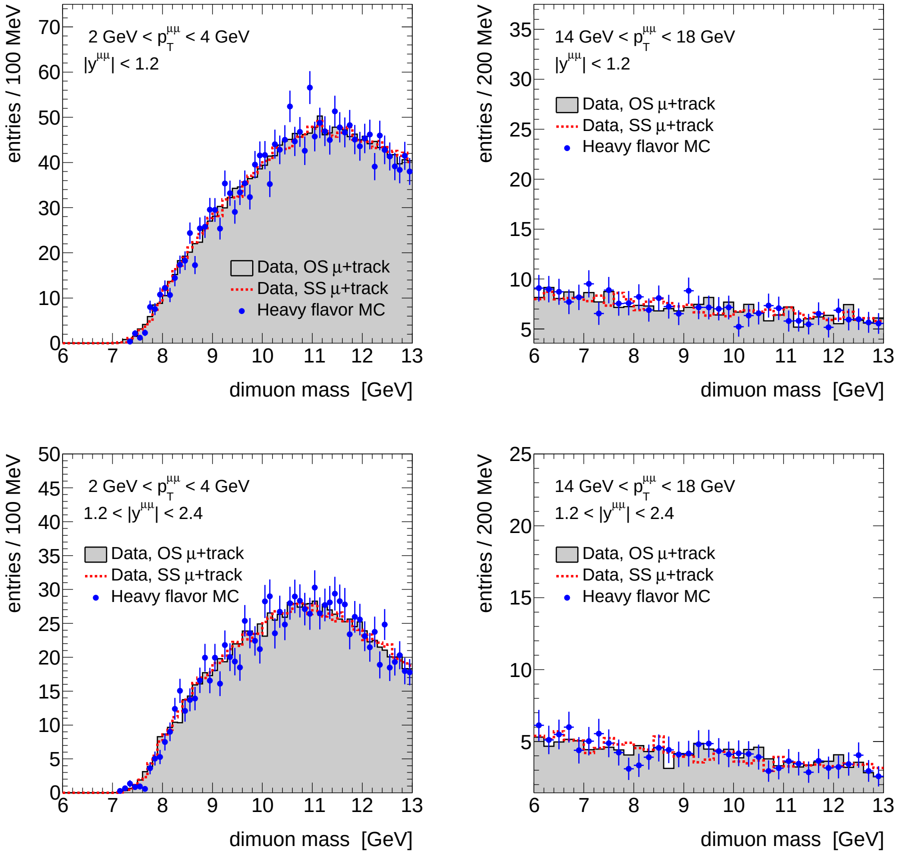

Figure 6.4: The three sets of background templates for four selected bins in $p_T^{\mu^+\mu^-}$ and $y^{\mu^+\mu^-}$. The MC templates have the largest statistical errors and are thus shown with markers. All histograms are normalized to the number of background events determined in the fit results, cf. Fig. 6.8.

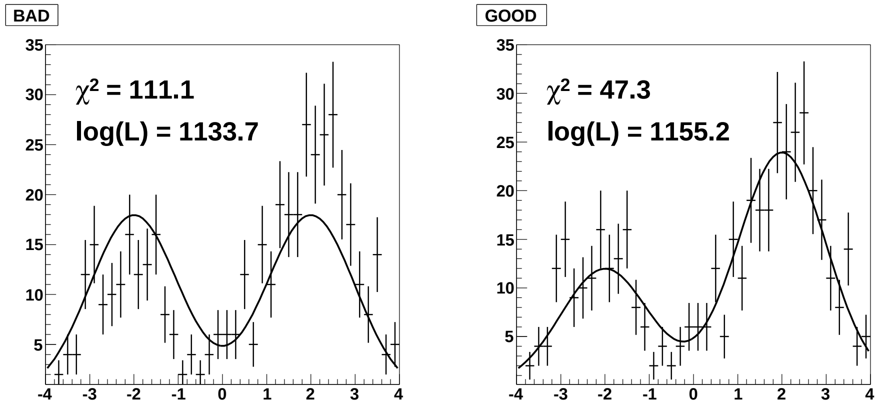

Figure 6.5: Two example fits for the same 'data' distribution. The fit shown at left overestimates the distribution at negative values of the independent variable and underestimates it at positive values. The fit shown at right is the function used to generate the distribution. Note that it has a lower χ^2 and higher likelihood than the fit at left.

in the ensemble: $\mathcal{L} = \prod_{x_j \in sample} f(x_j)$. (In the binned case this becomes $\prod_{j \in bins} f_j^{h_j}$.) With $\nu_L + \nu_R$ constrained to be equal to the size of the data sample[4], the likelihood technique then chooses the best fit by maximizing $\mathcal{L}$.

6.3.1 The Extended Likelihood

The function defined above is sometimes referred to as the classical likelihood [67]. The *extended* likelihood function is defined [68] as

$$\mathcal{L}_E = \frac{e^{-\nu}\nu^N}{N!}\mathcal{L} \tag{6.2}$$

where $\nu = \nu_L + \nu_R$ (there is now no longer any constraint on $\nu_L + \nu_R$) and N is the total size of the sample. The additional term is the probability of obtaining N events when the expectation value of the number of events is ν.

Eq. 6.2 emphasizes the differences between the two methods. However as in the example of the χ^2 method the briefest notation also obscures the dependence on the fit parameters and the dataset. Using the same example as above in which the model is a sum of two Gaussians ('left' and 'right') with fixed mean and width, i.e. only normalizations are varied, $\mathcal{L}_E$ is defined as

---

[4] Without this constraint the likelihood could be made arbitrarily large by making ν_L or ν_R arbitrarily large.

$$\mathcal{L}_E(\nu_L, \nu_R; \overrightarrow{x}) \quad = \quad \frac{e^{-(\nu_L+\nu_R)}(\nu_L+\nu_R)^N}{N!} \prod_{j=1}^{N} \nu_L g_L(x_j) + \nu_R g_R(x_j) \tag{6.3}$$

and the best fit condition is

$$\frac{\partial \mathcal{L}_E(\nu_L, \nu_R; \overrightarrow{x})}{\partial \nu_L} = 0, \qquad \frac{\partial \mathcal{L}_E(\nu_L, \nu_R; \overrightarrow{x})}{\partial \nu_R} = 0 \tag{6.4}$$

A clear discussion of the difference between the result obtained from maximizing $\mathcal{L}$ and that obtained from maximizing $\mathcal{L}_E$ is [67]. The crucial point is not the fitted yields themselves, which are identical in both formalisms, but the *error* on the yields. The classical likelihood provides the error on the yields given a fixed total: it answers the question 'what fraction of events should be assigned to a given PDF.' The extended likelihood answers a different question: 'how many events should be assigned to a given PDF.'

As a toy example to illustrate that these questions may imply identical central values with different errors, consider a sample of events that divide unambiguously into two categories with yields denoted P and F (for pass and fail.) The error on the number of passing events is $\sigma_P = \sqrt{P}$ and is independent of F. The error on the fraction of passing events is $\sigma_\epsilon = \sqrt{(\epsilon)(1-\epsilon)/(P+F)} = \sqrt{(PF)/(P+F)^3}$ where $\epsilon \equiv P/(P+F)$. The point is that $(P+F) \times \sigma_\epsilon$ is most certainly not equal to σ_P.

As the goal of this analysis is to count $\Upsilon(1S)$ events, and not, for instance, to measure a branching fraction, the signal is extracted by the extended ML technique.

6.3.2 Implementing the Likelihood Maximization

The fits were performed with the RooFit v3.12 libraries [69] in the ROOT 5.26/00e framework [70]. The core functionality of these libraries is to build the likelihood function given the PDFs and the data sample; the actual extremization is carried out with a port of the MINUIT program [71].

6.4 Pseudo Experiments

In order to check that the fit results are not biased, and that the statistical error calculation is performed correctly, 1000 pseudo experiments are performed for each p_T^{Υ} and y^{Υ} bin. For a given pseudo experiment and a fixed bin, 4 Poisson distributions are sampled with means equal to the yields observed in data for that bin: $N(\Upsilon(1S)), N(\Upsilon(2S)), N(\Upsilon(3S))$ and $N(BG)$. These 4 randomly generated numbers - for the jth pseudo experiment call them $N^j_{true}(\Upsilon(1S)), N^j_{true}(\Upsilon(2S))$ etcetera - then determine the 'true' yields of each of the 4 components in that pseudo experiment; the

$\Upsilon(1S)$ PDF is sampled $N^j_{true}(\Upsilon(1S))$ times, ..., and the background PDF is sampled $N^j_{true}(BG)$ times. Then all four distributions of pseudo data are added together.

The fit is then performed identically to how it is on data, and the pull $\equiv (N^j(\Upsilon(1S)) - N(\Upsilon(1S)))/\sigma^j(\Upsilon(1S))$ is recorded. Fig. 6.7 shows the distribution of pulls for two example p_T and y bins. A Gaussian fit is made to both distributions. In both cases the mean is consistent with 0 and the width is consistent with 1, to within statistical error.

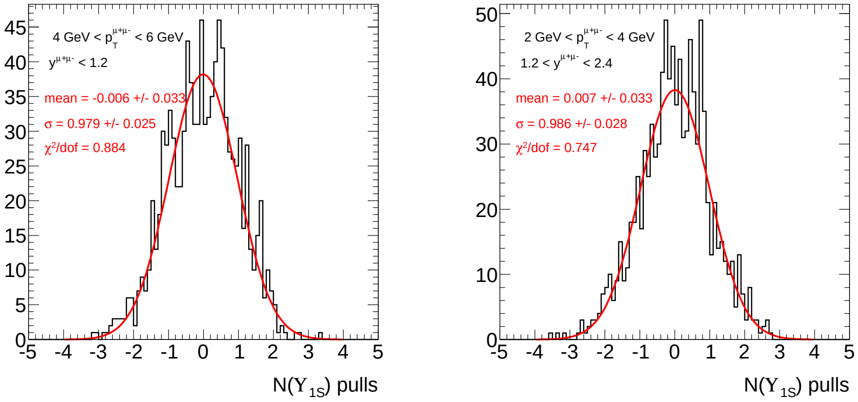

Figure 6.6: Pull distribution for two selected $p_T^{\Upsilon(1S)}$ and $y^{\Upsilon(1S)}$ bins. Shown also are the results of fitting a Gaussian function to both distribution.

Fig. 6.7 summarizes all 16 pseudo experiment ensembles. The means appear largely consistent with 0, though in several cases there are deviations approaching 10%[5]. Similarly, the widths are always within 10% of unity.

6.5 Results

The fits for the four 'standard example' bins (those corresponding to Figs. 5.19, 5.21 and 6.4) are shown in Fig. 6.8. It is seen that the background structure is quite well modeled by the OS μ+track templates. The mass resolution in the forward rapidity bins is notably worse than in the central bins, and this is reflected in the width of the signal PDFs.

Table 6.2 shows the fit results for $N(\Upsilon(1S))$, $N(\Upsilon(2S))$ and $N(\Upsilon(3S))$ for all 16 bins. The χ^2 and χ^2 probability are also shown. The χ^2 probability is never less than 6%, indicating reasonable goodness-of-fit.

---

[5] When discussing pulls, the units are in the statistical error of that bin. For most bins in this analysis the statistical error is ≈10%, meaning a 10% shift of the pulls is a 1% bias on the yield.

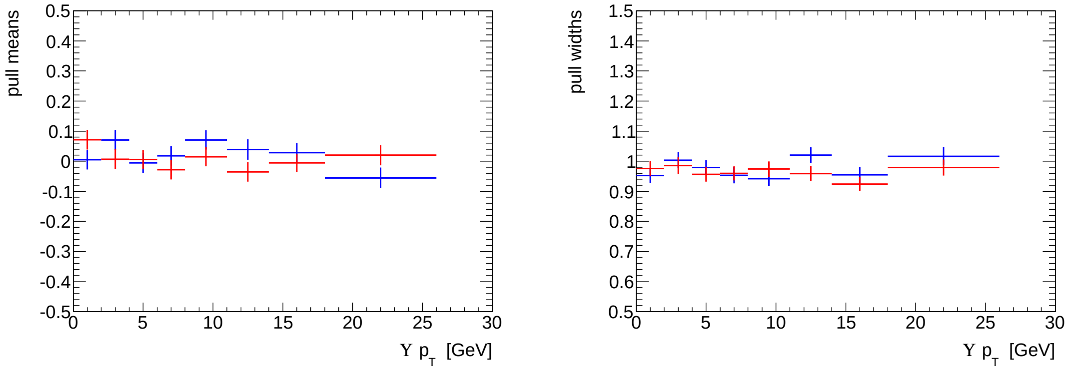

Figure 6.7: Mean (left) and width (right) of the Gaussian fits to the pull distributions for the fitted number of $\Upsilon(1S)$ events as a function of $p_T^{\Upsilon(1S)}$ for the central (red) and forward (blue) region.

6.6 Coda: Why Templates?

In the fits presented above, the likelihood function is maximized only with respect to *normalization* parameters such as $N(\Upsilon(1S))$ and $N(BG)$. This is in stark contrast to previous Υ cross section measurements [57, 58, 59], in which each resonance is modeled by an analytic function such as a Gaussian or a Crystal Ball function [72] and the background model is a simple polynomial. In these fits, the likelihood or χ^2 is also extremized with respect to shape parameters such as the detector resolution or the background dependence. As a practical matter this is easier to implement since it amounts to 'letting the data choose' the peak width or the background shape rather than having to construct the precise PDF beforehand. Why then does the current analysis not proceed in a similar fashion?

The primary reason concerns the comparatively hard p_T^μ cut of 4 GeV. In [59], for instance, the p_T^μ cuts are 3.5 and 2.5 GeV, depending on η^μ. Having a harder cut pushes the $\mu^+\mu^-$ minimum mass threshold to a correspondingly higher value and results in more rapidly varying background under the signal region. At 4 GeV, the background is sculpted to such an extent that a polynomial is insufficient to adequately describe the threshold part of the spectrum, as shown in Fig. 6.9.

In these fits a 4th degree polynomial is used for a background model. The signals are modeled with Gaussians. The widths are constrained such that $\sigma_{\Upsilon(1S)}$:$\sigma_{\Upsilon(2S)}$:$\sigma_{\Upsilon(3S)}$ = $\text{mass}_{\Upsilon(1\text{S})}$:$\text{mass}_{\Upsilon(2\text{S})}$:$\text{mass}_{\Upsilon(3\text{S})}$, but the single parameter unconstrained by the two mass ratios is freely varied.

Comparing Figs. 6.8 and 6.9 shows that while the high p_T^Υ bins have similar χ^2/ndf and similar yields for both procedures, the polynomial fit is clearly a nonstarter for the low p_T^Υ bins.

A second reason is the comparatively poor dimuon mass resolution - for *low* $\mu^+\mu^-$

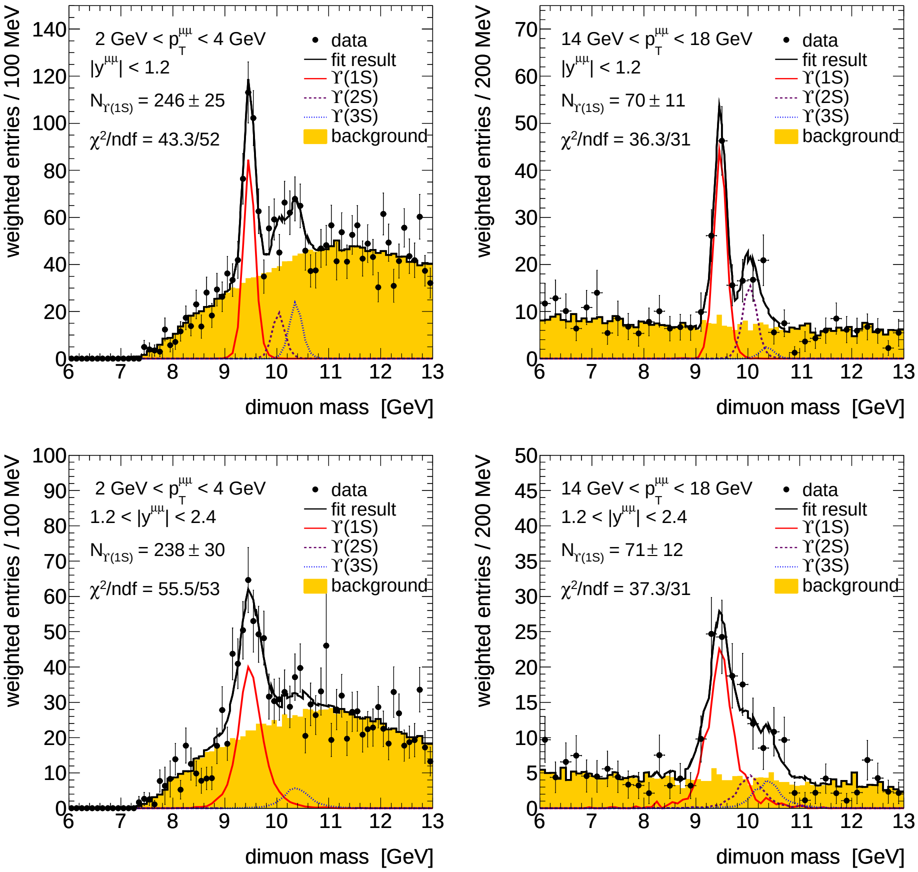

Figure 6.8: Fit results for four selected bins. The data (filled circles) are shown together with the result of the unbinned extended maximum likelihood fit (histogram) as explained in the text. The shaded histogram shows the default background PDF, the OS μ+track templates. The three other histograms show the signal PDFs. The jagged features noticable in the high p_T bins are due to limited Υ MC statistics.

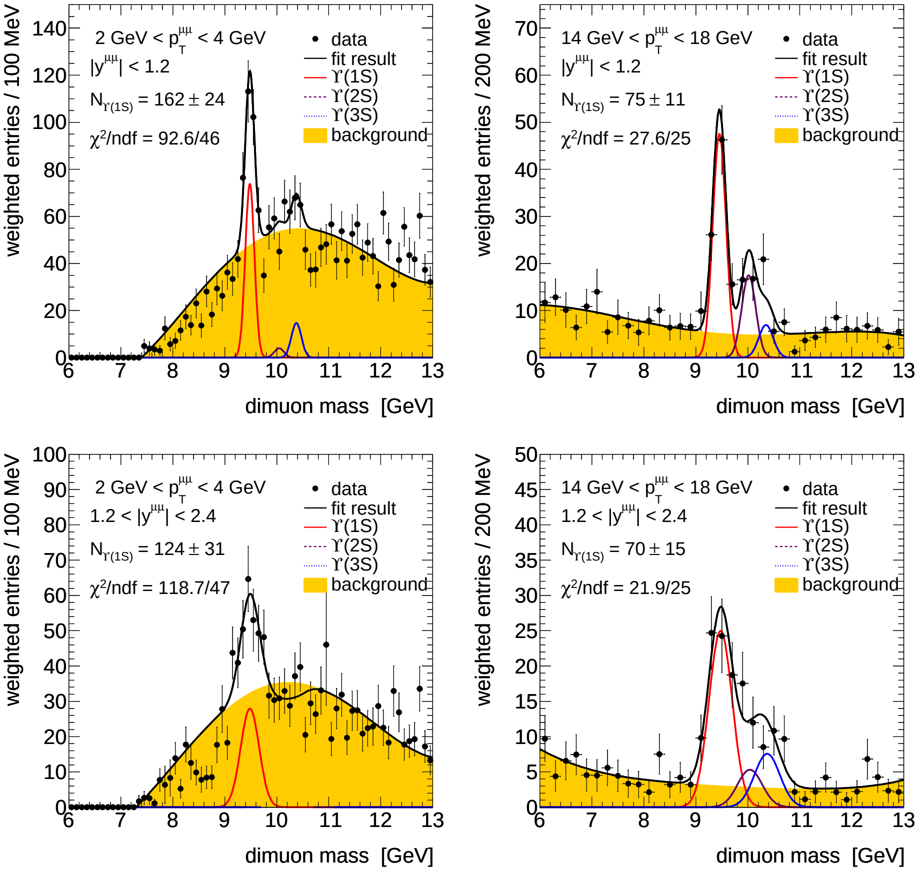

Figure 6.9: Fits results using an alternative - and unsuccessful - model as described in the text. These are included to motivate the background templates in a fashion similar to the historical development and do not enter into the final results.

p_T [GeV]	$\lvert y\rvert$	$N(\Upsilon_{1S})$	$N(\Upsilon_{2S})$	$N(\Upsilon_{3S})$	χ^2	ndf	$\mathcal{P}(\chi^2)$ [%]
0-2	$0.0-1.2$	213±26	66±22	31±21	51.3	46	27.2
2-4	$0.0-1.2$	246±25	60±20	76±22	43.3	52	80.1
4-6	$0.0-1.2$	190±22	55±18	44±19	52.3	64	85.3
6-8	$0.0-1.2$	175±20	53±16	15±13	72.1	66	28.2
8-11	$0.0-1.2$	203±21	32±15	54±17	42.8	31	7.7
11-14	$0.0-1.2$	110±15	33±11	12±10	42.6	31	8.0
14-18	$0.0-1.2$	70±11	27±9	4±7	36.3	31	23.4
18-26	$0.0-1.2$	17±6	18±7	12±6	8.2	10	60.8
0-2	$1.2-2.4$	163±28	34±32	2±30	31.3	47	96.2
2-4	$1.2-2.4$	238±30	-9±38	38±35	55.5	53	38.2
4-6	$1.2-2.4$	124±22	71±9	28±22	53.1	62	78.1
6-8	$1.2-2.4$	129±19	0±19	24±18	84.1	66	6.5
8-11	$1.2-2.4$	132±19	45±22	18±18	36.5	31	22.7
11-14	$1.2-2.4$	75±13	26±15	13±13	18.8	31	95.9
14-18	$1.2-2.4$	71±12	15±13	13±11	37.3	31	20.1
18-26	$1.2-2.4$	28±8	16±10	-6±8	11.1	10	35.0

Table 6.2: Three Υ yields for each bin in $p_T^{\mu^+\mu^-}$ and $y^{\mu^+\mu^-}$. Shown also is the χ^2 of the fit, the number of degrees of freedom (ndf) and the χ^2 probability ($\mathcal{P}(\chi^2)$.

mass - of the ATLAS detector compared with two[6] of the experiments used for the measurements cited above. In those cases ([57, 59]) the $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ are resolved well enough that there is essentially no overlap between the peaks in the reconstructed mass spectrum. In the present analysis, especially in the forward rapidity bins, there is significant overlap. A signal extraction procedure such as the fits shown in Fig. 6.9, in which the detector resolution is allowed to float, gives results for the width parameter that have large fluctuations from bin to bin, shown in Fig. 6.10.

This result strongly suggests the need to take the mass resolution from external sources rather than to fit for it; historically, a similar procedure and a similar observation led to the adoption of the signal templates.

---

[6] The remaining case [58] has a dimuon mass resolution at the mass of the $\Upsilon(1S)$ comparable to ATLAS yet does have one free resolution parameter in their fit. However this analysis also makes use of a much larger dataset.

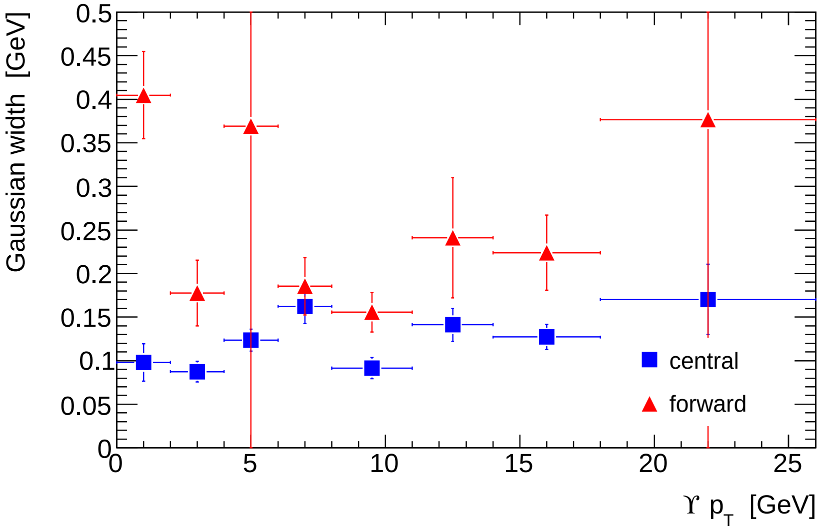

Figure 6.10: *Fitted* (though scarcely believable) dimuon mass resolution at the mass of the $\Upsilon(1S)$ as a function of dimuon p_T and y. The large variations suggest that the model has too many degrees of freedom and was a primary motivating factor leading to the construction of fixed width signal PDFs as described in Sec. 6.1.

Chapter 7

Systematic Uncertainties

Like the extraction of the cross section central values, the determination of the systematic uncertainty of the measurement divides - with two exceptions to be discussed in Sec. 7.3 - into two categories: systematic uncertainties related to the efficiency reweighting, and systematic uncertainties associated with the fits of the mass spectra.

Of all the systematics, those arising from the fit procedure are dominant.

7.1 Efficiency Systematics

The statistical error on the muon reconstruction efficiency directly propagates to a systematic uncertainty on the cross section. If the efficiency was a constant independent of muon charge and kinematics the relationship between the two would be $\delta_\sigma/\sigma = 2\delta_\epsilon/\epsilon$ where σ is the cross section and δ_ϵ is the error on the efficiency. Given an efficiency map with multiple bins one can fluctuate all values high and low by 1σ in order to carry out the error propagation in a straightforward manner. However in this case the error on the efficiency is almost entirely statistical. Because of this, such a procedure would overestimate the error as it assumes perfect correlation between the different map cells.

Instead, the error is propagated by constructing 1000 ‘fluctuated’ efficiency maps. Each cell in a fluctuated map is constructed from the central value map by adding $\delta_\epsilon \times g$ where g is random number that is Gaussian distributed with a width of 1. To each fluctuated map there then corresponds an average weight; the final error is the Gaussian width of the distribution of average weights[1].

Trigger efficiency uncertainties also propagate to the cross section results. The reasoning and procedure for the evaluation of these terms is identical to that described

---

[1] The above procedure is not strictly correct because it treats the full efficiency error, δ_ϵ, as being entirely do to the the statistical component; the systematic component *is* correlated from bin to bin. However for these momentum ranges the statistical error is the dominant source of uncertainty.

above for the muon reconstruction uncertainties. It should be noted however that - again because of the use of a single muon trigger - the relationship between the trigger uncertainties and the cross section uncertainties is not identical to the case of the muon reconstruction uncertainties. Instead of $\delta_\sigma/\sigma = 2\delta_\epsilon/\epsilon$ (in the 'flat' approximation), the dependence is instead given by $\delta_\sigma/\sigma = 2(1-\epsilon)\delta_\epsilon/\epsilon$.

With the IP efficiency systematic, the uncertainty was assigned by taking the difference in yields obtained from using the J/ψ method and the Z method (shown in Fig. 5.24). An alternative would have been understanding the systematics attending a single method, and then propagating those systematics in a manner similar to the trigger and muon reconstruction uncertainties. This was not done in order to be conservative as neither of the two methods can be deemed as robust and unambiguous as the various tag and probe methods.

For each of the 16 Υ bins, all three terms are shown in Fig. 7.1. Although the IP weights themselves are smaller than either the trigger or muon reconstruction weights, the uncertainty associated with these weights is the largest of the three.

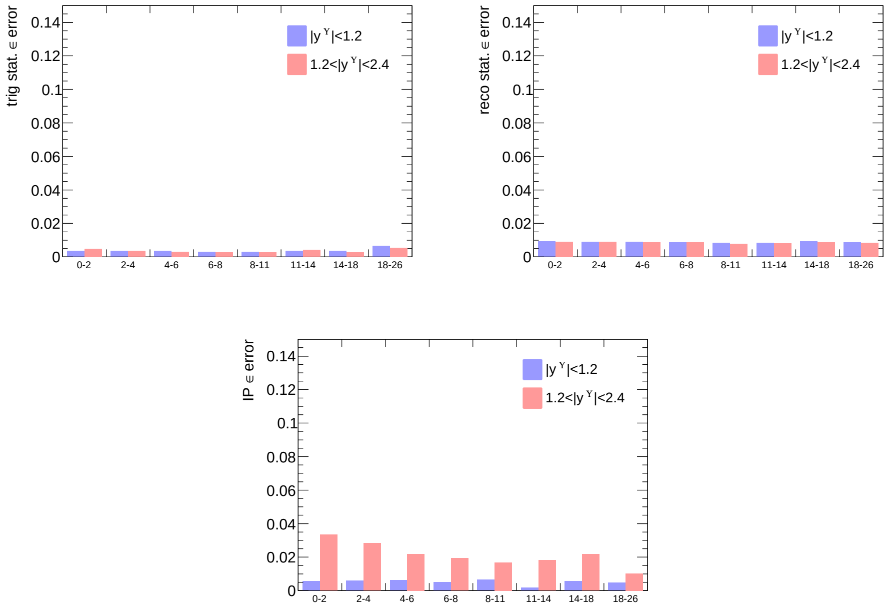

Figure 7.1: The uncertainties due to error on the efficiency maps. Clockwise from top left the three panels correspond to the trigger error, the MS reconstruction error and the ID reconstruction error. The format is identical to the average weight histograms in Figs. 5.29 and 5.30: each pair of bins corresponds one bin in $p_T^{\Upsilon(1S)}$.

7.2 Fit Systematics

This section will discuss systematic uncertainties associated with the fits used to extract $N_{\Upsilon(1S)}$. Five studies were performed, all variants of a common strategy. First, pseudo data is generated *similar* to the procedure described in Sec. 6.4. The fit is then performed in a fashion identical[2] to that used to extract the central values from data, and the $N_{\Upsilon(1S)}$ pull is recorded. The procedure is repeated 1000 times for each bin. Finally a Gaussian fit is made to the distribution of pulls; the product of the Gaussian mean and the statistical error[3] - the average yield bias - is then assigned as the systematic for that particular issue and bin.

The key difference lies in the generation of the pseudo data: rather than adopt the fit model itself when the random sampling is performed, a ‘mismatch’ is introduced between the PDFs entering the fits and the sampled functions. For example, in one of the 5 studies, the signal pseudo data is generated from signal templates that are wider templates than those described in Sec. 6.1 but the fit is performed with the default templates.

The motivation behind such a procedure is to evaluate the size of different effects without resorting to a method that could be influenced by statistical fluctuations in the data i.e. without repeatedly fitting the same dataset. Five effects were studied with this method.

- The possible bias due to mismodeling the background is ascertained using the two alternate set of templates described in Sec. 6.2. Pseudo events from both alternate sets of templates are generated separately and each are fit with the default (OS μ+track) templates; two biases for each bin are thus obtained. The larger of the two is the final uncertainty assigned for the background modeling.

- Uncertainty on the general functional form of the resolution function is evaluated by sampling the signal templates, but then using Crystal Ball functions as the signal PDFs. This is the only case in which the mismatch is introduced in the fit, as supposed to in the generation of pseudo data. Not proceeding in this order would necessitate fixing the parameters that describe the non-Gaussian tail. However to some degree any such choice would be-hoc. The current method avoids this drawback.

- To propagate the uncertainty on the muon momentum resolution, resonance templates are constructed with the smearing parameters increased by 1σ (see Table 6.1.)

- Because the likelihood is not maximized with respect to rigid translations of the PDFs, a constant offset in the reconstructed mass would lead to a bias.

---

[2] One study is an exception to this statement. This is discussed further in the itemized list in this section.

[3] Pull ‘units’ are in statistical error.

The possible size of such an offset has been constrained to less than 25 MeV by studying the position of the K_s^0, J/ψ and Z peaks [39, 73, 66]. Signal events are thus generated according to the default templates used in the fit, but then shifted by 25 MeV.

- Because the fit model is not allowed to vary by an overall change of scale, an incorrectly calibrated mass scale would also cause a bias. The above study is repeated, but with shifts of 0 MeV for the $\Upsilon(1S)$ sample, $0.01 \times \Delta m_{2S-1S}$ for the $\Upsilon(2S)$ sample, and $0.01 \times \Delta m_{3S-1S}$ for the $\Upsilon(3S)$ sample. A value of 1% is considerably more than what is allowed from the K_s^0, J/ψ and Z studies.

The results are shown in Fig. 7.2. Uncertainty on the background model varies from 2%-8%. Changing the signal model produces negligible biases in the central bins. For the forward bins it is an 4%-10% effect. The last three terms, the resolution uncertainties and the two mass scale uncertainties, are generally at or below the 3% level.

7.2.1 A Cross Check of the Fit Systematics

The two largest systematic uncertainties associated with the fit procedure are the background modeling systematic ,evaluated with the alternate templates, and the signal modeling systematic, evaluated by switching to Crystal Ball functions.

As a qualitative check not entering into the final uncertainty, the data itself is refit with the alternate models. That is, the fit systematics are evaluated a second time without use of pseudo experiments. The evaluation of the background uncertainty is 'apples-to-apples:' the fit is performed with both the SS μ+track and MC templates. For each bin the larger of the two yield differences is then taken as the systematic.

Fig. 7.3 shows the direct variation results side by side with the uncertainties obtained from the pseudo experiments. In the central bins, the Crystal Ball function refit generally results in a 2%-4% yield difference. Four examples of the Crystal Ball fits are shown in Fig. 7.4. The analogous results obtained from the pseudo experiments are <1%. This is not interpreted as an inconsistency as the direct variation refits are sensitive to statistical fluctuations in the data.

With this qualification noted, the rest of the direct refits give yield differences that are very similar to the pseudo experiment biases. In particular, these results indicate that the pseudo experiments do not underestimate the fit uncertainty.

7.3 Total Experimental Uncertainty

The final experimental uncertainty is the sum, in quadrature, of the statistical error (shown in Fig. 7.5) and 10 systematics. Eight of these (3+5) are the efficiency and fit systematics described above. The two remaining systematics considered are

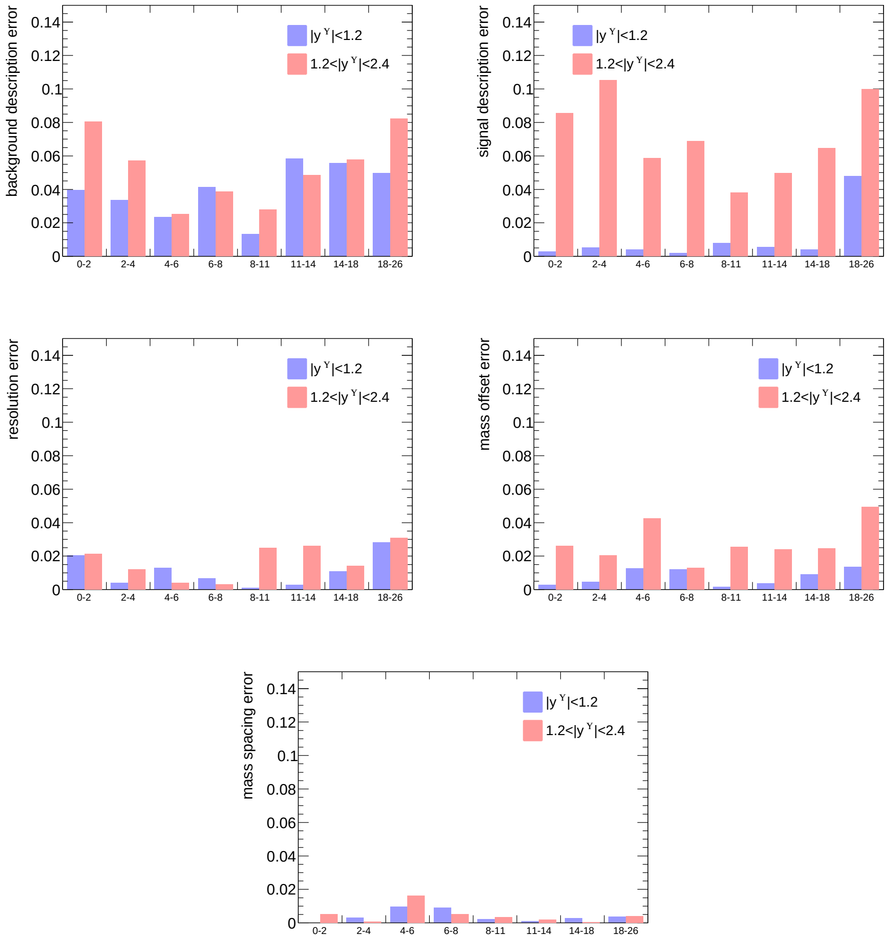

Figure 7.2: The uncertainties associated with the fit procedure. Read left to right from the top the ordering of the five effects corresponds to the listing in the text: background, functional form, resolution, mass offset, mass scale. The first two are dominant while the remaining three are 3% or less.

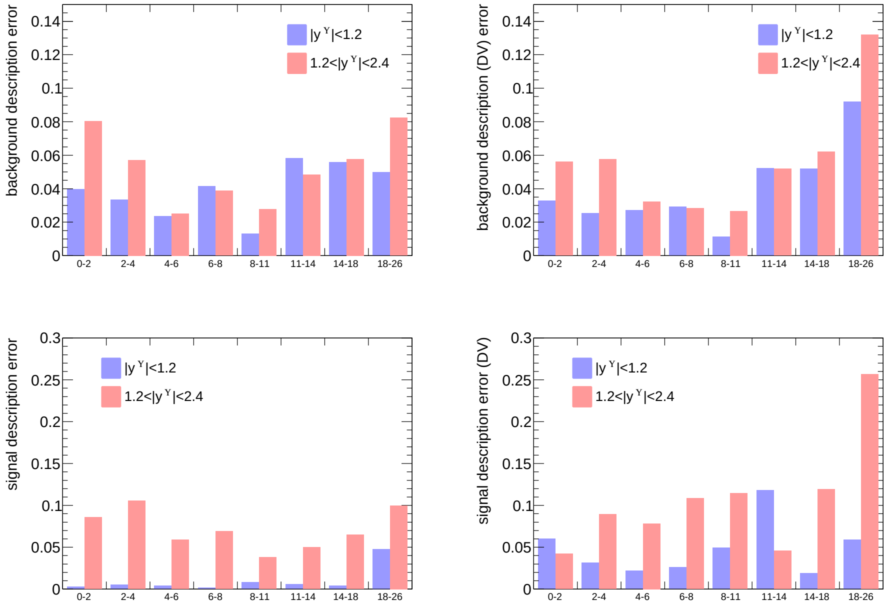

Figure 7.3: The left panels are identical (upto a change of scale) to those in Fig. 7.2 and are shown here for convenience. They are derived from pseudo experiments. The right panels show the corresponding ‘direct variation’ (DV) yield differences: the differences obtained by simply swapping the background templates or signal model and performing the fit on the data a second time.

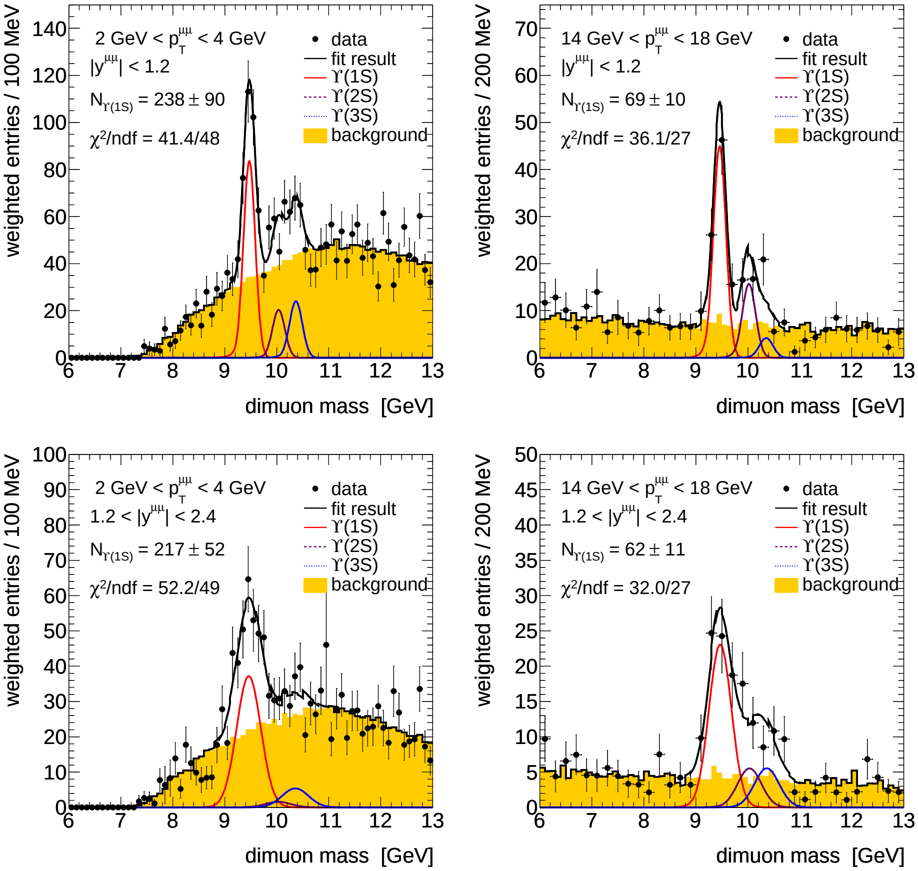

Figure 7.4: Four example bins illustrating the use of using Crystal Ball functions for the signal PDFs. (Compare with Fig. 6.8)

the luminosity uncertainty of 3.4% [61] and a flat figure of 2% for spectral distortions. These arise from muon fiducial migrations, dimuon bin migrations, and FSR radiation (since the final theoretical comparison will be before accounting for FSR.) The figure of 2% was assigned on the basis of Fig. 5.28.

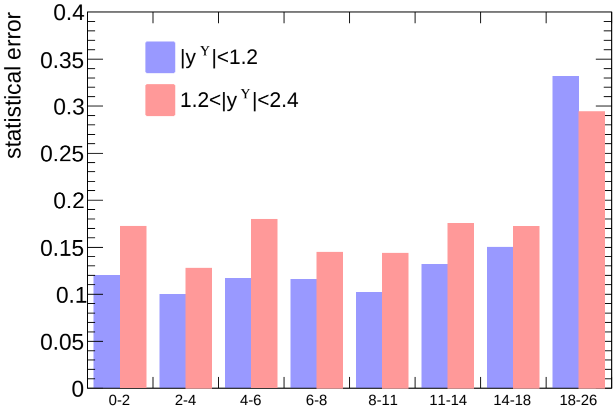

Figure 7.5: The statistical error for each $\Upsilon(1S)$ bin.

Fig. 7.6 shows the total error. With the exception of the two highest p_T^{Υ} bins, which are extremely statistically limited, the forward bin always has a larger uncertainty than the corresponding central bin. The central bin errors are approximately 10%-15% while the forward bin errors are approximately 15%-20%.

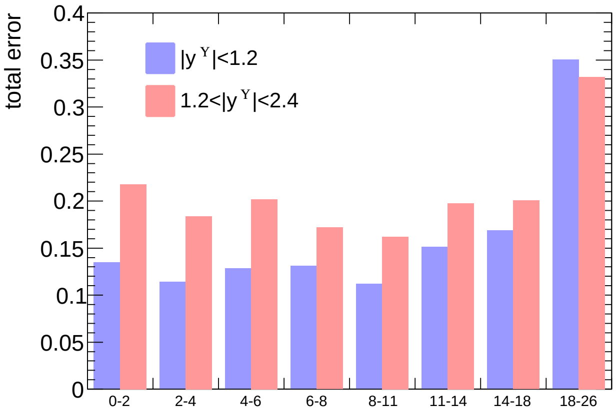

Figure 7.6: The total experimental error for each $\Upsilon(1S)$ bin.

Chapter 8

Results and Conclusions

8.1 Theoretical Comparisons

The results presented in the following section are compared with two theoretical predictions. The first is NRQCD as implemented by PYTHIA8 [74]. A general feature of PYTHIA is that it is is a leading order event generator. In the case of $Q\bar{Q}$ this means that the production amplitude is at tree level only; the effects of loops of virtual particles are not included. As implemented, these processes are also divergent in the limit $p_T^{\Upsilon(1S)} \rightarrow 0$; a reweighting scheme has been utilized to tame these divergences [75].

Specific settings are as follows: two of the long distance matrix elements for the process $(Q\bar{Q})_n \rightarrow \mathcal{Q}$ were altered such that they all of them conform to [76], and the parameters describing the low $p_T^{\Upsilon(1S)}$ reweighting were set to values recommended by [77][1]. The parton distribution functions were switched to CTEQ5L [78]. All remaining settings were those of default PYTHIA8.135.

Forty-six percent of the predicted PYTHIA cross section is due to *feed down*: the decay of a more massive bottomonium particle. In reality such feed down can be due to a variety of processes such as $\Upsilon(2S) \rightarrow \Upsilon(1S)\pi^0\pi^0$ or $\chi_b \rightarrow \Upsilon(1S)\gamma$. As PYTHIA does not model the $\Upsilon(2S)$ or $\Upsilon(3S)$ this prediction accounts only for the latter.

The second prediction is the color singlet model (CSM) at next-to-leading (NLO) order [79]. The CSM is obtained from NRQCD upon taking the limit $v \rightarrow \infty$. As the name implies, $Q\bar{Q}$ pairs produced in a color octet state are no longer included in the sum in Eq. 2.8. This prediction is 'direct only': it does not include any feed down,

---

[1] The syntax to affect these changes is

```
pythia.readString("Bottomonium:OUpsilon3P08 = 0.02")
pythia.readString("Bottomonium:Ochib03P01=0.085")
pythia.readString("PhaseSpace:pTHatMin = 1.0")
pythia.readString("PhaseSpace:pTHatMinDiverge = 0.5")
UserHooks* oniumUserHook = new SuppressSmallPT(1.0, 3, false)
pythia.setUserHooksPtr(oniumUserHook)
```

from χ_b decays or otherwise. However it does include next-to-leading order processes. These include both loops from virtual particles and emission of additional real gluons and light quarks. These additional terms are illustrated in diagrams (b) and (c) of Fig. 8.1. For comparison, diagram (a) is at leading order.

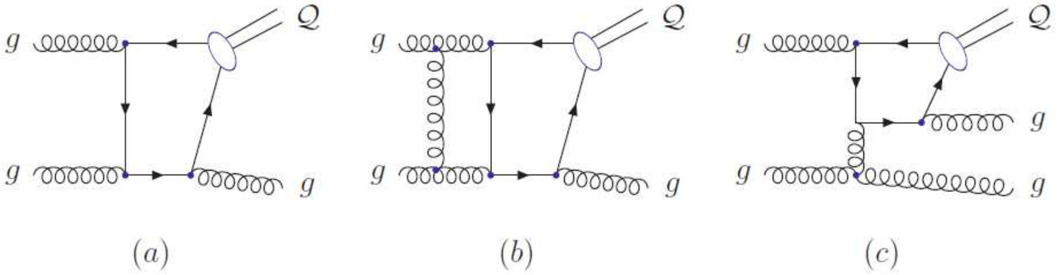

Figure 8.1: Diagram (a) shows ${}^3S_1$ quarkonium production at leading order. (Two gluons have even parity so that it really is leading order even though show it shows a real gluon in the final state.) Diagrams (b) & (c) shows representative diagrams that enter at NLO. The former includes a virtual gluon loop. The latter includes emission of an additional real gluon. Reproduced from [79].

The CSM prediction is implemented with MCFM [80]. The default calculation corresponds to using $\sqrt{4m_b^2 + p_T^{\Upsilon(1S)}}$ with $m_b = 4.75$ GeV for the renormalization and factorization scales. An 'error' on this prediction is assessed by varying both scales between one-half and twice that used used for the default.

8.2 Results and Conclusions

The results are shown in Fig. 8.2 and Table 8.1. The measurement is made in two bins of rapidity, $|y^{\Upsilon(1S)}| < 1.2$ and $1.2 < |y^{\Upsilon(1S)}| < 2.4$, and 8 bins in $p_T^{\Upsilon(1S)}$ extending upto 26 GeV. The differential cross section falls by a factor of approximately 50 over this $p_T^{\Upsilon(1S)}$ range. The dependence on rapidity is weak: the forward (high rapidity) yield averages 86% that of the central.

The CSM prediction underestimates the data everywhere by at least a factor of 2.5 and in some places by more than a factor of 10. This discrepancy can at least partially be attributed to the lack of feed down. Indeed in [81], it was found that both a flat factor of 2 for feed down [82] *and* higher order corrections termed NNLO* were needed to reproduce the cross section in 1.96 proton-antiproton collisions as measured by CDF[2].

The PYTHIA prediction results in a spectrum that is too hard compared with the data: with increasing $p_T^{\Upsilon}(1S)$ it is increasingly in excess of the measured differential

---

[2] This naturally raises the question of why the current results are not compared with a CSM prediction calculated to NNLO*. The answer is that no such prediction was available.

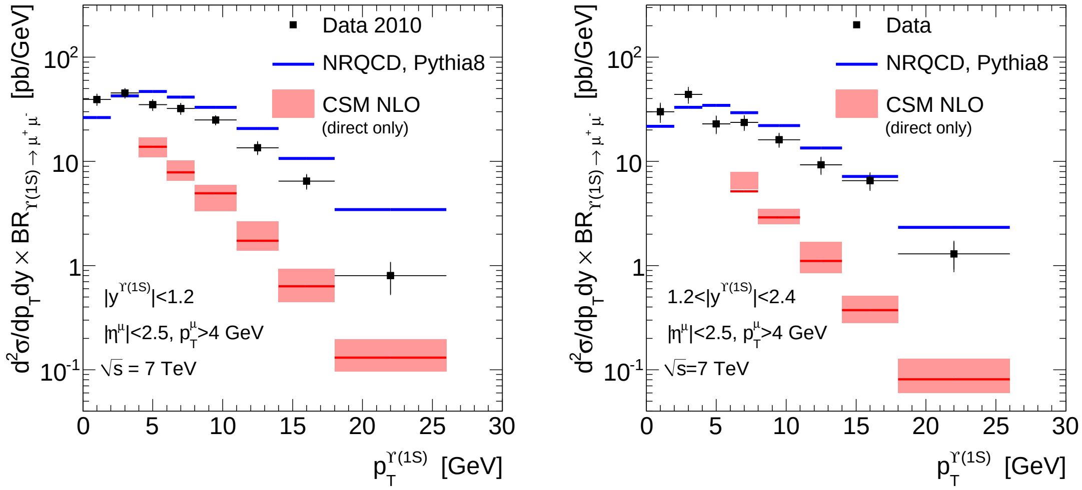

Figure 8.2: Measured differential cross section for the $|y^{\Upsilon(1S)}| < 1.2$ (left) and $1.2 < |y^{\Upsilon(1S)}| < 2.4$ (right) are shown as squares. Also shown are predictions from the PYTHIA8 implementation of NRQCD and the color singlet model at next-to-leading order in lines and filled markers, respectively. The darkened line within each marker shows the default choice of scales; the vertical extent of the marker is the 'error' assessed with the scale variations explained in the text.

cross section. However the normalization at low $p_T^{\Upsilon}(1S)$ is approximately reproduced: with one exception[3], the prediction is within 34% of the data in the 8 lowest $p_T^{\Upsilon}(1S)$ bins.

---

[3] The noted case is also a visible outlier; the bin spanning 4 GeV$< p_T^{\Upsilon}(1S) <$ 6 GeV in the forward rapidity region.

$p_T(\Upsilon(1S))$ [GeV]	$N(\Upsilon(1S))$	$d^2\sigma/dp_Tdy$ [pb/GeV]	δ_{stat} (%)	δ_{syst} (%)	δ_{tot} (%)
$\lvert y(\Upsilon(1S))\rvert < 1.2$					
$0-2$	213	39.3	12	6	13
$2-4$	246	45.4	10	5	11
$4-6$	190	35.0	12	5	13
$6-8$	175	32.2	12	6	13
$8-11$	203	24.9	10	4	11
$11-14$	110	13.6	13	7	15
$14-18$	70	6.5	15	7	17
$18-26$	17	0.80	33	9	34
$1.2 \leq \lvert y(\Upsilon(1S))\rvert < 2.4$					
$0-2$	163	30.0	17	13	22
$2-4$	238	43.8	13	13	18
$4-6$	124	22.9	18	9	20
$6-8$	129	23.7	15	9	17
$8-11$	132	16.2	14	7	16
$11-14$	75	9.3	18	9	20
$14-18$	71	6.5	17	10	20
$18-26$	28	1.3	29	15	33

Table 8.1: Measured differential cross section of $\Upsilon(1S)$ production in the fiducial acceptance ($p_T^\mu > 4$ GeV, $|\eta^\mu| < 2.5$) and the fractional statistical (δ_{stat}), systematic (δ_{syst}) and total (δ_{tot}) uncertainties. Also given is the corrected yield of $\Upsilon(1S)$ mesons, $N(\Upsilon(1S))$, in each kinematic bin.

A Summary for the Layman

This dissertation studies the production of a particle called the $\Upsilon(1S)$ in 7 TeV proton-proton collisions with a scientific instrument called ATLAS. Doing so helps us to understand a bit better the nature of the force that acts between hadrons and their constituents; more fundamental particles called quarks and gluons. The basic structure of this force has been known for several decades, but working out the actual details is in some cases - and this case is certainly one of them - difficult.

First we'll cover all the specialized terms used above. Then we'll explain how the experiment actually works and how the data is analyzed to arrive at the results.

As a point of departure, we'll recall a few facts from elementary high school chemistry: matter is made up of atoms, all of which are composed of electrons orbiting a nucleus composed of protons and neutrons. The reason the electrons stay in orbit is because they're negatively charged and the protons in the nucleus are positively charged, and opposite charges attract. But why do the protons stay stuck together in the nucleus? After all, they're the same charge and like charges should repel each other. So regardless of what the answer is, we know there has to be something else; some new force that keeps atomic nuclei stuck together.

Besides the electron, proton and neutron there are now hundreds, perhaps thousands, of known subatomic particles. A few of these are like the electron: they don't feel the force that keeps nuclei stuck together. But the large majority of subatomic particles do feel this force. They are called hadrons.

Beginning in 1961 and culminating in the early 1970s, a number of experiments and theoretical advances made it clear that hadrons are composed of more fundamental particles called quarks. The force law between quarks was also figured out during this time. Suprisingly, it's in many ways very similar to the laws of electricity and magnetism. One very important difference, however, is that it's non-linear. If one antenna transmits a certain pattern of radio waves that are recieved by your radio as $r_1(time)$ - the notation emphasizes that the 'volume' can be changing with time - and a second antenna transmits another pattern, $r_2(time)$, later in the day, then were they both to have been transmitting at the same time the received pattern would have been $r_1(time) + r_2(time)$. That may sound obvious enough, but it's not true for the quark force law!

This difference is due to the behavoir of the particles that transmit each force. The particle that transmits the electromagntic force is called the photon. (The name arises because it's an indivisible unit of light.) Photons 'see' and bounce off of charged particles, but not each other. The analogous particles in the case of the the quark force law are called gluons, and they do bounce off each other; that is the source of the non-linearity. This force law is called Quantum Chromodynamics (QCD), and essentially because of the non-linearity of QCD it can be very difficult to use to calculate physical quantities.

Momentarily putting aside QCD, we still need to define what we mean by an $\Upsilon(1S)$ particle. Not suprisingly given the above discussion, it is a hadron. In some ways it's a very special hadron. First of all it is composed of a quark orbiting its antimatter partner. Just as a planet can orbit a star very quickly in a close orbit, or slower in a larger orbit, there are different ways in which the same two quarks can combine to form different hadrons. If a quark and its antimatter partner - an antiquark - are orbiting each other in the right way, they can annilihate and briefly turn into a single photon. This is a very convenient property for a particle to have! The photon can rematerialize into an electron and its antimatter partner, or it can rematerialize into a muon - a sort of heavy electron - and its antimatter partner, and both electrons and muons[1] are fairly straighforward to see with our particle detectors. So a particle that can turn into a photon is easy to spot.

Hadrons with this property are technically called vector mesons, though the origin of this term isn't so important. There are vector mesons made up of light quarks, vector mesons made up of charm quarks - which have a 'medium' mass - and vector mesons made up of bottom quarks. These are the heaviest; the top quark, while heavier, is so heavy that it decays before it can begin orbiting anything at all, whether another top quark or a light quark. It doesn't make hadrons. Being heavy is another good thing: some of the challenging details of calculating with QCD go away in the case of heavier quarks.

So now we can put it all together: the $\Upsilon(1S)$ is made from heavy quarks, so it's a physical system that theorists can make predictions about. It can turn into a photon, so it's easy for experimentalists to see it in the laboratory. And finally, it's the lightest example (hence the '1') in which this is all true.

———

The Large Hadron Collider is a tool for colliding protons against protons at the highest energy ever achieved in a controlled way. Collisions of vastly higher energy

---

[1] From now on 'muon' should be taken to mean both true muons and their antimatter partner. Same is true for use of the word 'electron' in this field, but electrons won't come up again.

occur naturally when cosmic rays impinge against the atmosphere, but if you want collisions on demand then the LHC is the premier facility.

Four large detectors are studying what happens in these collisions, though none of them were optimized for the purpose of studying the $\Upsilon(1S)$ and QCD. ATLAS, along with the Compact Muon Solenoid (CMS), was built to carry out a general search for new physical processes at energies from hundreds of GeV to a thousand GeV. (For comparison, the mass of the $\Upsilon(1S)$ is just under 10 GeV, so it's on the low end of the range of things ATLAS was designed to look at.) Observing the direct production of the Higgs particle - Higgs particles, possibly - is of particular interest.

However, that doesn't stop us from using ATLAS to look at the $\Upsilon(1S)$. Specifically, we want to determine how often the $\Upsilon(1S)$ is produced in an LHC collisions. If the accelerator settings are changed to produce more or less collisions per second, then there will obviously be more or less $\Upsilon(1S)$ particles produced in proportion. The cross section (upto a small fudge) *is* this constant of proportionality: (rate of all collisions) × (cross section) = rate of $\Upsilon(1S)$. If that doesn't make sense, here's a second explanation of what the cross section means: some proton collisions produce an $\Upsilon(1S)$; most don't. The fraction of times we *do* get an $\Upsilon(1S)$, one in a million, one in a billion, whatever it is; that's the cross section. (We're jumping the gun a bit here, but depending on precisely[2] how you define the ratio the number is one in 35.9 million.)

So this is the number we want to know, and then we'll compare it with different theoretical predictions to see how good they do. Science!

We get this number in three steps. First, we look for collisions which have two muons in them. This means we really only count the times in which the $\Upsilon(1S)$ turns into a photon which then turns into two muons, but that isn't a limitation because from other experiments we know that that happens to 1/40th of all $\Upsilon(1S)$ particles. One part of ATLAS is built from many layers of steel, so thick that it stops all particles except for muons[3]. This lets us pick out muons from all the other particles; there are detectors behind the steel, and if these detectors see something, then it must be a muon.

In our data sample we had 13,059 events with two muons in them. How do we know how many came from an $\Upsilon(1S)$? We do it by looking at the shape of the *mass distribution.* Everytime you have a collection of particles, you can add up all of their kinetic energy and use $E = mc^2$ to calculate a mass for that physical system. If you do that for all of the 13,059 pairs of muons you get this plot:

---

[2] We've taken the numerator as the number of $\Upsilon(1S)$ decays to $\mu^+\mu^-$ in which both muons are in the fiducial acceptance region discussed in Ch. 4. $\Upsilon(1S)$ events that are easily visible, in other words. For the denominator, we've taken the total inelastic cross section measured in [83] as 69.1 mb.

[3] Muons are very good at going through material, though we don't have time to explain why this is so

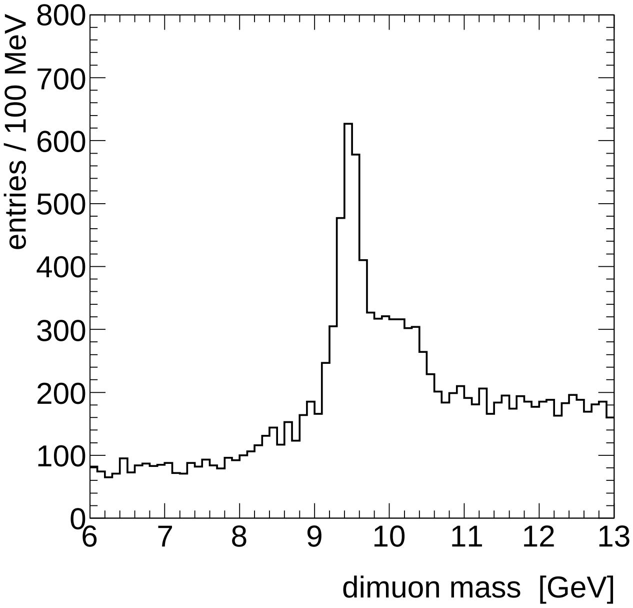

Figure i: Histogram of the dimuon mass distribution. Every system of particles has a unique number associated with it: its mass. Each entry in this plot is the mass of one pair of oppositely charged muons. The central message is that there are more pairs of muons with a mass of 9.5 GeV than any other mass; this is the telltale signature of the $\Upsilon(1S)$.

It shows that there are events with all different masses, but there are lots of events with a mass of about 9.5. That, not coincidentally, is the mass of $\Upsilon(1S)$: there are more muon pairs with that mass than anywhere else because of $\Upsilon(1S)$ decays. The blob on the right hand side of the $\Upsilon(1S)$ peak that stretches to 10.5 is due two other particles just like the $\Upsilon(1S)$, but a bit heavier. Our 'eyes' are too blurry to see them seperately, but we know from other experiments with clearer vision that there are two seperate spikes there.

The second and third steps will probably sound like minor details, but they constitute most of the work of the author for the last year! So we'll briefly mention them. The second step is to figure out how many times we missed one or more of the muons. The detector is highly efficient i.e. it doesn't miss much. But no scientific instrument is perfect and that goes for ATLAS too. There are actually many steps in seeing a muon where it could potentially get lost, and they all have to be tracked down and accounted for. When this is done, and when we calculate how many events with two muons we *would* have had if none were lost, then we get the dotted points on the left of Fig. ii.

The last step is to quantify how many events are in the big spike. When we know this, we'll know how many $\Upsilon(1S)$ particles there were. Then, since we know how many collisions there were in total, we'll know the cross section. This is a fairly intricate process, but the right panel of Fig. ii gives some indication of how it's done. We break the shape of the graph down into 4 seperate elementary pieces: the yellow, red, purple and blue pieces. The yellow represents background processes; there are background events at all masses. The red is how the $\Upsilon(1S)$ looks in the detector while the purple and blue are how the heavier versions of the $\Upsilon(1S)$ look. Then we find the best way to put the four pieces back together to get the original graph. When we've done this, the amount of red that we had to use is the answer. We get 2,174.

* * *

So what was learned? We compared our results to two theoretical calculations, shown in Ch. 8. Both of them are based on QCD but they make different approximations by ignoring different effects. What we found is that neither did a great job of predicting the results but that one of them, the *Pythia* prediction - the name refers to a computer program written by a collabortion of theorists - did a good job in some cases. We've actually split the cross section results up depending on if the $\Upsilon(1S)$ has a little or a lot of momentum, and depending on what angle the particle has compared with the accelerator. If the momentum is small, regardless of the angle, the Pythia prediction is fairly accurate.

Pythia ignores quantum fluctuations when the bottom quarks are first produced, but it doesn't ignore *feed down.* Sometimes two bottom quarks can first come together in a different orbit than the one that corresponds to a $\Upsilon(1S)$. When one of these higher orbit hadrons decays to an $\Upsilon(1S)$ it's called feed down. An example of one of

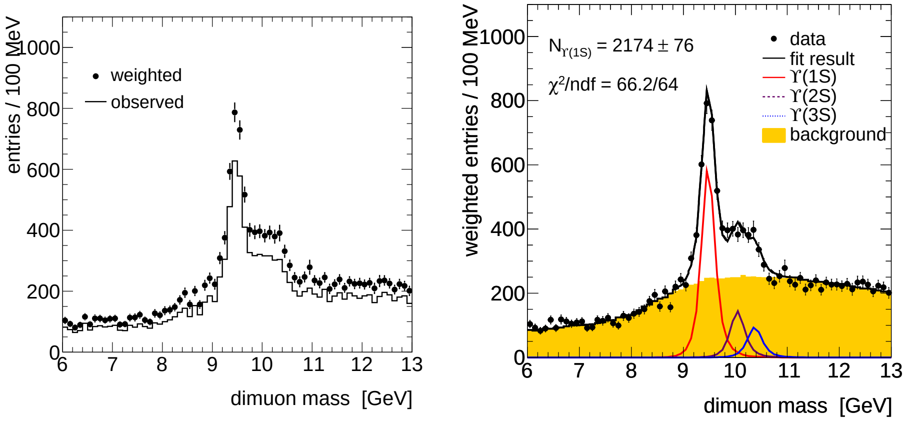

Figure ii: Summary of the data analysis. The plot at left illustrates the efficiency corrections: the solid line represents events with two muons that the detector actually saw while the dotted points represent our calculation of what the detector would have seen if it never missed anything. The plot at right illustrates how the area of the spike is quantified, see text.

these particles is the χ_b. So we can summarize by saying Pythia simulates both of the following processes:

$$protons \rightarrow \Upsilon(1S) \rightarrow two\ muons \tag{8.1}$$

$$protons \rightarrow \chi_b \rightarrow \Upsilon(1S) + other\ things,\ \Upsilon(1S) \rightarrow two\ muons \tag{8.2}$$

The other theoretical prediction - the CSM NLO prediction - does take into account some quantum fluctuations, but it ignores feed down[4]. So it doesn't have the process indicated above by Eq. 8.2. Depending on the momentum, but again not depending so much on the angle, the CSM NLO prediction underestimates the cross section by factors of 2.5-10.

---

[4] For the cognoscenti, it also has no color-octet contributions.

Bibliography

[1] JADE Collaboration, W. Bartel et al., "Observation of Planar Three-jet Events in e^+e^- Annihilation and Evidence for Gluon Bremsstrahlung." *Phys. Lett.* **91B** (1980) 142.

[2] MARK J Collaboration, D. P. Barber et al., "Discovery of Three-jet Events and a Test of Quantum Chromodynamics at PETRA." *Phys. Rev. Lett.* **43** (1979) 830.

[3] PLUTO Collaboration, Ch. Berger et al., "Evidence for Gluon Bremsstrahlung in e^+e^- Annihilation at High Energies." *Phys. Lett.* **86B** (1979) 418.

[4] TASSO Collaboration, R. Brandelik et al., "Evidence for Planar Events in e^+e^- Annihilation at High Energies." *Phys. Lett.* **86B** (1979) 243.

[5] UA-2 Collaboration, M. Banner et al., "Observation of Very Large Transverse Momentum Jets at the CERN pp Collider." *Phys. Lett.* **118B** (1982) 203.

[6] S. W. Herb et al., "Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton Nucleus Collisions." *Phys. Rev. Lett.* **39** (1977) 252.

[7] J.-E. Augustin et al., "Discovery of a narrow resonance in e^+e^- annihilation." *Phys. Rev. Lett.* **33** (1974) 1406.

[8] J. J. Aubert et al., "Experimental observation of a heavy particle J." *Phys. Rev. Lett.* **33** (1974) 1404.

[9] M. L. Perl et al., "Evidence for Anomalous Lepton Production in e^+e^- Annihilation." *Phys. Rev. Lett.* **35** (1975) 1489.

[10] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson." *Phys. Rev. Lett.* **13** (1964) 138.

[11] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interactions." *Progr. Theor. Phys.* **49** (1973) 652.

[12] R. Cahn and G. Goldhaber, *The Experimental Foundations of Particle Physics (2nd ed.)* New York: Cambridge University Press, 2009.

[13] J. D. Jackson, “Lectures on the New Particles” in *Proc. of Summer Institute on Particle Physics*, Stanford, CA, Aug. 2-13, 1976, M. Zipf, ed.

[14] W. Chinowsky, “Properties of Psionic Matter.” *Ann. Rev. Nucl. Sci.* **27** (1977) 393.

[15] G. Goldhaber et al., “Observation in e^+e^- Annihilation of a Narrow State at 1865 Mev/c^2 Decaying to Kπ and K$\pi\pi\pi$.” *Phys. Rev. Lett.* **37** (1976) 255.

[16] K. Berkelman, “Upsilon Spectroscopy at CESR.” *Phys. Rept.* **98** (1983) 145.

[17] S. Behrends et al., “Observation of Exclusive Decay Modes of b Flavored Mesons.” *Phys. Rev. Lett.* **50** (1983) 881.

[18] C.N. Yang and R. Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance.” *Phys. Rev.* **96** (1954) 191.

[19] O. Greenberg, “Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons.” *Phys. Rev. Lett.* **13** (1964) 598.

[20] L.D. Faddeev and V.N. Popov, “Feynman Diagrams for the Yang-Mills Field.” *Phys. Lett.* **B25** (1967) 29.

[21] D.J. Gross and F.Wilczek, “Ultraviolet Behavoir of Non-Abelian Gauge Theories.” *Phys. Rev. Lett.* **30** (1973) 1343.

[22] H.D. Politzer, “Reliable Perturbative Results for Strong Interactions?” *Phys. Rev. Lett.* **30** (1973) 1346.

[23] M. Srednicki. *Quantum Field Theory.* Cambridge, UK: Cambridge University Press, 2007.

[24] The H1 Collaboration, “Measurement of Inclusive Jet Production in Deep-Inelastic Scattering at High Q^2 and Determination of the Strong Coupling.” *Phys. Lett.* **653B** (2007) 134. © Elsevier B.V. Figure reproduced with permission.

[25] D. Gross, “The Discovery of Asymptotic Freedom and the Emergence of QCD.” The Nobel Prize in Physics 2004. The Nobel Foundation. Stockholm. December 8, 2004.

[26] M. Breidenbach et al., “Observed Behavior of Highly Inelastic Electron Proton Scattering.” *Phys. Rev. Lett.* **23** (1969) 935.

[27] S. Bethke, “The 2009 World Average of α_s.” *Eur. Phys. J.* **C64** (2009) 689. http://arxiv.org/abs/0908.1135

[28] R.P. Feynman, “The Behavior of Hadron Collisions at Extreme Energies.” in *High Energy Collisions: Third International Conference, Stony Brook, N.Y., 1970*, Gordon & Breach.

[29] J. Bjorken and E. Paschos, “Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon.” *Phys. Rev.* **5** (1975-1982) 185.

[30] G. Salam, “Elements of QCD for Hadron Colliders.” 2009 European School of High Energy Physics. European Organization for Nuclear Research and the Joint Institute for Nuclear Research. Bautzen, Germany. June 2009. http://arxiv.org/abs/1011.5131

[31] G. T. Bodwin, E. Braaten and G. P. Lepage, “Rigorous QCD Analysis of Inclusive Annihilation and Production of Heavy Quarkonium.” *Phys. Rev. D* **51** (1995) 1125. http://arxiv.org/abs/0704.2599

[32] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider.” *JINST* **3** (2008) S08003.

[33] L. Evans, (ed.), P. Bryant, (ed.), “LHC Machine.” *JINST* **3** (2008) S08001.

[34] B. Richter, “From the Psi to Charm - The Experiments of 1975 and 1976.” The Nobel Prize in Physics 1976. The Nobel Foundation. Stockholm. December 10, 1976. © The Nobel Foundation. Figure reproduced with permission of Nobel media office.

[35] CMS Collaboration. “CMS 3-D image for inclusion in presentations - black background.” April 15, 2011. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=2715. © CMS Collaboration. Figure reproduced with permission of CMS Outreach. (retrieved August 11, 2011)

[36] L. Evans, (ed.), P. Bryant, (ed.), “LHC Machine.” *JINST* **3** (2008) S08001.

[37] “Tevatron Luminosity.” September 2, 2011. http://www.fnal.gov/pub/now/tevlum.html.

[38] ATLAS Collaboration. “ATLAS Experiment - Public Results: Luminosity Public Results.” September 2, 2011. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[39] ATLAS Collaboration. “Charged-Particle Multiplicities in pp Interactions Measured with the ATLAS Detector at the LHC.” *New J.Phys.* **13** (2011) 053033.

[40] ATLAS Collaboration. "Studies of the Performance of the ATLAS Detector Using Cosmic-Ray Muons." *Eur. Phys. J.* **C71** (2011) 1593. http://arxiv.org/abs/1011.6665.

[41] A. Korn. "ID Performance in the 2009 Cosmic Ray Data." ATLAS ID Tracking and Performace Weekly. Geneva. March 12, 2010.

[42] J. Wotschack. "ATLAS Muon Chamber Construction Parameters for CSC, MDT, and RPC Chambers." April 15, 2008. http://cdsweb.cern.ch/record/1099400

[43] D. Adams et al., "Track reconstruction in the ATLAS Muon Spectrometer with MOORE 007," May 28, 2003. http://cdsweb.cern.ch/record/685500.

[44] ATLAS Collaboration. "Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics," September 10, 2008. http://cdsweb.cern.ch/record/1125884.

[45] ATLAS Collaboration, "Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at $\sqrt{s}$=7 TeV," July 28, 2010. http://cdsweb.cern.ch/record/1281363.

[46] T. Cornelissen. *Track Fitting in the ATLAS Experiment.* Ph.D Thesis, University of Amsterdam, 2006.

[47] N. G. Piacquadio, K. Prokofiev, and A. Wildauer, "Primary vertex reconstruction in the ATLAS experiment at LHC," in *Proc. Int. Conf. Computing in High Energy Physics*, Victoria, BC, Canada, Sep. 27, 2007. © IEEE. Figure reproduced with permission.

[48] R. Frühwirth, W. Waltenberger, P. Vanlaer, "Adaptive Vertex Fitting." *J. Phys. G: Nucl. Part. Phys.* **34** (2007) N343.

[49] ATLAS Collaboration, "Performance of Primary Vertex Reconstruction in proton-proton Collisions at $\sqrt{s} = 7$ TeV in the ATLAS Experiment," July 21, 2010. http://cdsweb.cern.ch/record/1281344

[50] ATLAS Collaboration, "ATLAS Detector and Physics Performance Technical Design Report," May 25, 1999. http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html

[51] S. Hassani, et al., "MuTag: New algorithm for Low p_T Muon Identification." ATLAS Flavour Tagging Combined Performance. Geneva. December 8, 2005.

[52] T. Sjöstrand, S. Mrenna and P. Skands, "PYTHIA 6.4 Physics and Manual." *JHEP* **05** (2006) 026.

[53] M. Smizanska, S.P. Baranov, J. Hrivnac and E. Kneringer, "Overview of Monte Carlo Simulations for ATLAS B-physics in the Period 1996-1999," September 29, 2000. http://cdsweb.cern.ch/record/684060.

[54] S. Agostinelli et al., "GEANT4 - A Simulation Toolkit." *Nucl. Inst. & Meth. in Phys. Res. A* **506** (2003) 250.

[55] ATLAS Collaboration, "The ATLAS Simulation Infrastructure." *Eur. Phys. J.* **C70** (2010) 823.

[56] ATLAS Collaboration, "Charged-particle Multiplicities in pp Interactions at $\sqrt{s}$ = 900 GeV Measured with the ATLAS detector at the LHC." *Phys. Lett.* **688B** (2010) 21.

[57] CDF Collaboration, D. Acosta et al., "Υ Production and Polarization in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV." *Phys. Rev. Lett.* **88** (2002) 161802.

[58] D0 Collaboration, V. M. Abazov et al., "Measurement of Inclusive Differential Cross Sections for $\Upsilon(1S)$ Production in $p\bar{p}$ Collisions at $\sqrt{s}$=1.96 TeV" *Phys. Rev. Lett.* **94** (2005) 232001.

[59] CMS Collaboration. "Measurement of the Inclusive Υ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV." *Phys. Rev. D* **83** (2011) 112004.

[60] The ATLAS Collaboration. "Luminosity Determination in pp Collisions at $\sqrt{s}$ = 7 TeV Using the ATLAS Detector." *Eur. Phys. J. C* **71** (2011) 1630.

[61] The ATLAS Collaboration. "Updated Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector." June 15, 2011. http://cdsweb.cern.ch/record/1334563.

[62] The ATLAS Collaboration. "A Measurement of the ATLAS Muon Reconstruction and Trigger Efficiency using J/ψ decays." March 18, 2011. http://cdsweb.cern.ch/record/1336750.

[63] S.C. Hsu. "Tracking Performance for Muons." ATLAS Workshop on Tracking. Ringerberg. January 2011.

[64] H. Evans et al., "Measurement of the $\Upsilon(1S)$ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV." June 12, 2011. (ATLAS internal document ATL-COM-PHYS-2011-170)

[65] B. Heinemann, personal communication. June 2011.

[66] The ATLAS Collaboration. "Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010." June 15, 2011. http://cdsweb.cern.ch/record/1338575.

[67] L. Lyons, W.W.M. Allison, J. Pañellas Comellas. “Maximum Likelihood or Extended Maximum Likelihood? An Example from High Energy Physics.” November 4, 1985. *Nucl. Inst. & Meth. in Phys. Res. A* **245** (1986) 530

[68] J. Orear, Berkeley preprint UCRL-8417 (1958) unpublished.

[69] W. Verkerke and D. Kirkby. “The RooFit Toolkit for Data Modeling.” Computing in High Energy and Nuclear Physics 2003. La Jolla, California. March 2003. http://arxiv.org/pdf/physics/0306116

[70] R. Brun and F. Rademakers. “ROOT - An Object Oriented Data Analysis Framework,” Proceedings AIHENP ’96 Workshop, Lausanne, September 1996, *Nucl. Inst. & Meth. in Phys. Res. A* **389** (1997) 81 See also http://root.cern.ch/

[71] F. James, M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations.” *Comput. Phys. Commun.* **10** (1975) 343

[72] M.J. Oreglia. *A Study of the Reactions* $\psi' \to \psi\gamma\gamma$. Ph.D Thesis, Cornell, 1980.

[73] The ATLAS Collaboration. “J/ψ Performance of the ATLAS Inner Detector.” July 28, 2010. http://cdsweb.cern.ch/record/1281369

[74] T. Sjöstrand, S. Mrenna, P. Skands. “A Brief Introduction to PYTHIA 8.1.” *Comput. Phys. Commun.* **178** (2008) 852.

[75] A.C. Kraan. “Experimental Aspects of Heavy Quarkonium Production at the LHC.” `arXiv:0807.3123 [hep-ex]`. http://arxiv.org/abs/0807.3123

[76] P. Nason et al., “Bottom Production.” `arXiv:0003142 [hep-ph]` http://arxiv.org/abs/hep-ph/0003142

[77] T. Sjöstrand *via* B. Heinemann, personal communication. June 2011.

[78] H.L. Lai et al. (CTEQ Collaboration) “Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions.” *Eur. Phys. J. C* **12** (2000) 375.

[79] J. Campbell, F. Maltoni, and F. Tramontano “QCD Corrections to J/ψ and Υ Production at Hadron Colliders.” *Phys. Rev. Lett.* **98** (2007) 252002. Figure reproduced with permission.

[80] J.M. Campbell and R.K. Ellis. “MCFM for the Tevatron and the LHC.” *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10.

[81] P. Artoisenet et al. “Υ Production at Fermilab Tevatron and LHC Energies.” *Phys. Rev. Lett.* **101** (2008) 152001.

[82] CDF Collaboration, T. Affolder et al., "Production of $\Upsilon(1S)$ Mesons from χ_b Decays in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV." *Phys. Rev. Lett.* **78** (2000) 2094.

[83] L.A. Tompkins. *A Measurement of the proton-proton inelastic scattering cross-section $\sqrt{s}$=7 TeV with the ATLAS Detector at the LHC.* Ph.D Thesis, University of California, Berkeley, 2011.

[84] L. Skinnari, personal communication. June 2011.

Appendix A

The Z Method for Determining the Impact Parameter Resolution Function

This appendix will present an alternative to the J/ψ method (presented in Sec. 5.5) used to determine the impact parameter (IP) signal efficiency. The results are used to ascertain the systematic uncertainty of the size of IP weights (see Sec. 7.1).

The first step is the selection of two opposite sign muons within the range 60 GeV $< M(\mu^+\mu^-) <$ 120 GeV. The muon selection differs from the cross section muon selection in three regards. First, the minium p_T^μ cut is raised from 4 GeV to 20 GeV; though this is not strictly necessary it is still low compared with the characteristic scale of the process ($M(Z^0)/2 \approx$ 45 GeV). Second, both muons are required to be combined. Third, an isolation requirement is imposed on both muons; $p^{cone40}/p_T < 0.2$. This means that when additional tracks with a non-neglible amount of momentum are nearby to the muon, the muon is vetoed. (The motivation is to reject muons coming from the decay products of hadrons in a hadronic jet.)

Because the Z^0 is so massive compared with all the other scales involved in this analysis, there is also no reason to restrict to the EF_mu4 dataset; nearly the entire 40 pb^{-1} collected in 2010 is used.

The next step is simply a binned fit of the Gaussian core of the d_0 distribution in both data and Z^0 MC. This gives the high p_T resolution. After performing these fits it is found that the resolution is about 35% (50%) worse than the simulated resolution in the central (forward) region. Fig. i shows the distributions.

This intrinsic resolution is largely determined by the pixel pitch and the alignment but because most of the muons in this analysis have $p_T \approx 4$ GeV the effect of multiple scattering cannot be ignored. To parametrize the momentum dependance of both the asymptotic resolution and the effect of multiple scattering the following expression is used:

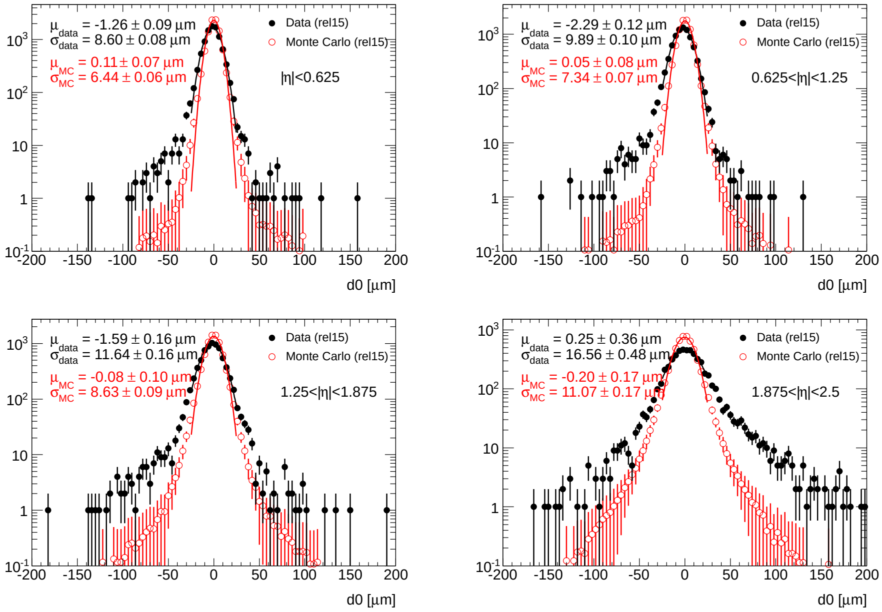

Figure i: The d_0 distribution in $Z^0 \rightarrow \mu^+\mu^-$ events in both data and MC for four different regions in $|\eta|$. The ratio of the widths in data to the widths in MC are used as input parameters in Eqs. A.1. From [64, 84].

$$\sigma_{d0}^{ideal} = \sigma_\infty \sqrt{1 + \left(\frac{p_X}{p_T}\right)^2} \tag{A.1}$$

$$\sigma_{d0}^{Z+mat} = \sigma_\infty \sqrt{(r_{data/MC})^2 + \left(\frac{p_X\sqrt{\delta_{mat}}}{p_T}\right)^2} \tag{A.2}$$

$$\frac{\sigma_{d0}^{Z+mat}}{\sigma_{d0}^{ideal}} = \frac{\sqrt{(r_{data/MC})^2 + \delta_{mat}(p_X/p_T)^2}}{\sqrt{1 + (p_X/p_T)^2}} \tag{A.3}$$

where $r_{data/MC}$ is the ratio of the Gaussian cores from the fits. p_X is the p_T value at which the multiple scattering and the intrinsic contributions are equal in a perfect detector; 14 GeV (20 GeV) in the central (forward) region [32]. The factor of δ_{mat} has been inserted to account for the possible addition of more material and is taken to be 1.1 for $|\eta| < 1.0$ and 1.3 for $1.0 < |\eta| < 2.5$ [39].

The above equations yield a unique expression for the ratio of the resolutions between perfect MC and the resolution observed in data. Motivated by this observation there is no subsequent smearing; the reconstructed d0 in $\Upsilon(1S)$ MC is simply scaled[1]

---

[1] This is not the most physically motivated of decisions. However in the limit that the d_0 resolution

by the expression $\sigma_{d0}^{Z+mat}/\sigma_{d0}^{ideal}$ in order to arrive at the resolution function. The results are shown in Fig. 5.24.

---

function is a perfect Gaussian, smearing and scaling are identical operations. In the present case this is not a bad approximation.