FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

MEASUREMENTS OF THE DIFFERENTIAL CROSS SECTIONS FOR THE INCLUSIVE PRODUCTION OF A PHOTON AND HEAVY FLAVOR JET

By

DANIEL J. DUGGAN

A Dissertation submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2009

The members of the Committee approve the Dissertation of Daniel J. Duggan defended on March 26, 2009.

Horst Wahl
Professor Directing Dissertation

Chris Schatschneider
Outside Committee Member

Todd Adams
Committee Member

Joseph Owens III
Committee Member

Efstratios Manousakis
Committee Member

Approved:

Mark Riley, Chair
Department of Physics

Joseph Travis, Dean, College of Arts and Sciences

The Graduate school has verified and approved the above named committee members.

ACKNOWLEDGEMENTS

This paper would never have been possible without the support, help, and friendship of many, many people. First, to the members of my defense committee, especially the non-high energy physicists Chris Schatschneider and Stratos Manousakis, for wading through a very large dissertation and for their valuable critique and help. To all of my Florida State University crew, George, Scott, Pat, Rob, Takahisa, Mica and countless others. These guys are awesome, and a guy could be so lucky to be in their company. All of my highly energetic friends, Paulo, Jadranka, Norm, Silke, Fernando, Azar, Jackson, Jose, Sinjini, Daekwang, Haryo, Edgar, Trang, who have all helped shape my FSU HEP horizons. A special thanks to Fernando, who is truly a genius of quantum field theory. I owe a large amount of gratitude to all of the High Energy Physics faculty at Florida State, Harrison Prosper, Howie Baer, Laura Reina, Susan Blessing, Todd Adams, Yuri Gershstein and both Vasken and Sharon Hagopian for taking the time to listen and provide valuable insights into this process. I may not have taken all of their advice, but when I didn't I sure learned it the hard way eventually. Jeff Owens, for both providing such an interesting research topic to pursue and to foster the communication between theory and experiment that is too often lacking in the field. Tvzetalina, who was somehow able to manage the vast amounts of stress I piled on her throughout our time together. I can't think of Tallahassee without thinking of Casey, who has been like a little brother to me and there for me every step of the way, and always at the airport to pick me up during my numerous trips back to campus and to provide some good philosophical debate.

I would be remiss if I did not thank the Fermilab community for their helping hands and making the lab such a welcoming place. Jimmy Degenhardt, my very good friend who kept me impossibly sane throughout my time at Fermilab, even when worlds away. My comrades in arms Angela, Mike, and Satish, who not only helped me adjust to lab life, but who made it so feel so much like home. Outside of the lab, I had an amazing support crew: Doctor Rob,

who made Thursday nights the best night of the week for me and countless others. Marshall, my roommate who indulged all of my crazy ideas, as well as countless other friends whose company has been nothing short of stellar.

A large part of my time at DØ was spent in the control room, and my expertise really came to fruition thanks to Dean Schamberger and George Ginther, who pushed me when I needed it and gave me opportunities to make a difference. My postdocs, Oleksiy, Dmitry, and Jochen, taught me not only how to do analysis at DØ, but how to do it right. It's hard to believe that when I showed up at Fermilab, I barely knew how to compile code and what a calorimeter exactly *does*, and for every question or problem, Andrew Askew had a clear and correct response. Not only providing sage wisdom beyond his years, but failing that, providing enough homebrew to bring the wisdom into focus. I cannot begin to thank him enough. Throughout my time at Florida State, I have often strayed into different projects and gotten myself into more trouble than I care to remember, but this whole adventure in particle physics would not have been possible without the support of my advisor Horst. He made sure that I had the opportunities to excel in my career, providing wisdom and guidance every step of the way. In doing so is to thank for the physicist I have become, and I am forever indebted.

Most of all, this thesis is dedicated to my family, who has gone to great and amazing lengths to support me and my crazy career. Honey has brightened my days and helped me finish this long road, with care packages, frequent visits and amazing constant support. My cousin Josh provided ample distractions, doses of excitement and adventure, and even Thanksgiving swamp-filled fun. My sisters and my brother have always given whatever they could to help me out, including much needed breaks. They've flown me out, driven in, abducted me for vacations (one involving a narrow escape in a ball pit) and generally have been my backbone in these last five years and change. You all mean the world to me. And finally, to my parents, who always challenged me to live up to my potential and to excel in something I love. They have always put me before themselves, and have sacrificed immensely to give me these amazing opportunities.

I have been very lucky in my life to be surrounded by such wonderful people, and to be given the chance to live out my dreams.

— Dan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis presents the first measurement of the differential production cross section of a heavy flavor (bottom or charm) jet and direct photon at the Fermilab Tevatron. These measurements were performed using data recorded with the DØ detector from proton-antiproton collisions at a center of mass energy of $\sqrt{s} = 1.96$ TeV. These results probe a kinematic range for the photon transverse momentum of $30 < p_T^{\gamma} < 150$ GeV and rapidity of $|y^{\gamma}| < 1.0$ and for jet transverse momentum $p_T^{jet} > 15$ GeV and rapidity of $|y^{jet}| < 0.8$. These results are compared to next-to-leading-order theoretical calculations.

CHAPTER 1

INTRODUCTION

The most fundamental questions of existence have endured throughout the history of man, “Why am I here?”, and “What is *here*?” Though these questions have lingered for generations upon generations, considerable work has been produced in a large variety of fields to provide powerful arguments and robust responses. The first of these two questions certainly contains an air of philosophy, a melding of self purpose with the natural instinct of survival, not to mention chords that strike the very core of both faith and religion. It has become the subject of many experts, both poets and priests have devoted their lives to its understanding.

However, the second of these two can be much simpler to quantify. This basic inquiry has driven our understanding of the world from the largest galaxies to the smallest of particles. It is an adventure that is at the heart of progress, that has fostered the flame of invention and revolutionized past and present cultures. What is more, this question naturally sparks yet another seemingly innocuous, but powerful query, “How does *here* work?”

The questions still remain, however, and it seems that the more we learn, the more dark alleys and unknowns seem to appear. But it is the nature of science to systematically and methodically prick and prod these mysteries to gauge their depths and to reveal their secrets. In every facet, the intricacies of the universe are unraveled one measurement at a time, providing valuable clues that provide a gateway into a deeper understanding of this magnificent and beautifully complex world.

Particle physics takes this mantra of exploration and understanding and forges ahead in the understanding of the smallest of these puzzle pieces, the particles that form the foundation of everything. Although small in size, these clues can be paramount in resolving questions as large as galaxies, or as small as the pieces of the proton.

1.1 Particle Physics

High Energy Particle Physics is the branch of physics that attempts to elucidate the properties of the smallest constituents of matter. This is important not only to understand the particles themselves, but also to understand how they interact, providing a window to the beginning and early development of the universe. By studying particles we gain insight into this quantum world: its properties, its laws and its consequences for other areas of physics as well as for the universe. In the last forty years there has been a revolution in the understanding in this field with the emergence of a picture generally referred to as the “Standard Model” of particle physics. This theoretical description has been remarkably successful and powerful. Through it, three of the four known forces (electromagnetism, weak, and strong) are described, as well as the modeling of protons, neutrons and the multitude of existing composite particles. The Standard Model was used to predict the existence and masses of the force mediating particles of the weak force, and continues to be the guide in search for other Standard Model particles.

The major experimental goal in particle physics is to study particle collisions at the highest possible energies, with the aim to subject the Standard Model to further scrutiny by testing its predictions with measurements of increasing precision. Many sophisticated, high precision experiments have been performed to test the Standard Model with the hope of finding discrepancies, and it has withstood all attempts of being invalidated. On the other hand, certain questions still remain, such as the origin of particle masses and the structure they appear to assume, and the physical interpretation of the gauge groups used within it. Thus, the Standard Model in its current form can only be an approximation (albeit a very good one) of a more general theory, which itself is an “extension” of the Standard Model.

The Standard Model (SM) encompasses the known fundamental particles and the forces that govern their interactions (with the exception of gravity). There are two types of particles in the SM: fermions and bosons. Fermions are spin 1/2 particles while bosons have integer spin. Fermions are separated into three generations. Each incremental generation has particles that are more massive than its predecessor. Fermions are further divided up into two classifications: quarks and leptons (Table 1.1). The forces included in the SM are the electromagnetic force, the weak force, and the strong force and are mediated by bosons (Table 1.2). Charged particles interact electromagnetically through the exchange of photons.

Leptons will weakly interact by the exchange of $W^{\pm}$ and Z^0 bosons[1]. Quarks are the only fermions that interact through the strong force, which the gluon mediates. All six flavors of quarks are charged, thus they interact through all of the forces the SM describes, which increases the challenge to study them.

Quarks and gluons are the only fundamental particles that interact via the strong force. The study of these interactions is known as Quantum Chromodynamics (QCD). To better understand QCD, it is necessary to study the properties and interactions of both quarks and gluons. There are six types ("flavors") of quarks (Table 1.3). The three heaviest quarks are charm (mass of $\sim$1.3 GeV/c^2), bottom ($\sim$4.3 GeV/c^2), and top ($\sim$172 GeV/c^2) [1]. Quarks are the fundamental constituents of hadrons. Hadrons are divided into two groups, mesons and baryons. A meson is the bound state of a quark and anti-quark pair and a baryon is the bound state of three quarks. For instance, a proton is the bound state of two up quarks and one down quark. Quarks are never seen alone in nature because of a phenomenon known as color confinement. Color is an attribute of a quark and a gluon that is somewhat analogous to charge in electromagnetism. So far only colorless particles have ever been observed. Mesons contain a color-anticolor pair, whereas a baryon will have one of each color[2] (for antiparticles they will contain one of each anticolor). If the bound state of a hadron is disrupted and quarks are emitted, quark-antiquark pairs are produced that couple to quarks to create mesons in order to maintain color confinement.

According to the picture that has emerged from experimental and theoretical efforts during the last forty years, when protons (or other hadrons) are probed in scattering processes, they behave as if they were made of "partons" (quarks and gluons), each carrying a fraction of the proton's momentum. The division of a hadron's momentum among its partons is described by parton distribution functions (PDFs). A large fraction of the momentum is carried by valence quarks[3], but in addition there are also gluons and "sea quarks" (quark-antiquark pairs originating from intermediate gluons). Since these distribution functions cannot be calculated from perturbative QCD alone, numerical techniques using data are used to aid in the determination of functional fits for these distributions.

Because protons are really a distribution of partons, when they hit each other it is really

---

[1] Charged leptons and the $W^{\pm}$ will also exchange photons from the electromagnetic interaction.

[2] The presence of all three of these colors in the baryon will yield a colorless state.

[3] These quarks carry the hadron's quantum number and thus identify the hadron.

Table 1.1: Summary of Standard Model fermions.

Quarks	u d	c s	t b
Leptons	ν_e e^-	ν_μ μ^-	ν_τ τ^-

Table 1.2: Summary of Standard Model gauge bosons.

Gauge Boson	Interaction	Charge (Q/e)	Mass (GeV/c^2)
Photon (γ)	EM	0	0
W	Weak	±1	80.4
Z	Weak	0	91.2
Gluon (g)	Strong	0	0

the partons that interact. When the momentum transfer in the parton-parton interaction is large ("hard scattering"), the recoiling partons from the collision may have large transverse momentum [4] The location of the partons' interaction is commonly referred to as the primary interaction vertex, or simply the primary vertex. When the partons try to separate after the collision, confinement makes it energetically favorable for new quark-antiquark pairs to be created out of the vacuum. This is because the strong interaction between quarks, antiquarks, and gluons has the feature that, within the short range of the interaction, the interaction strength increases with distance. Thus a chain of mesons is formed, with each meson having less energy than the previous one. The chain ends when the remaining energy of the final quark is not sufficient for another pair creation. Since the momentum components orthogonal to the direction of the original parton tend to be small, the cloud of newly created hadrons appears collimated around the direction of the hard-scattered parton. This appears as a "jet" of particles. This process of turning quarks and gluons into hadrons, called "hadronization" or "fragmentation", is modeled from experimental data. Due to singularities that arise in the theory, it is not possible to calculate hadronization from first principles; instead, experimental observations are used to tune the hadronization models.

---

[4] Transverse momentum is the momentum component perpendicular to the direction of the incident particles.

Table 1.3: Standard Model quark properties.

Quark Flavor	Charge (Q/e)	Mass(GeV/c^2)
Up (u)	+2/3	∼.0025
Down (d)	−1/3	∼.004
Charm (c)	+2/3	∼1.3
Strange (s)	−1/3	∼0.110
Top (t)	+2/3	∼172
Bottom (b)	−1/3	∼4.3

The direct photon itself serves as a natural probe of the hard scattering process at the primary vertex as it does not suffer from ambiguities due to jet identification and fragmentation processes. Thus its energy can be measured with higher precision, and that energy is closely related to that of the original interaction.

Studying interactions resulting in a charm or bottom flavor quark and an associated photon is important, as information can be obtained to further validate perturbative techniques used in theory, as well as to obtain a much deeper understanding of the charm or the bottom quark dynamics in these events. Much work has been done from the theoretical side to address these types of interactions (Refs. [2]-[11]), and experimental measurements are important to constrain and test them.

Previous measurements have been performed that probe these final states (Refs. [12]-[19]), which include measurements from the Tevatron, LEP, and deep inelastic scattering experiments. However, there are currently no published results for the differential production of photon and heavy flavor jets at energies relevant for the Tevatron. Additionally, there are related measurements previously performed at DØ [20, 21] to which this analysis provides a natural extension.

Finally, the understanding of photon and heavy flavor jet production is not only important for physics research at the Tevatron, but also for the physics to be studied at the Large Hadron Collider (LHC)[5]. The better these quantities are understood, the more confidence can be placed in these signals for new measurements and for searches for new physics[6].

---

[5] The LHC has an estimated turn-on date of October 2009

[6] New physics is considered to be any particle or interaction that either violates the SM or is not described by it.

CHAPTER 2

THEORY

The main goal in performing these measurements is to provide results that can be compared to theoretical predictions. These comparisons serve to further validate the Standard Model, and hopefully to provide stronger constraints for charm and bottom parton distribution functions. Theoretical predictions come in the form of a calculated hadronic cross section of a photon and either a charm or bottom quark. The cross section calculation can be separated due to the factorization theorems [22] into two parts: one that describes the force at short distances, and the other that describes the force at large distances. The former part can be calculated perturbatively and the latter part accounts for non-perturbative effects.

The partonic cross section is the part of the hadronic cross section that is perturbatively calculable, and can be written as a series in terms of its force couplings. The interactions that give rise to the photon plus heavy flavor quark final state are governed solely by two forces, the strong force and the electromagnetic force. Because of the dominance of the strong force over the electromagnetic force, it is advantageous to write this series in terms of α_s, the strong coupling constant. This series takes on the form shown in Equation 2.1 for the final state of a direct photon and a heavy flavor quark jet.

$$\hat{\sigma} = \sigma^{LO}\alpha_{EM}\alpha_s + \sigma^{NLO}\alpha_{EM}\alpha_s^2 + ... \tag{2.1}$$

In this equation, α_{EM} is the electromagnetic coupling constant and σ is the partonic cross section at both leading order (LO) and next-to-leading order (NLO) with the coupling terms factored out. It is important to note that $\hat{\sigma}$ does contain ultraviolet, soft, and collinear divergences that need to be addressed in order to arrive at a finite result. Ultraviolet singularities, which arise when a particle's momentum becomes infinite in a loop diagram, are absorbed into α_s through renormalization. Thus, α_s becomes dependent on the characteristic energy of the interaction (Q), and the one-loop expression is defined as

$$\alpha_s(Q^2) = \frac{2\pi}{b_0 \ln(Q^2/\Lambda^2)}, \tag{2.2}$$

where b_0 is a constant and Λ^2 is the effective mass scale.

Soft divergences occur when the gluon's energy tends toward zero, and these are canceled with the lower bounds of the exposed poles of the loop integral through dimensional regularization. Collinear singularities arise when the angle between two massless particles approaches zero. These divergences are absorbed into PDFs and fragmentation functions (FFs), which are measured experimentally as the dependence on a parton's momentum with respect to its parent hadrons cannot be perturbatively calculated.

The non-perturbative part of the theoretical hadronic cross section consists of both initial state effects, which are handled using PDFs, and final state effects, handled using FFs. Because free quarks have not been observed, their properties must be understood within the context of their hadrons. PDFs describe the incident distribution of these constituent quarks and gluons, but require experimental information in order to be calculated. The experimental input required is the fractional momenta of quarks with respect to their parent hadrons (x) and the energy scale at which these measurements were made (Q^0). This information is used to set the boundary conditions for the DGLAP equations [23], which are differential equations whose solutions describe the evolution of these quark and gluon parameterizations as a function of Q^2. Analogous techniques are used to derive parameterizations for fragmentation functions, and in these cases the hadronized quarks and gluon distributions are derived.

Calculations that provide a complete treatment of all subprocesses through next-to-leading order have been performed [24], and numerical techniques [25] to solve the DGLAP equations were used. Parameterizations of the PDFs were taken from the CTEQ collaboration [26]. To fully appreciate these results, the processes that comprise the final state of the photon plus heavy flavor quark are reviewed in more detail.

At leading order there is only one subprocess that contributes to this final state, Compton-like scattering:

$$\mathrm{Q} + g \rightarrow \mathrm{Q} + \gamma, \tag{2.3}$$

where Q represents either a charm or bottom quark, and g is the gluon. This subprocess contributes from two experimentally indistinguishable channels, shown in Fig. 2.1, and

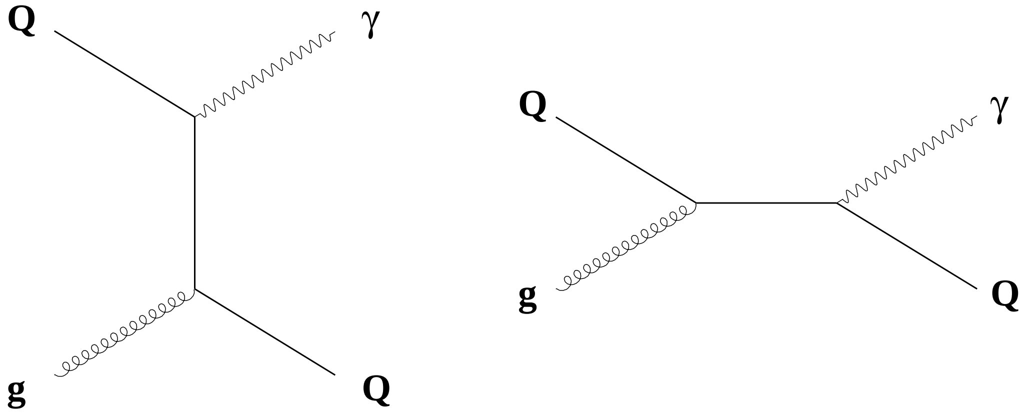


Figure 2.1: Both the t channel (left) and s channel (right) contributions to the photon plus heavy flavor quark cross section coming from the Compton-like scattering subprocess of an incident heavy flavor quark (Q) and gluon (g).

accounts for up to $\sim$ 95% of the cross section at a photon transverse momentum of $p_T^\gamma \sim 20$ GeV. The incident heavy flavor quark is assumed to be radiatively produced from a gluon splitting into a quark-antiquark pair, and the most commonly used theoretical parameterizations do not account for non-zero contributions of either charm or bottom quarks. Certain theories predict an intrinsic, or non-perturbative charm component, to the proton [27]. The cross sections are potentially sensitive to these contributions, providing an interesting point in these comparisons.

There are additional contributions to the cross section other than Compton scattering, and they must also be addressed. Most notably, subprocesses where a gluon or quark fragments into a photon will contribute to the cross sections at all orders of α_s. Fig. 2.2 shows the leading order diagrams that come from fragmentation. Since in the fragmentation processes the photon is emitted in close proximity of other partons, the experimental isolation requirements suppress these contributions. Because of this, only the components not arising from fragmentation processes are emphasized.

The next step is to calculate the next-to-leading order corrections to the cross section. At energies relevant for the Tevatron (1.96 TeV), one next-to-leading order subprocess mainly contributes to the photon plus heavy flavor jet final state, the quark-antiquark annihilation

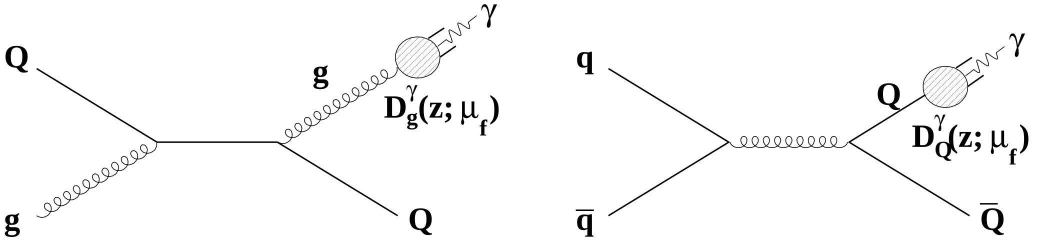


Figure 2.2: Examples of subprocesses where the gluon (g) or heavy flavor quark (Q) fragments into a photon. The contribution from these diagrams is largely suppressed due to photon isolation requirements.

diagrams:

$$q + \bar{q} \rightarrow \gamma + \mathrm{Q}\bar{\mathrm{Q}}, \tag{2.4}$$

where the heavy flavor quarks in the final state (Q and $\bar{\mathrm{Q}}$) come from gluon splitting, as seen in Fig. 2.3. The large fraction of events coming from this interaction is due to the proton-antiproton collisions at the Tevatron and the large abundance of valance antiquarks from the incident antiprotons. The contribution from this subprocess to the overall cross section rises as p_T^γ increases. From this subprocess, there is a direct sensitivity to the fraction of heavy flavor quarks coming from gluon splitting. Previous measurements have been performed measuring these fractions, such as results from LEP [19]. Although agreement with respect to b quarks is seen between data and theory, excesses in data on the order of 1.5–2 times the theoretical predictions were measured for the c quark case.

We now have a general understanding of not only the theoretical processes that go into this cross section, but also a basis for how the theory itself can be calculated in such a way to be directly comparable to the experiment measurement. With this in hand, we can now begin to focus on the experimental foundations, techniques, and tools that guide the measurement itself.

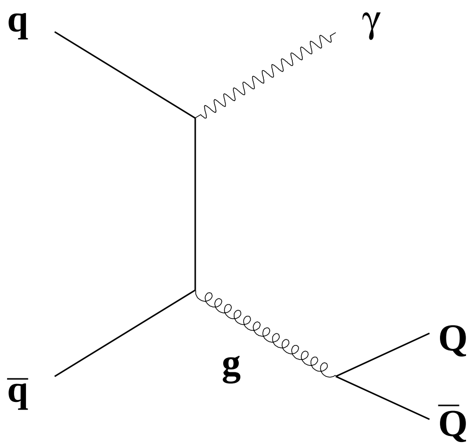


Figure 2.3: The quark-antiquark annihilation process where the heavy flavor quarks (Q and $\bar{Q}$) in the final state arise from gluon splitting.

CHAPTER 3

EXPERIMENTAL APPARATUS

The data collected by the DØ detector originate from proton-antiproton ($p\bar{p}$) collisions. Protons and antiprotons travel around the roughly four mile Tevatron ring in about 21 microseconds and collide at a $\sqrt{s} = 1.96$ TeV center-of-mass energy.

Both protons and antiprotons travel in orbits within the ring designed such that they meet only twice upon completing an entire circuit, at BØ and DØ, which are home to the two multipurpose detectors at Fermilab. Three superbunches of particles travel around the ring, with a spacing of 2.6 microseconds in between them. Inside each superbunch, there are twelve individual bunches, separated by 396 nanoseconds. Each bunch initially contains up to ~ 300 billion protons and ~ 100 billion antiprotons, where the limiting factor is the number of antiprotons available for use. Collisions occur when these bunches meet at their crossing points along the ring and the individual protons and antiprotons interact.

The measure of the rate of these interactions, called luminosity ($\mathcal{L}$), is based on how many protons and antiprotons are contained in a bunch at any given time. Luminosity is calculated using the following equation:

$$\mathcal{L} = \frac{N}{\sigma}, \tag{3.1}$$

where N is the rate of collisions and σ is the effective cross section for $p\bar{p}$ interactions. Luminosities of greater than 300×10^{30} cm^{-2}s^{-1} have been recorded at DØ. However, as more and more collisions occur, the total number of protons and antiprotons in orbit decreases over time. This in turn results in a steady decline of the luminosity.

The period of time from the point when collisions begin to when the beams are taken out of the Tevatron is referred to as a store. Stores typically last 24 hours, and on average there is a two to three hour window from the end of one store to the beginning of the next.

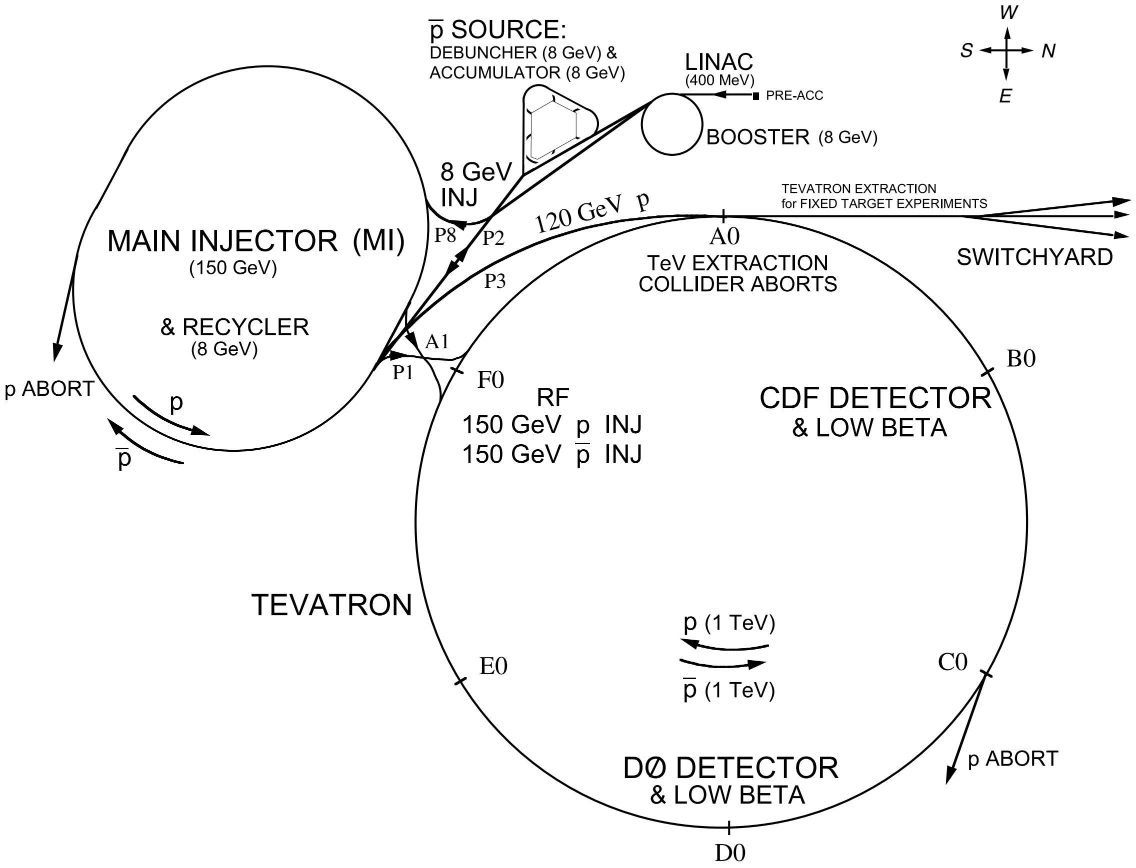


Figure 3.1: The Tevatron accelerator chain.

Integrated luminosity is the measure of how much luminosity was present in a given time period, which provides a metric for the amount of data available for study from the detector.

3.1 The Accelerator Chain

Before investigating the data from $p\bar{p}$ interactions that the DØ detector has collected, it is important to first gain an appreciation for the mechanisms of the accelerator chain (Fig. 3.1) that make these collisions possible.

The Cockcroft-Walton is the first stage of the acceleration process. This machine takes negative hydrogen ions at essentially at rest and accelerates them to an energy of approximately 750 keV. This is performed using a combination of diodes and capacitors to increase the voltage in stages. These negative ions, consisting of one proton and two electrons, are then sent to the Linac.

The Linac, a linear accelerator, is the second step on the proton's tour. At a length of five hundred feet, this accelerator increases the proton's energy from 750 keV to 400 MeV, a change greater than a factor of 500. Oscillating fields are applied using cylindrical electrodes in such a way that the ions always are accelerated when they cross the gap between the electrodes. At the end of the Linac, these ions reach energies of 400 MeV and are sent through a carbon foil that strips them of their electrons, leaving only the proton to continue to the Booster.

The Booster is a rapid-cycling synchrotron that increases the protons' energies from 400 MeV to 8 GeV. A synchrotron is an accelerator with a fixed orbital radius and variable strength magnets separated into dipoles and quadrupoles. These magnets utilize the basic principle of the Lorentz force equation, $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$, where the force acting on a charged object ($\vec{F}$) is proportional to its charge (q) and is dependent on the present electric ($\vec{E}$) and magnetic ($\vec{B}$) fields, as well as the object's velocity ($\vec{v}$). Using this relationship, the quadrupoles and dipoles serve to focus and curve the beam throughout its orbit, respectively. Once the protons' energies reach 8 GeV, they are ejected from the booster. From here protons head to the next destination along the accelerator chain, the Main Injector.

The Main Injector is another synchrotron accelerator, and it serves three main purposes:

- To accelerate protons from 8 GeV to 150 GeV
- To produce 120 GeV protons used by the antiproton source
- To inject 150 GeV protons and antiprotons into the Tevatron

The antiproton source consists of the Target Station, the Debuncher, and the Accumulator. The Target Station is where the 120 GeV protons, provided by the Main Injector, smash into a nickel target and produce antiprotons along with a large number of other particles. A pulsed magnet deflects negative secondary particles of momentum 8 GeV into a transport line leading to the Debuncher. At this point, only $\sim 1\%$ of the particles in the Debuncher are actually antiprotons. The Debuncher is a synchrotron accelerator, and as the particles begin to make their revolutions, particles such as π^- mesons and muons decay, leaving only the antiprotons. The final rate for antiprotons to enter the Debuncher from the initial proton on nickel collisions is on the order of 2×10^{-5} antiprotons per collision[1], and

---

[1] That is to say, for every 100,000 protons hitting the target, two antiprotons are stored.

these are stored in the Accumulator, a small antiproton accelerator. Transfers of antiprotons from the Accumulator can be made directly to the Main Injector for immediate use in the Tevatron. However, the process of storing antiprotons ("stacking") becomes less efficient as the number of total antiprotons rises, and from here antiprotons are typically transferred to the Recycler until they are used in the Tevatron.

The Recycler is another synchrotron accelerator whose main functionality and role have both evolved and increased from its original designs. The Recycler was built to act as a storage ring for antiprotons after stores terminated in an effort to reclaim them for use in the following cycle of stores. Because store durations and peak instantaneous luminosities are limited largely by the number of antiprotons available for collisions, reclaiming antiprotons is an attractive option. As the Tevatron made sizable gains in both its production rates of antiprotons and its ability to attain and maintain increasingly high luminosities, the Recycler assumed a much more vital role. Instead of mainly recycling previously used antiprotons from ending stores, it became a storage ring, much larger than the Accumulator, for newly created antiprotons awaiting injection into the Tevatron. Using a technique called electron cooling, it is able to store large amounts of antiprotons with a small spread of their bunches' sizes in both momentum and space. In fact, the standard mode of operation is for antiprotons to be collected in the Accumulator until reaching a total number of $\sim 80 \times 10^{10}$ antiprotons, then they are added to the collection of antiprotons in the Recycler. From here the antiprotons, via the Main Injector, are inserted into the Tevatron at energies of 150 GeV.

The final destination for both protons and antiprotons in the acceleration process is the Tevatron. The Tevatron is a synchrotron accelerator with a circumference of roughly four miles. It uses superconducting magnets to provide the intense magnetic fields needed to bend the paths' of protons and antiprotons and to maintain their momenta for times up to and occasionally exceeding 24 hours, the typical store duration. The Tevatron, like the Booster, uses dipole magnets to bend the particles' paths, and quadrupole magnets to minimize the spread in distance of the particles within any given bunch. Because protons and antiprotons have opposite charges, they will move in opposite directions under the influence of the same magnetic field. This is advantageous for the Tevatron, and because of this fact can utilize the same set of magnets and beam pipe to accelerate both the positively charged protons and the negatively charged antiprotons in opposite directions. It greatly reduces the redundancy of the system and the power needed for its operation. Their orbits are such that these two paths only intersect at two places along the Tevatron's ring, BØ (where CDF resides) and DØ. These collision points are located at the center of both experiments, and these proton-antiproton interactions at a center-of-mass energy of 1.96 TeV provide the basis for the physics that is performed there.

3.2 The DØ Detector

The DØ detector provides information about particles originating from the proton-antiproton collisions provided by the Tevatron. To develop the most complete picture of these interactions, this massive detector uses multiple subdetectors, each optimized to provide specific information from interesting collisions. From tracking particles' paths to producing high resolution energy measurements, these systems combine to form a very powerful tool for investigating the universe at its smallest scale. As it is important to understand both the strengths and weaknesses of an experimental apparatus, the DØ detector will be over-viewed; however, for a more detailed description of the entire system, the reader is encouraged to review Ref. [28].

The DØ detector is composed primarily of several major systems[2]: a tracking volume encapsulated by a solenoid magnet, a three-cryostat hermetic liquid argon calorimeter, a muon detection system, and luminosity monitoring detector. DØ realizes its high resolution position measurements near the beam pipe using the Silicon Microstrip Tracker, and these, in combination with information from the Central Fiber Tracker, allow the reconstruction of particle paths, called tracks, through the detector. A solenoid magnet is used so that charged particles' paths curve inside the tracking volume. The liquid argon calorimeter provides precise energy measurements optimized for both electromagnetic and hadronic particle showers. The muon system surrounds the calorimeter as the outermost detector in the assembly, identifying muons with scintillating and drift tube detectors. Additionally, a luminosity system is situated at large distances from the center of the detector, but near the beam pipe, to estimate the luminosity seen from the collisions to high precision. The entire detector schematic can be seen in Fig. 3.2.

---

[2] Various components of the detector were upgraded or added in the Summer Shutdown of 2006, such as the addition of another layer of silicon and an upgrade to the calorimeter trigger. Although these changes are both interesting on their own right and important for increased precision of the detector, the data used in this analysis taken before these additions, which are detailed in Ref. [29, 30].

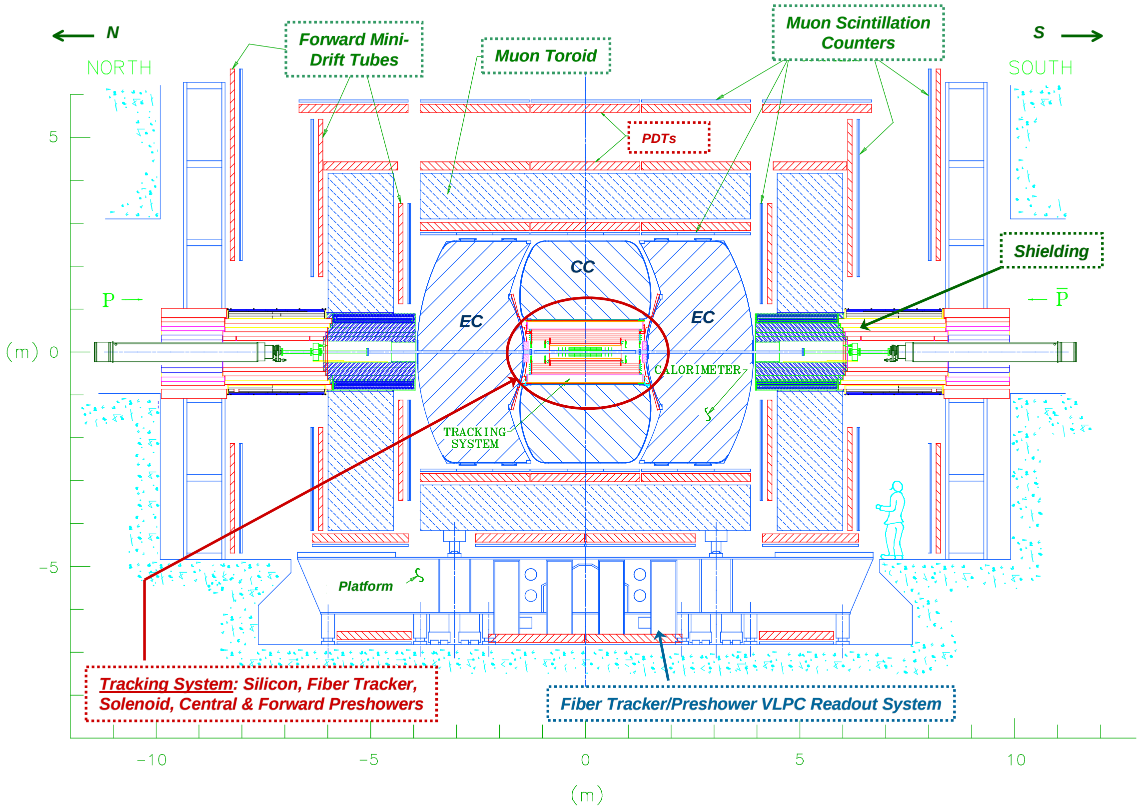


Figure 3.2: The DØ detector broken up into its detector components.

3.2.1 DØ System of Units

Before delving too deeply into the DØ detector and its subsystems, it is important to understand *what* is being measured and *how* that information is conveyed.

The first step is to understand how to navigate the detector, and to do this, it is important to have a common physical orientation, as well as a coordinate system in which it can be easily described. The natural coordinate system for the detector is cylindrical, as much of it is cylindrical in shape and symmetric about the beam pipe. The origin of this geometric system lies at the center of the detector. The positive z-axis corresponds to the axis along the beam pipe in the direction of protons. The azimuthal angle ϕ is defined to be between zero and 2π in a plane perpendicular to the z-axis, and r is the radial distance from the beam pipe. Rapidity is defined as

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right], \tag{3.2}$$

where E is the energy of the particle and p_z is the component of the particle momentum parallel to the z-axis. Rapidity is a dimensionless quantity, and it is advantageous to use as differences in rapidity (Δy) are Lorentz invariant. In the relativistic limit, rapidity can be approximated by pseudorapity (η):

$$\eta = -\ln[\tan(\theta/2)], \tag{3.3}$$

where θ is the polar angle. As a particle's energy becomes much larger than its mass ($E \gg M$), its rapidity approaches the pseudorapidity ($y \to \eta$), and for massless particles $y \equiv \eta$.

It follows naturally that the units to best describe particles' energies, time scales, and masses may not be the ones we use in our daily lives. In the frame of System Internationale (SI) units, we measure the world in kilograms, meters, and seconds. For the world of particle physics, on the other hand, the objects of interest are on the order of 1×10^{-27} kg (roughly the mass of the proton and neutron), of 1×10^{-15} meters (known as a femtometer), and of 1×10^{-25} seconds. To add to the confusion, since the beams of protons and antiprotons are traveling approximately the speed of light, relativistic equations are used to understand their energies, angles, and momentum transfers. These calculations often contain higher orders of constant terms specifically the speed of light (c) and Planck's constant ($h = 2\pi\ \hbar$). To simplify terms involving these constants, a notation of *natural units* has been employed. In this framework, c $= \hbar = 1$. As a consequence, mass, energy, and momentum share the same units. The most natural units to use for high energy physics is the GeV[3]. The inherent applicability in these units is apparent when looking, for instance, at the proton's mass: $1.66 \times 10^{-27} = 0.938$ GeV. Length and time in this system are somewhat more confusing still. They now share the same units, the GeV^{-1}. To give a sense of scale, 1 GeV^{-1} $= 1.975 \times 10^{-15}$ meters $= 6.59 \times 10^{-25}$ seconds.

Now that our orientation is complete, the remainder of the section will take a deeper look at the DØ subdetectors, exploring and explaining both their construction and functionality. Since a measurement is only as good as the tools used to make it, understanding the

---

[3] In English, this is the **giga-electronVolt**, where 1 GeV $= 1.6 \times 10^{-10}$ Joules.

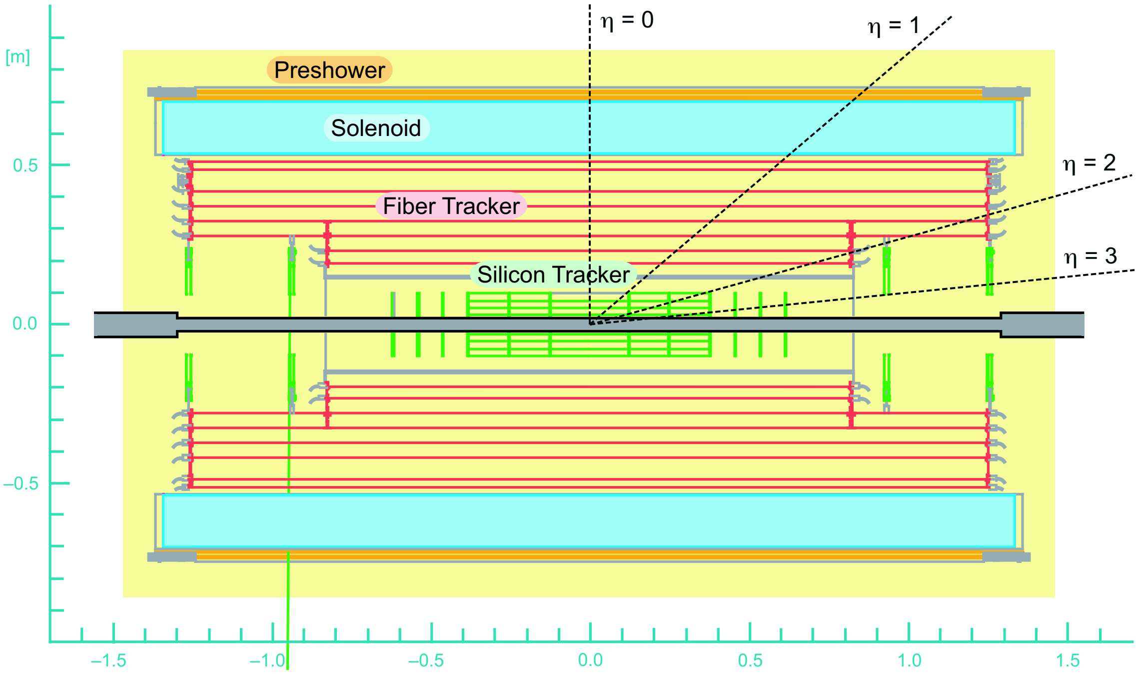


Figure 3.3: Two dimensional schematic view of the DØ tracking system, consisting of a silicon system and a fiber tracker within a solenoid magnet, as well as a preshower detector outside of it.

information produced from them provides a solid experimental foundation on which to build a result.

3.2.2 Tracking System

The tracking system combines high density silicon with the larger volume of a fiber tracker that are used together to resolve particles' tracks and determine the primary interaction vertex position within 35 μm. This system is subject to the solenoid's constant two Tesla magnetic field running parallel to the beam pipe, which causes charged particles trajectories to bend in the transverse plane as they traverse this volume.

Silicon MicroStrip Tracker

The Silicon Microstrip Tracker (SMT) is the innermost detector with respect to the beam pipe and is essential for high resolution track reconstruction near the interaction point. This resolution is extremely important for the reconstruction of displaced vertices, which are formed in part from decays of mesons containing either c or b quarks. The impact parameter is the minimum distance of a track to the primary interaction vertex and is a major component to distinguishing heavy flavor decays. The combined resolution of the entire tracker can resolve the impact parameter's length within 15 μm.

To attain such a high resolution, the SMT uses the semiconductor properties of silicon. A semiconductor conducts currents when its electrons are in the conduction band of the material, which occurs when its electrons are excited. When an electron is excited, it is free to propagate its negative charge and it leaves a hole in the valence band of the silicon, which in turn acts as a positive charge propagator. Silicon is used due to the small energy gap between the valence and conduction bands, translating into a large number of freed electrons for a given amount of energy. When a biased voltage is applied to the silicon, the electrons and their hole partners create an electrical current, which is used as the the signal of a charged particle passing through the material. The SMT uses the silicon wafers, which correspond to a total of 792,576 readout channels, to detect charged particles coming from the $p\bar{p}$ collisions. The wafers and are mounted on cylinders ("barrels"), and disks, both centered about the beam pipe. The SMT is composed of six barrels in its central readout that extend to $z = 38.7$ cm, and 16 disks for extended η and barrel gap coverage, as shown in Fig. 3.4. Each barrel contains four silicon readout layers and within each layer, silicon wafers are situated back-to-back with a stereo angle between them. This back-to-back wafer configuration is called a ladder. The first two layers house 12 ladders and the second two house 24 ladders each, combining for a total of 432 ladders in the central barrel readout.

The disk assembly for the SMT is divided into 12 F disks and four H disks. F disks are located at each end of the detector (three on each side) and at the end-to-end meeting of two barrels. The exception to this being that there is no F disk located at $z = 0$ cm. These disks are composed of "wedges", providing similar functionality as the silicon wafer ladders, only with a different geometry. For the F disks, the wedges are doubled-sided silicon wafers, with each side reading out independently.

In addition to the F disks, the SMT employs the use of four H disks [4] to complete its detector assembly. These disks are designed to increase the η range of the SMT, and are split evenly on each side of the three F disks. These disks are each composed of 24 wedges. These wedges are similar to those used in the F disks, with the exception that each wedge

---

[4] The two furthest H disks in z were removed in the spring of 2006.

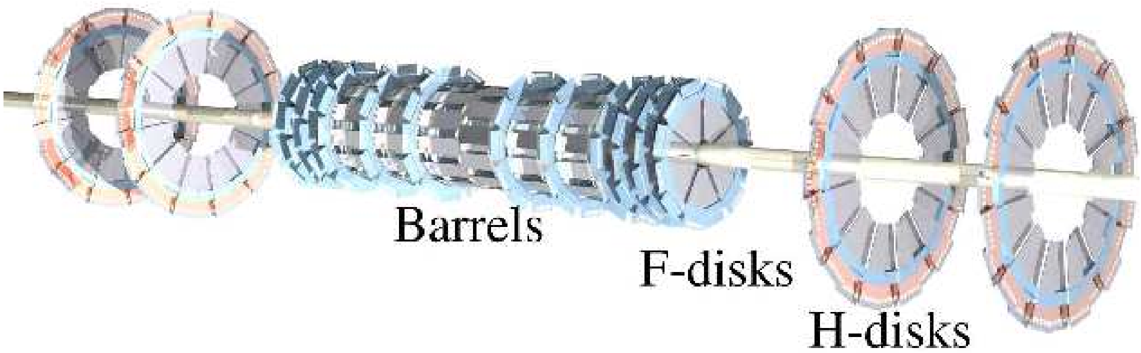


Figure 3.4: Three dimensional view of the SMT system and beam pipe. The SMT is composed of barrel and two types of disk detectors.

is two single-sided silicon wafers that are back-to-back.

In order to maximize the effectiveness of the readout channels, a portion of the wafers were positioned in a way to provide stereo coverage. For the F disk double-sided wafers, the stereo angle is 30°. For the barrel detectors, the two sides of the wafers have an opposite bias (negative and positive) with a 2° angle between them. These two wafers comprise one 12 cm long ladder. The single-sided wafers, placed back to back on H disks, have a stereo angle of 15°.

The core of the SMT electronics lies with the SVXIIe chip. Each chip contains 128 readout channels and performs the majority of the data collection operations. This includes the preamplification, analog delay, and digitization of the signal as well as data sparsification. The analog delay can be as long as ~ 4.2 microseconds, corresponding to a superbunch of 12 $p\bar{p}$ crossings. Once an event is deemed to be of good quality, this signal charge is digitized and stored as hits that, when used in combination with the information coming from the fiber tracker, form the basis of the offline track reconstruction algorithm (Section 4.2.1).

Central Fiber Tracker and Preshowers

The Central Fiber Tracker (CFT) is the second innermost layer of the DØ detector and provides the majority of the position measurements of particle paths from energetic collisions using scintillating plastic fibers. Charged particles passing through scintillating plastic excite its electrons. When one of its excited electrons returns to its ground state, it emits a photon of a characteristic wavelength corresponding to the energy between the ground and excited states. The wavelength of this photon in the scintillating fiber is in the visible spectrum, and the measured number of produced photons corresponds to the signal for the intereacting charged particles. The information from each scintillating fiber is recorded to later be combined to recreate charged particle paths' as they traverse the tracking volume.

The CFT is composed of eight barrel superlayers and has an outer barrel coverage of $|\eta| < 1.7$. Each of these superlayers contains an axial layer of fibers, running parallel to the beampipe. The fibers are arranged in a doublet format, having fibers in between the gaps of the first single layer of fibers. There is also a doublet stereo layer of fibers per superlayer, placed at a stereo angle of 3.0° with respect to the beampipe. This layer's stereo angle is positive (+3.0°) for odd numbered superlayers, and negative (−3.0°) for even numbered ones. The fibers themselves are scintillating, and are doped to quickly produce photons within a few nanoseconds. The fibers are also wavelength shifting such that these photons, of wavelength ~ 340 nm, are in turn reabsorbed and re-emitted as ones with a peak emission spectrum of 530 nm. The smaller attenuation of these photons is important as the fibers range in length from 1.66 to 2.52 meters. They are connected at one end to waveguides and are mirrored at the other end, so that the signal may be reflected back. This is done to increase the efficiency of light detection for the detector. They are constructed to be minimally attenuating, which is important as the signal at this point is still analog in nature. The signal photons are sent through clear fibers that act as waveguides, which lead to the Visible Light Photon Counters (VLPCs), sophisticated avalanche photodiodes that increase the gain of the signal.

The preshower detectors are composed of a central preshower (CPS) and two forward preshowers (FPS) with pseudorapidity coverage of $|\eta| < 1.1$ and $1.5 < |\eta| < 2.5$ respectively. The central preshower detectors are located between the calorimeter cryostat and the solenoid, and the forward preshowers are mounted on the faces of the end cap calorimeter. The preshowers were added not only as an attempt to regain the precision in energy measurement to the calorimeter that was lost by adding a solenoid in the Run II upgrade, but also to provide a very fine spatial resolution of shower development before the calorimeter.

The CPS is constructed from three layers of triangular plastic scintillating strips, one whose strips are parallel to the beampipe (X layer) and two that are angled with respect to the beampipe. These are stereo angles are measured to be 24.0° and -23.7° for the second (U layer) and third (V layer) layers respectively. Each layer contains 2560 strips and is split at $z = 0$ cm.

The FPS is similar in construction, using the same triangular plastic scintillating strips, but with disc-shaped panels that sit perpendicular to the beampipe, mounted directly onto the calorimeter endcap detectors. This is in contrast to the cylindrical ones used in the CPS geometric design. In addition, two layers of scintillators are used to form the depths of each disc, with an additional layer of lead absorber (X_0[5]) that was put in between each layer of the scintillating strips to maximize particle showers in the second layer.

From the energy deposition and strip crossings from all layers, a position in (ϕ, z) can be determined along with the sum of the energy from all associated strips. This process is known as 3-D clustering, and is performed for both the CPS and FPS from their strip information.

Because the CPS and FPS are both scintillating detectors producing photonic signals, they utilize the same light detection readout setup as the CFT, including the VLPCs. From the digitized information coming from the VLPCs, 3-D clustering is performed using the position and energy information from both preshower systems separately.

Solenoid

The tracking detectors are surrounded by a superconducting solenoid that generates a 2 Tesla magnetic field parallel to the beam direction. The paths of charged particles are bent under the influence of a strong magnetic field ($\vec{v} \times \vec{B}$), and this technique provides a simple and effective way of deducing particles' momenta from their tracks' curvatures. The magnet uses two layers of superconductor to achieve the linear current density needed for the 2 T field. It is 1.42 meters in diameter and 2.73 meters in length, and accounts for 1 X_0 at $\eta = 0$, and close to 4 X_0 at $\eta \sim 1.7$. The measured magnetic field is within 0.09% of its calculated value, and its overall variation is on the order of ~0.3% in the fiducial tracking volume. The magnet is capable of dual polarity, and is operated roughly an equal amount of time with the field in each direction.

---

[5] X_0 is the general notation for radiation length, which is the amount of material needed to reduce an electron's energy by a factor of e. For photons, the radiation length is 7/9 of the mean free path for pair production.

3.2.3 Calorimeter

The purpose of a calorimeter is to measure particles' energies. This is achieved by putting a known amount of material in the particle's way to make it expend its kinetic energy in the production of new particles. The interactions of the particles with material that creates the subsequent ones particles is referred to as showering. The calorimeter exploits this effect with multiple layers of very dense material and in between a liquid that, when ionized, will give a measured signal proportional to the energy deposited. From this measurement the particle's original energy can be estimated.

The DØ calorimeter is a liquid argon sampling calorimeter. It is housed in three separate crystostats, the central giving coverage of $|\eta| < 1.1$, and the entire calorimeter providing coverage of $|\eta| < 4.2$, shown in Fig. 3.5. The electronics were upgraded between Run I and Run II in order to handle the high luminosities and decreased bunch spacings of the Tevatron. However, the crystostats have remained unchanged from the original Run I design [31].

In addition to original Run I calorimeter, both the Inner Cryostat Detector (ICD) and the previously discussed preshower sytems are used in Run II for particle energy measurements. The ICD was installed to give additional coverage between cryostats, specifically in the region of $1.1 < |\eta| < 1.4$. The preshower detectors, both central and forward, were added to provide additional particle energy and position measurements as compensation for the added material in front of the calorimeter in Run II as previously described.

Particles that interact electromagnetically, specifically photons and electrons, deposit the majority of their energy passing through a much smaller amount of material than do hadronic particles. The DØ calorimeter is longitudinally segmented into four electromagnetic (EM) layers, four fine hadronic (FH) layers and up to three coarse hadronic (CH) layers to optimize the energy response for various types of particles. A schematic drawing of the entire calorimeter is shown in Fig. 3.6. The calorimeter uses a combination of metal absorbers and active liquid to accurately measure particles' energies that pass through it. Liquid argon (LAr) is the active sampling medium and the absorber layers, as well as the thicknesses, are listed in Table 3.1.

Each layer of the calorimeter is comprised of a 2.3 mm gap filled with LAr between an absorber plate and a G10 insulator board coated with an epoxy with high resistivity. Particles passing through the gap ionize the active medium and a potential of ~2 kV creates

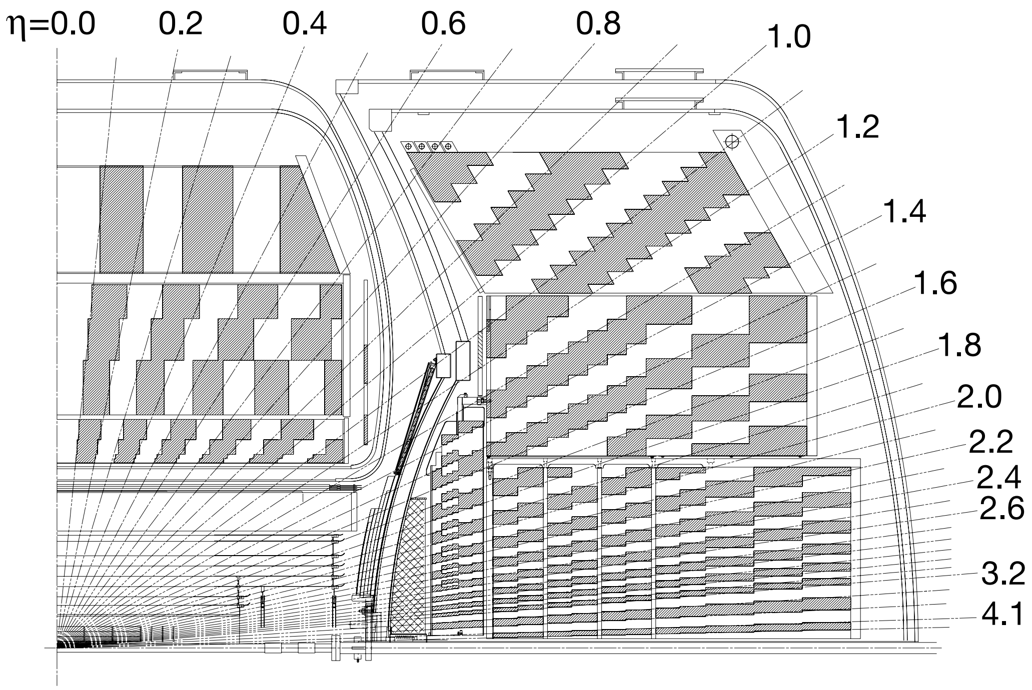


Figure 3.5: The longitudinal view of the DØ calorimeter, showing both the segmentation of the electromagnetic and hadronic layers and the η coverage of the entire calorimeter.

Table 3.1: Amount of material before each layer of the central (CC) and end cap (EC) calorimeters for the electromagnetic (EM), the fine hadronic (FH) and coarse hadronic (CH) layers. The thicknesses are listed in terms of radiation lengths (X_0) for the EM layers and in absorption lengths (λ_A) for the hadronic layers.

Layer	Region	Thickness	Material
EM	CC	2, 2, 7, 10	Uranium (3mm)
EM	EC	0.3, 3, 8, 9	Iron(1.4mm) + Uranium (4mm)
FH	CC	1.3, 1, 0.9	Uranium (6mm)
FH	EC	1.3, 1.2, 1.2, 1.2	Uranium (3mm)
CH	CC	3	Copper (46.5mm)
CH	EC	3, 3, 3	Iron (46.5mm)

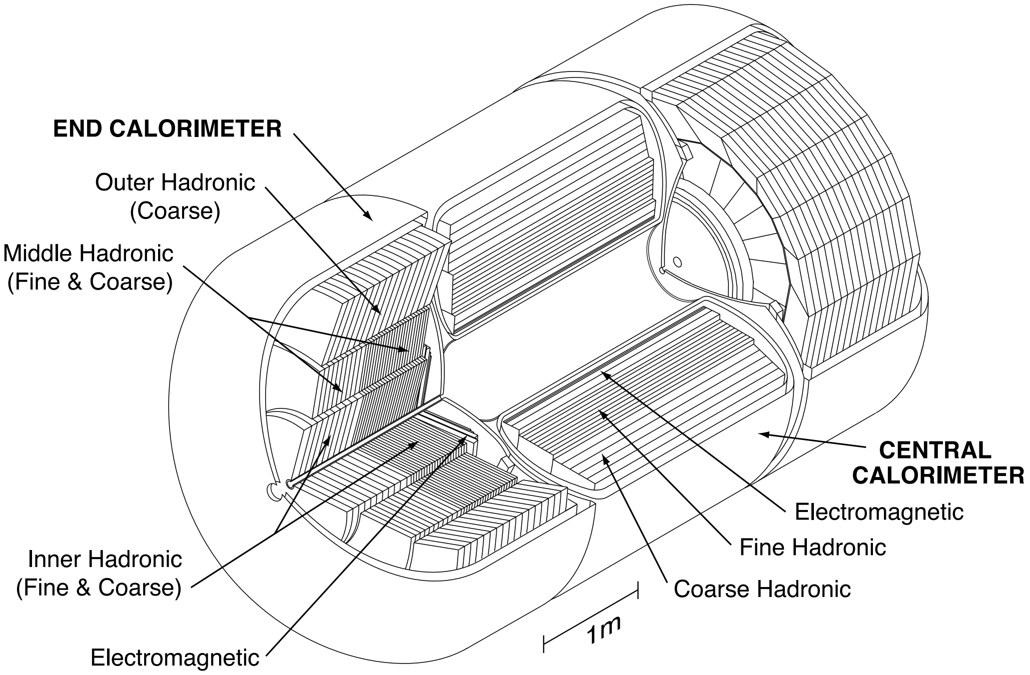


Figure 3.6: The three cryostat, hermetic liquid argon sampling calorimeter at DØ.

a current of electrons towards the resistive pad with an average drift time of ~450 ns. Etched copper pads sit between the insulator boards and collect the produced image charges, and pads are ganged together to form individual readout cells. Figure 3.7 shows a representation of the calorimeter cell design.

For the entire calorimeter, there are 47,364 active readout channels. The cells corresponding to these channels are finely segmented, 0.1×0.1 ($\Delta\eta \times \Delta\phi$), with increased segmentation for the maximal EM shower layer to 0.05×0.05. Up to 12 cells form pseudo-projective towers in each $\eta - \phi$ position.

Once the charge is collected at the copper pads within the detector, it is sent via low impedance (30 Ω) cables to the preamplification system. The preamplifiers are designed to match the cables' input impedance to prevent reflections in the signal. Fourteen types, or species, of preamplifiers are used to minimize the effects of cell-to-cell capacitance differences, which can be up to almost a factor of ten. This is done to create an integrated signal from the preamplification system that is independent of the input cell characteristics. This integrated signal is then sent to the Baseline Subtraction System.

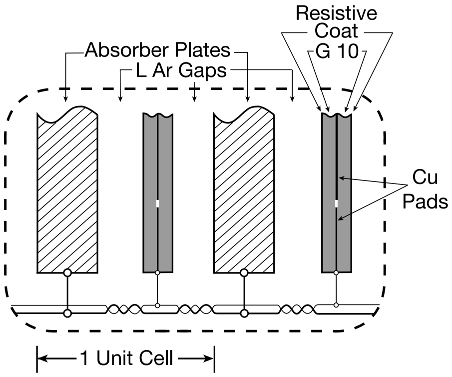


Figure 3.7: Schematic of a calorimeter cell.

The Baseline Subtraction (BLS) system receives the voltage signal from the preamplifier system and determines the energy deposited in each cell on an event-by-event basis. When the signal arrives to the BLS system, it is first intercepted from the Level 1 Calorimeter (L1Cal) trigger pick-off. The L1Cal sums the analog energy in towers of 0.2×0.2 ($\Delta\eta \times \Delta\phi$) for use in trigger decisions of event selection and is described in greater detail in Section 4.1.1.

The signal continues on to the signal shapers. The signal is shaped such that only the rising edge of the input signal is used in its energy estimation, which is approximately two thirds of the total signal. This is done because an integration of the entire signal is impossible given the timing constraints imposed by the Tevatron's 396 ns bunch crossings. The signal is then sampled at its newly shaped peak.

The shaped signal is sampled every 132 ns, and in order to remove the remaining energy from previous events, the BLS system subtracts the signal 396 ns before the peak, which is called baseline subtraction. This modified signal energy is stored until the event passes the second level of the three-tiered trigger system (Level 2). Due to the high volume of events processed and awaiting Level 2 trigger decisions, Switched Capacitor Arrays (SCAs) are used for analog storage. The SCAs are also used for storage of signal information along the two gain paths, $\times 1$ and $\times 8$. The use of two gain paths increases the dynamic range for the analog to digital conversion from 12 to 15 bit. The correct gain path is chosen by the BLS circuitry and if the event passes the Level 2 trigger requirements, it is sent to the Analog to Digital Converters.

The Analog to Digital Converters (ADCs) transform the analog signal from the calorime-

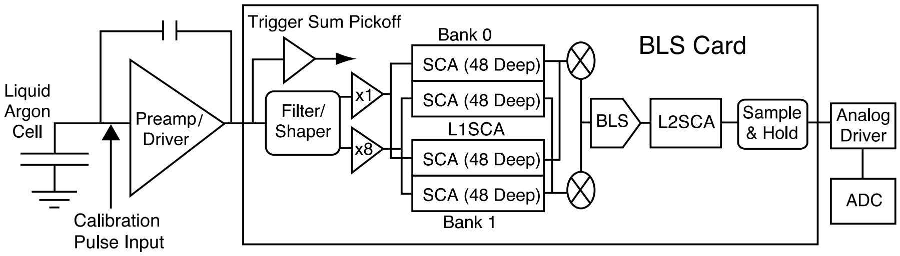


Figure 3.8: Schematic of the analog signal through the BLS circuitry using two independent gain paths.

ter cell into a digital energy value. It is here where signals are subject to a 2.5σ threshold in order to further suppress noise and signals from overlapping minimum bias events [6]. The ADC system also controls the timing and control operations of the entire calorimeter. It is synchronized to the accelerator clock and sends commands to the BLS system to regulate the signal and background sampling and subtraction. It also coordinates event processing among the ADCs and communications to the overall DØ detector framework. The digitized energy is sent to be used in event reconstruction.

3.2.4 Muon System

The muon system at DØ was developed to efficiently detect muons in both the central ($0 \lesssim |\eta| \lesssim 1$) and forward ($1 \lesssim |\eta| \lesssim 2$) regions of the detector. Both regions are optimized to provide superior resolution and there are multiple layers for confirmation of muon trajectories through these detectors.

The central muon system is primarily composed of three layers of proportional drift tubes (PDTs), scintillation counters and a toroidal magnet. The first (A) layer of the PDTs is located within the toroid, and the B and C layers are outside of this magnet. The PDTs themselves each consist of an anode wire, held at a high electric potential, with vernier cathodes placed above and below that read out to provide additional position information from the ionizing particle passing through the chamber. With each PDT "hit", its position within the chamber is calculated from the readout time difference from each end of the wire. This information, combined with the charge deposited on the cathode pads, can result in

---

[6] Minimum bias events do not contain large transverse energy.

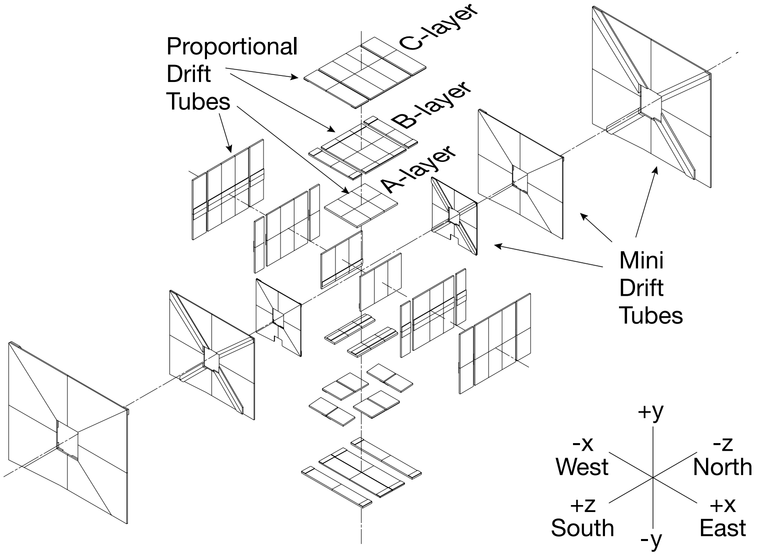


Figure 3.9: Schematic view of the central muon system, including both the central proportional drift tubes (PDTs) and the forward mini drift tubes (MDTs).

achieved resolutions of up to ∼1 mm. An exploded view of the muon system's wire chambers can be seen in Fig. 3.9.

The scintillating counters are physically divided into three separate areas in the central muon detector: the cosmic cap, the bottom scintillator counters, and the A-ϕ counters, which are located with the A layer PDTs. All are made of plastic scintillating tiles, grooved to accommodate waveshifting fibers that bring the light to the PMTs (256 cosmic cap, 116 bottom, and 630 A-ϕ). A schematic picture of their geometry as it fits together with the PDTs can be seen in Fig. 3.10.

The A-ϕ scintillator counters are used to provide information to the first level of the DØ trigger system. Their segmentation in ϕ matches that of the CFT readout in order to facilitate a combination of their information to create a more powerful muon candidate selection at this stage in the trigger.

The forward muon system, covering a pseudorapidity range of $1.0 \lesssim |\eta| \lesssim 2.0$, uses much of the same technology as the central system. It is composed of both three layers of Mini Drift Tubes (MDTs) and three layers of scintillating counters. The scintillating counters provide the information to the triggering system, and the MDTs are optimized for the precision

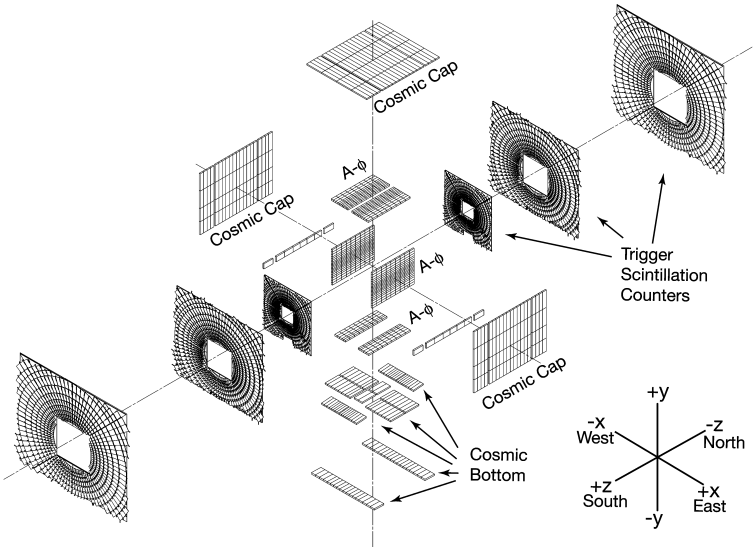


Figure 3.10: The scintillating section of the muon detector, which includes the cosmic caps as well as the scintillation counters used for triggering.

readout. The MDT system consists of 6,080 drift tubes, with each tube containing eight wires. This combines for 48,640 readout channels for the MDTs and, like the central readout system, is divided into octants. For the drift tubes, the electron drift time is measured to be less than 132 ns, and the resolution per cell is $\sigma < 1$ mm. This level of precision is of the same magnitude as seen in the PDTs for central muons, and this entire system sets the foundation for high precision muon identification especially when combined with particle path information from the tracking volume.

3.2.5 Luminosity Detector and Luminosity Determination

Luminosity at DØ is monitored using two arrays of scintillating counters placed at $z = \pm 140$ cm. Each array covers a pseudorapidity range of $2.7 < |\eta| < 4.4$ and contains 24 scintillating counters with PMTs mounted on the scintillators, as shown in Fig. 3.11. The counters for each array are housed in two light-tight enclosures, with additional preamplifiers that increase the gain by a factor of five. The timing and charge information from the arrays is digitized and combined to create time of flight information from collisions for each bunch crossing. Because the detector's clock is in sync with the accelerator, the timing of each

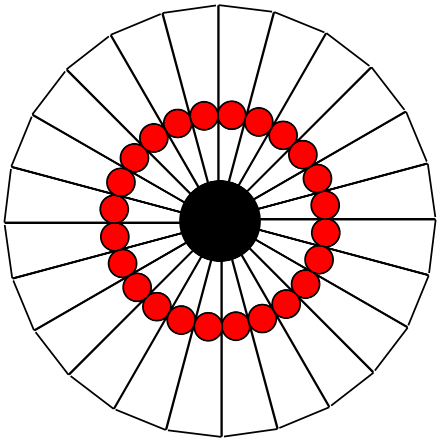


Figure 3.11: Transverse view of one of the two scintillating arrays that comprise the DØ luminosity monitoring system, including 24 long scintillating strips with mounted PMTs (small circles).

bunch crossing is known to very high precision. Thus any difference in the time of flight between the two arrays gives rise to a displaced primary interaction vertex z position (PV_z). This difference can be determined using Equation 3.4, where t_+ (t_-) is the time of flight (TOF) for the array at the positive (negative) z position.

$$PV_z = (\mathrm{c}/2)(t_- - t_+) \tag{3.4}$$

The luminosity is determined from the average number of interactions per bunch crossing ($\tilde{N}_{LM}$), the beam crossing frequency (f), and the effective cross section "seen" by the luminosity monitors (σ_{LM}):

$$L = \frac{f\tilde{N}_{LM}}{\sigma_{LM}}. \tag{3.5}$$

Because multiple $p\bar{p}$ interactions can take place in any given bunch crossing, $\tilde{N}_{LM} \gtrsim 1$, and additional handles are needed to accurately determine the luminosity. The solution is to count the number of bunch crossings without interactions, and calculate $\tilde{N}_{LM}$ from Poisson statistics [32]. The combination of this with the time of flight information yields the final number of interactions per second at the detector.

CHAPTER 4

DATA ACQUISITION AND RECONSTRUCTION

4.1 Trigger System

Roughly 2.5 million bunch crossings occur every second when the Tevatron's $p\bar{p}$ beams are in operation, and the events from these crossings are largely filled with soft scattering processes and noise, as well as the occasional large momentum transfer leading to the more interesting physical observables. Due to limitations of data recording and object reconstruction, only the most interesting events are selected to be stored. This filtering procedure is performed in a three-tiered fashion [28], reducing the total number of events per second from 1.7 MHz down to roughly 50-100 Hz. Each tier, or level, is optimized based on the amount of time in which decisions must be made.

4.1.1 Level 1 Trigger

The Level 1 (L1) trigger is the first filter in this system and is designed to reduce the event rate from 1.7 MHz down to $\sim$ 2 kHz, a rejection of almost a factor of 1000. The decision time at the L1 trigger stage is 4.2 μs, and employs event buffering, built into the subsystem electronics, to satisfy this time constraint. Each subsystem has its own L1 trigger system and all communicate with the trigger framework (TFW), as shown in Fig. 4.1. The exception to this is the SMT, as it is limited by its relatively long readout time of 100 μs. Thus its information is only considered after the L1 trigger determines that an event satisfies at least one trigger term requirement. Although the analysis performed only utilizes calorimeter trigger information, all major L1 systems will be briefly reviewed.

The calorimeter's L1 trigger system (L1Cal), bases trigger decisions on the amount of energy deposits in the calorimeter and, to a lesser extent, the shape of these deposits. Because of time constraints, the detector information used for these decisions is taken from an analog

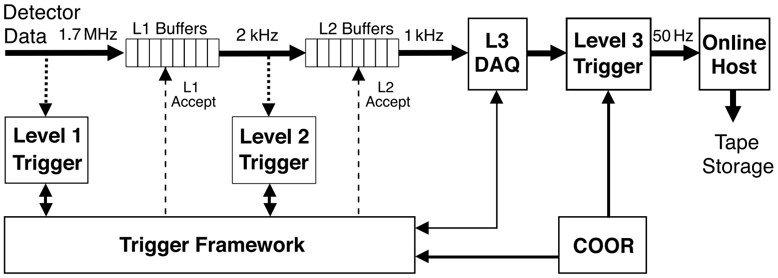


Figure 4.1: The three-tiered trigger system layout that is coordinated using the trigger framework.

trigger pickoff located at the BLS system, but before the baseline subtraction procedure is performed. All cells are grouped in pseudo-projective trigger towers of size 0.2 × 0.2 ($\Delta\eta \times \Delta\phi$), four times the size of the precision readout towers.

All towers are still divided into their EM and hadronic components, and the L1Cal system uses information from 1280 separate EM and hadronic[1] trigger towers. EM objects and hadronic jet trigger terms take advantage of this separated tower information. In each event, the information from all energetic towers is stored and then combined to form basic hadronic and electromagnetic energy clusters. If the energy of one of these clusters meets the trigger requirements, such as minimum cluster energy or number of energetic towers, then the event will continue to the second level of triggering. The L1Cal trigger was upgraded in Spring 2006 and now contains more sophisticated algorithms for EM object and jet selection; however, because this analysis does not use data after this upgrade period, these methods are not detailed here. They can instead be found in Ref. [29].

The tracking system uses information from three separate areas to form the basis for its first level trigger, the Level 1 Central Track Trigger (L1CTT). These are separated into the various parts of the detector from which they gather their information. The first is the stereo information from the CPS and the second is the entire FPS. The third and the one most commonly used as the basis for track triggering at DØ uses the information coming from the CFT and CPS axial layers. Each of these layers is divided into 4.5° sectors in the azimuthal angle ϕ, with additional information from neighboring sectors used to extrapolate tracks crossing sector boundaries. Fiber hits of all layers in each sector are compared to

---

[1] The hadronic energy in trigger decisions only comes from the fine hadronic section of the calorimeter.

$\sim$20,000 predefined track equations, which are stored in look-up tables. The found tracks are subject to a 1.5 GeV threshold, then are ordered in p_T and divided into four p_T ranges, 1.5–3, 3–5, 5–10 GeV, and greater than 10 GeV. In each of these p_T ranges, the six highest p_T tracks are stored. If there are fewer tracks than six in a given p_T range, then they will all be saved. This equates to up to 24 tracks (6 tracks × 4 ranges) stored per sector.

The next stage in the L1CTT track reconstruction is the track merging, where tracks are matched to others, as well as to CPS clusters. Sectors are combined into ten super-sectors, each covering 36° in ϕ. Here the initial tracks are combined when possible, and their momenta are redetermined. After this track combination is finished, these new tracks are stored in a p_T ordered track list for each super-sector. In each of these, variables forming the basis of the 55 available trigger terms are calculated. These variables include the total number of tracks, the number of all CFT fibers active, and also the number of high p_T isolated tracks[2].

The muon system uses a combination of its wire chambers, scintillating counters, and found tracks from L1CTT to find patterns consistent with those of muons. The first level of the muon trigger (L1Muon) uses 32 trigger terms and takes as input $\sim$60,000 muon channels and up to 480 L1CTT tracks from every bunch crossing. The triggering is divided into three physical regions, the North, South, and Central. Each region is further divided into octants.

There are two main modes of running L1Muon's muon identification. The first is track-matching, which takes advantage of the information coming from both the L1CTT and muon scintillators in each octant. Ten 4.5° sectors, plus an additional sector on either side, correspond to one octant. Each octant is accessed for its list of p_T ordered tracks. These tracks are matched to hits in the layers of the muon scintillator, using the same track p_T ranges as defined from the L1CTT requirements, and depending on the number of layers that are matched to each track, a quality assessment is determined. For example, if a track is matched to the scintillator's A layer only, its quality is "loose"; however, if it is matched to the A and B layers both, it becomes "tight". The second form of L1Muon's muon identification is very similar to the first in that it again uses the scintillator information, but in this case the matching is performed with respect to the wire chamber hits instead of L1CTT tracks.

The Trigger Frame Work (TFW) is the system that coordinates all the event information coming into the first level of triggering. The TFW makes the decisions on whether events

---

[2] Isolation here refers to the lack of CFT fiber hits that exist in an area about a given reconstructed track.

pass or fail based on the trigger terms coming from all of the Level 1 (L1) trigger systems. It also executes specific vetoes of triggers and trigger prescaling, which limits the number of those trigger accepted for a stable reduction of rate. A trigger that is prescaled by a factor of 3 indicates that only 1/3 of those triggered events will be passed[3]. The TFW's trigger decisions are composed of 256 "AND-OR" terms, specific individual conditions that form trigger terms through their various combinations. For an event to be passed on to the next level of the trigger decision, it must meet the requirements for both a combination of these trigger terms and a beam condition of good quality. Once these are met, an L1 accept is issued and the event is sent to the second level of the trigger system.

4.1.2 Level 2 and 3 Triggers

The Level 2 (L2) trigger was designed to reduce the event rate from $p\bar{p}$ collisions from ~ 2 kHz down to 500–1000 Hz. The second level of the calorimeter trigger (L2Cal) provides basic reconstruction of both jets and EM objects. Jet clustering is performed by first finding a seed tower, which forms the center of the energy cluster. Seed clusters are ordered in decreasing E_T and must pass a threshold energy of $E_T \geq 2$ GeV to be considered. A circle of area 1.0×1.0 in $\Delta\eta \times \Delta\phi$ (or 5×5 trigger towers), is centered on the seed tower. The energies of all hadronic towers within this circle are summed and approximate the jet's original energy.

The L2 EM object reconstruction works in a very similar fashion to that of the hadronic jets. However, in this case seed towers are formed with an energy requirement of $E_T \geq 1$ GeV, and each seed is combined with its neighboring EM trigger tower containing the largest amount of energy. The 3×3 area of EM tower energies surrounding the EM seed tower is taken to approximate the EM object's energy. An additional requirement can be made at this level, which is a very loose form of energy isolation. This isolation is done by taking the two most energetic towers' energies and dividing it by the entire energy associated to this object from the 3×3 trigger tower estimate. For direct photons and electrons, this number should approach one[4].

Finally, after all EM object and hadronic jet candidates have been found, the missing transverse energy for the event is calculated. This is done by taking the vector sum of all the found energies coming from the L1 trigger towers. This provides not only the size of the

---

[3] This analysis requires that only unprescaled triggers are fired in a given event.

[4] EM energy isolation was not required on any trigger used in this analysis.

missing transverse energy, but also determines its direction.

For all calorimeter triggers, the reconstructed EM object or hadronic jet's reconstructed energy is used to determine if an event passes the second level of the trigger. If an accept is issued for the event, it is passed to the third and last level of the triggering system.

The Level 3 (L3) trigger was designed to bring the event rate of $\sim$ 500–1 kHz down to a recording rate of 50–100 Hz. Because L3 is the final triggering level, events undergo much more intense scrutiny. Using over 200 farm nodes (an arsenal of dedicated computers), events are quickly and fully reconstructed using all available detector information. This includes a more precise treatment of object energy reconstruction. For instance, precision calorimeter towers 0.1×0.1 $(\eta - \phi)$ are used at this level to reconstruct object energies. An accepted event is transferred to the final data block for storage at a size of 250 kB. The recorded data are grouped in time intervals called Luminosity Blocks. Luminosity Blocks contain event information for 60 seconds of continuous running, or when the data acquisition system is in a state of transition[5].

4.2 Event Reconstruction

Once an event passes all levels of the trigger system, it is stored for use in data analysis. But before its information can be useful, the data must be translated from detector hits and energy deposits into physics objects. To this end, DØ has developed very sophisticated reconstruction algorithms that recreate the energy and paths of particles as they traverse the detector and decay. Specifically, the areas of interest for this analysis are the methods of reconstruction for tracks, primary interaction vertices, photons, jets and missing energy.

The program responsible for coordinating the processing of this data and the running of the algorithms is DØRECO. The reconstruction is done in stages, since the data must first be unpacked and translated into basic detector quantities. As the stages progress, the algorithms become increasingly complex and sophisticated, and are necessary for high level particle identification and rejection of all types of background. The following section will review the most pertinent algorithms relating to the reconstruction of objects used in this analysis.

---

[5] No data are collected during these transition periods, so these Luminosity Blocks are not used in data analysis.

4.2.1 Track Reconstruction

Particle paths are reconstructed in the DØ tracking volume as tracks. At their most basic level, these tracks are composed of basic position measurements from both the SMT and CFT, discussed in Section 3.2.2. The position measurements are combined in such a way that the deflection of the charged particles' paths can be determined to estimate the particles' momenta in flight.

Track reconstruction [33] is performed by first taking hit information from the SMT and from it forming a pool of track hypotheses. The initial track hypothesis requires its first hit to be either from one of the barrels or an F disk. The second hit must be located in a layer farther from the beam pipe than the first hit, and they must be close in ϕ such that $\Delta\phi(1^{st}, 2^{nd}) < 0.08$. The third hit must be in a layer further from the beam pipe than the first two, and the radius of the circle in the transverse plane fitted to the hits must be bigger than 30 cm, corresponding to a track p_T of 180 MeV.

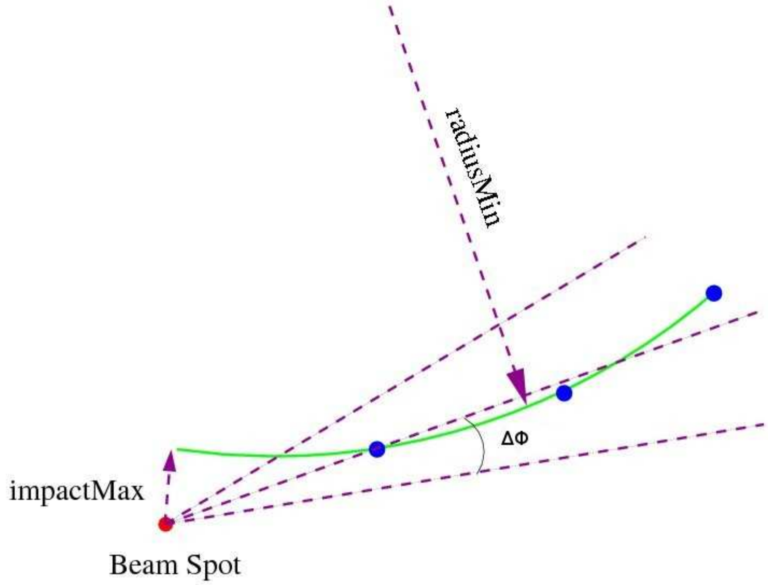


Figure 4.2: Schematic of the track hypothesis method showing SMT hits (blue dots) fitted with a track hypothesis. The track hypothesis must satisfy the conditions of a minimum radius of 30 cm (radiusMin), an impact parameter size less than 2.5 cm (impactMax), and a $\Delta\phi$ between the first and second hit less than 0.08 (the maximum $\Delta\phi$ allowed is shown).

The fit of these points to the track hypothesis must be of a decent quality, such that its

χ^2 satisfies $\chi^2 < 16$. The minimum transverse distance of the track hypothesis to the beam spot (impact parameter) is required to be less than 2.5 cm. These requirements are shown schematically in Fig. 4.2. If these conditions are satisfied, the track is extrapolated to the rest of the tracking volume. Hits found within the CFT are added to the track hypothesis as long as the resulting track hypothesis's $\chi^2 < 16$ condition is still met. If multiple hits match a track hypothesis in a given layer, then those track layers can be split into two separate hypotheses. Each formed track hypothesis must have at least four hits, including both axial and stereo hits. The hit layers are examined, and each track hypothesis is subject to additional requirements from this information. If a layer contributes no hits to a track hypothesis, then that layer is said to be a "miss". Within a track hypothesis, no more than two "misses" can be from the SMT and the total number of misses (N_{Miss}) must satisfy $N_{\text{Miss}} \leq 3$ and $N_{\text{Miss}} \leq \frac{N_{\text{Hits}}}{5}$.

Once a track hypothesis is formed, namely that it has passed the above requirements, it is then added to a final list of track hypotheses. The last stage of track selection is to compare the shared hits among the track hypotheses in the final list. There are two main requirements for hit sharing. The first is that the number of shared hits (N_{Shared}) is less than two thirds of the total number of hits (N_{Tot}) for a given track hypothesis, $N_{\text{Shared}} < \frac{2}{3}N_{\text{Tot}}$. The second requirement is that one of the two following conditions are satisfied:

- $N_{\text{Shared}} \leq \frac{1}{5}N_{\text{Tot}}$.

OR

- $N_{\text{Tot}} - 4 \geq N_{\text{Shared}}$.

If a track hypothesis satisfies these final requirements, then it is added to the list of tracks available for that event.

4.2.2 Primary Vertex Reconstruction

Due to multiple proton-antiproton interactions that happen in any given event, it is crucial to not only be able to reconstruct the primary interaction vertex from track information, but also to be certain that the reconstructed vertex corresponds to the hard scattering process that triggered the event.

The current process to find the primary vertex is performed in a two-pass mode, that is to say that the list of tracks is looped over twice. This procedure [34] requires preselected tracks as inputs. Tracks are considered if:

- Track's $p_T \geq 0.5$ GeV.
- Number of SMT hits associated to the track ≥ 2.
- Track's distance of closest approach (dca) significance (dca/σ_{dca}) ≤ 100.0 with respect to a common vertex (taken to be (0,0) in x–y coordinates).

These tracks are looped over to create a preliminary list of primary vertex candidates. Each primary vertex's position has been determined such that its fit χ^2, based on its associated tracks' impact parameters, has been minimized. The impact parameter (IP) is the minimum distance from a track to the primary vertex, and the smaller the IP, the more likely the track originated from that vertex.

Once a preliminary list of primary vertices has been established, the beam spot information is considered for the determination of track dca. In this pass, tracks still unassociated to any primary vertex candidate are subjected to a much more restrictive selection of $dca/\sigma_{dca} < 3$. Tracks passing this additional requirement are re-fit to each of the primary vertices. Thus the two-pass method leads to more tracks (on average) attached to a given vertex than from using a single pass technique, which is an important factor in the determination of the correct primary vertex. The more high transverse momentum tracks associated to it, the higher the confidence can be placed in its choice as the correct primary vertex.

Now that a list of primary vertices has been established, the next step is to select the *correct* primary vertex. A selection technique is used to determine the primary interaction vertex associated with the hard scattering process. For this technique, it is important to note that tracks coming from soft scattering process and minimum bias events have smaller p_T's than those coming from hard scattering processes. A distribution in $\log_{10} p_T$ of these minimum bias processes is used to determine the probability of a track coming from one of these events. So for each track of $p_T > 0.5$ GeV, a probability is assigned. The probability of the vertex is the product of these individual track probabilities: $\prod_i P^i_{track} = P_{PVz}$. The probability is then weighted to be independent of the vertex's track multiplicity.

4.2.3 Missing Energy Reconstruction

Partons involved in the original hard scattering process have large longitudinal momenta, but their transverse momenta before colliding are approximately zero. Due to momentum conservation, the total transverse momentum from the particles created after the initial collision must then also equal zero. However, whenever a particle escapes the detector and consequently does not deposit its energy, then there is an imbalance of transverse energy. The amount of "missing" energy in an event is energy needed to restore that balance. Missing energy is typically caused by neutrinos, which have a very low probability of interacting within the detector's volume. From measuring the missing transverse energy (E_T^{miss}), it is possible to estimate the amount of energy the neutrino carried. This is done by first calculating the total amount (and direction) of the transverse energy in the calorimeter with respect to the primary vertex z position on a cell-by-cell basis, where all energetic cells are looped over in a given event. Each cell's detector location (in $\eta - \phi$) is used and combined with the primary vertex position information to yield a direction for the energy ($\vec{\eta}^{\text{ cell}}$), where $|\vec{\eta}| \equiv 1$. The transverse direction $\vec{\eta}_T$ is associated to the energy measured in the cell (E^{cell}), such that $\vec{E}_T^{\text{cell}} = E^{\text{cell}} \cdot \vec{\eta}_T^{\text{ cell}}$. The missing energy in an event is simply the negative sum of all cells' transverse energy:

$$\vec{E}_T^{\text{miss}} = -\sum_{i=0}^{N_{\text{cells}}} \vec{E}_T^{\text{cell}_i} \tag{4.1}$$

Splitting up the missing energy into its transverse components can be done from the following relation:

$$\begin{aligned} (E_T^{\text{miss}})^2 &= (E_{T_X}^{\text{miss}})^2 + (E_{T_Y}^{\text{miss}})^2 \\ E_{T_X} &= \sum_{i=0}^{N_{\text{cells}}} E_i \cos\theta_i = -E_{T_X}^{\text{miss}} \\ E_{T_Y} &= \sum_{i=0}^{N_{\text{cells}}} E_i \sin\theta_i = -E_{T_Y}^{\text{miss}} \end{aligned}$$

It is important to note that only energetic cells are included in the above calculations. As noted in Section 3.2.3, cells are subject to a noise suppression of $\sigma_{\text{cell}} > 2.5$, thus any hardware malfunction resulting in an unphysically large cell energy is easily seen as a spike in the missing transverse energy distribution.

4.2.4 EM Object Reconstruction

The DØ calorimeter, as mentioned before, is optimized so electrons and photons will deposit the large majority of their energy in its first four layers. These particle showers most often consist of photons converting into electron-positron pairs and in turn these particles will produce photons through bremsstrahlung. These showers not only are typically constrained to the EM section of the calorimeter, but the transverse width of these showers is small in comparison to jets containing hadrons.

To reconstruct the energy of an EM object at DØ, the energies of the calorimeter's cells, as well as their resulting towers, are used. There are two treatments of calorimeter towers for EM energy reconstruction. The first is the "Simple Cone" (SCone) algorithm [35]. The second is CellNN, and although this second algorithm provides valuable techniques of EM object energy reconstruction at low photon momentum ($E_T^{\gamma} < 10$ GeV), it was not used in this analysis. However, a more in-depth discussion of this technique can be found in Ref. [36]. the SCone algorithm will be briefly overviewed, as it was not only used for EM object energy reconstruction, but its technique of tower energy clustering is very similar for the methods used to determine jet energy depositions in the calorimeter.

Simple Cone Energy Clustering

The SCone algorithm takes a list of energetic towers, ordered in decreasing E_T, and loops over each. Each tower with EM energy over a threshold of 500 MeV is taken to be a seed tower and is removed from the list of towers. A circle of radius 0.4 in η–ϕ is formed about the position of the seed tower, and its energy is taken to be the initial cluster energy. The remainder of the energetic tower list is looped over to find additional towers within this circle. This circle can be equivalently made at all layers of the calorimeter, and because layers farther away from the beam pipe are larger due to their increased radial distance, these circles vary in size and combine to form a cone about the energetic tower. If an additional tower from the list of energetic towers is found to be within this cone, its energy is added to the total energy of the cluster, and the energy weighted position of this larger energy cluster is recomputed. This process continues until the list of towers has been exhausted, yielding a total energy and position for this EM object energy cluster. The cluster is next subjected to quality criteria: it must be composed of at least two towers and it must have an energy exceeding 1 GeV.

This pre-selection is done to further suppress noise and low energy clusters. Once it passes these requirements, this cluster is added to the list of EM objects.

After the first round of EM clustering has finished, the list of recently computed EM objects is subjected to further quality criteria. First, its energy is recalculated for a cone of radius 0.2 (R = 0.2) and this revised energy must pass a threshold of 1.5 GeV. The fraction of cluster's energy from its EM section must satisfy

$$\frac{E^{EM}_{CLUSTER}}{E^{TOT}_{CLUSTER}} > 90\%. \tag{4.2}$$

The cluster must also be well-isolated within the detector, which translates into a dearth of energy in the area surrounding the cluster. DØ uses a fractional isolation, which is formed by subtracting the EM energy in the cluster energy in a cone of R = 0.2 from the total energy in a larger cone of R = 0.4, as shown in Equation 4.3. The schematic of this isolation requirement using reconstruction cones and the cone's mapping for an individual calorimeter EM layer can be seen in Figs. 4.3, 4.4.

$$Iso = \frac{E^{TOT}_{R=0.4} - E^{EM}_{R=0.2}}{E^{EM}_{R=0.2}} \tag{4.3}$$

If these requirements have been satisfied, the cluster is stored to a final list of EM objects for the event and its information is used to compute the values of various EM object identification variables.

Electromagnetic Energy Scale

The energy scale for photons and electrons is computed using the Z boson. The invariant mass peak of the Z is known to high precision, and so by using electrons and positrons from $Z \to e^+e^-$ decays, the absolute peak can be matched. To minimize the width of the Z peak, the energy contributions coming from each layer are chosen as the ones that minimize the width of the Z peak. It is interesting to compare the results of the found layer weights in Run II of DØ compared to Run I. In Run I, since there was significantly less material before the calorimeter, the EM showers reached a maximum energy deposition at the 3^{rd} EM layer. This also explains the interesting segmentation choice of the EM layers in the calorimeter, as they were chosen to optimize the resolution for where the maximum sensitivity was needed. During the upgrade for Run II, a silicon system, fiber tracker, solenoid and layer of lead were

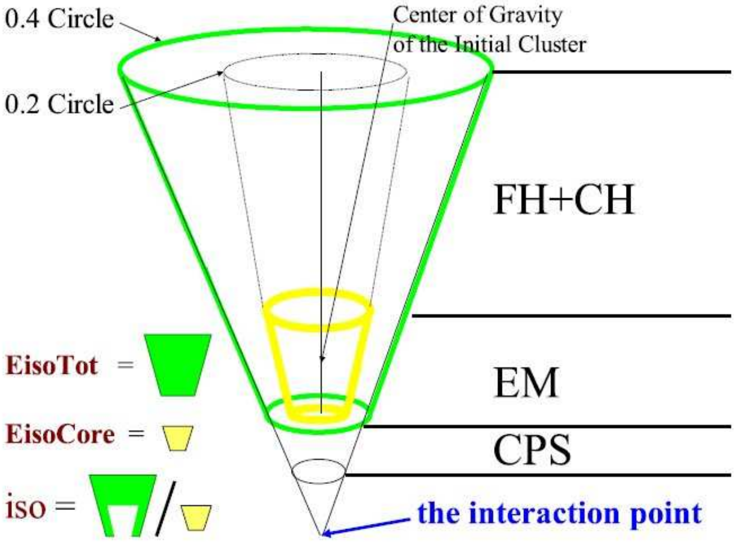


Figure 4.3: Schematic view of DØ's fractional isolation requirement using reconstruction cones.

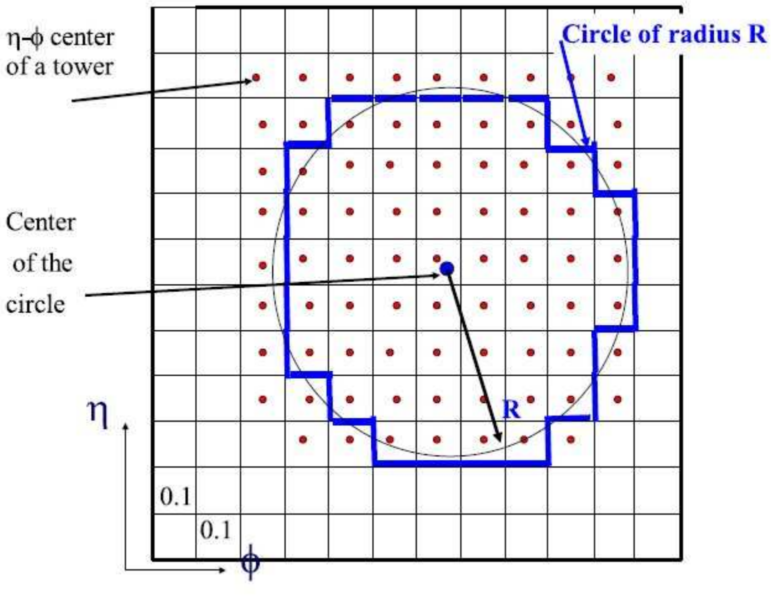


Figure 4.4: Transverse view of a calorimeter layer showing both its segmentation in η–ϕ and the mapping of a reconstruction cone on it.

placed in front of the detector, and as a result the maximum shower layer has moved closer to the 1^{st} EM layer.

There is an additional correction to photons due to a small difference in energy response between photons and electrons. This effect was studied in detail in Ref. [37], and it shows that the EM energy scale systematically over-estimates photon energies. The correction for photons varies from -2.0% for $p_T^\gamma \sim 30$ GeV to 0.1% for $p_T^\gamma \sim 150$ GeV, with a corresponding uncertainty coming from this correction between 0.5–0.9% that depends on p_T^γ.

Basic EM Object Identification Algorithms

The dominant background for electron and photon production comes from dijet events, where one jet deposits a large part of its energy in the EM layers of the calorimeter. Although the total fraction of dijet events having this signature is small, the cross section for dijet production is much larger than that of final states containing photons and electrons. The result is a large number of fake photon candidates, and sophisticated algorithms have been developed to discriminate against these objects. The two most important variables in EM object identification have been discussed before in Section 4.2.4, the isolation and EM fraction requirements.

Besides these, there are two more widely used discriminants for photon selection at DØ. The first is based on the transverse shower width in the 3^{rd} EM layer of the calorimeter and is calculated as follows:

$$\sigma_{r\phi} = \sum_{i=0}^{Ncells} \frac{E_i \times R_{Cal}^2 \times \sin^2(\phi_c - \phi_i)}{E_c} \tag{4.4}$$

where R_{Cal} is the radial distance from the beam pipe to the 3^{rd} calorimeter EM layer, E_i is the calorimeter cell energy, E_c is the cluster energy, ϕ_i is the cell's ϕ position and ϕ_c is the EM cluster's ϕ position. The second discriminant uses the fact that photons do not leave tracks in the detector's tracking volume, whereas background jets typically have multiple tracks associated to their clusters. For a given EM cluster, the list of reconstructed tracks is looped over and a spatial matching procedure is performed. A χ^2 value is computed to determine the quality of the track to cluster fit, and the χ^2 is transformed into a probability. The highest probability of a matched track is kept, and for photons this probability will still be small.

4.2.5 Jet Reconstruction

Jets are reconstructed in the calorimeter using the aptly named Run II Cone Algorithm [38]. This technique is divided into three stages, where the first stage draws many parallels to the SCone method, from Section 4.2.4, used for clustering photon energies. This stage is called preclustering, and its prescription is as follows:

- Form a list of all energetic towers with $E_T > 500$ MeV, ordered in decreasing E_T.
- For a given tower, form a circle of radius [6] ($\mathcal{R}$) 0.3 about its position.
- Loop over all energetic towers within a distance of the seed cluster of $\Delta\mathcal{R} < 0.5$ and add the tower's energy to the seed cluster.
- Add cluster to final list of preclusters if the number of towers is greater than one ($N_{Towers} > 1$).

Once a list of preclusters has been formed, it is re-ordered in terms of decreasing E_T. The second stage of the algorithm takes this list of preclusters and forms proto-jet candidates. This method is detailed below:

- For each precluster P_i, calculate the distance to each of its spatially neighboring preclusters P_j, where $P_i \neq P_j$.
- If $\Delta\mathcal{R}(P_i, P_j) > 0.25$, a proto-jet candidate cone of radius $\mathcal{R} = 0.5$ is formed.
- The list of energetic towers is looped over, and the energy E_T of each tower within this cone is summed with that of the proto-jet and its position is recalculated.
- This process continues until one of the two scenarios occurs:
    1. all towers are looped over;
    2. the position of the proto-jet's position is stable, such that the $\Delta\mathcal{R}$ between two iterations of its position is less than 0.001;
    3. the limit of 50 towers iterated over is reached.

---

[6] Here $\mathcal{R}$ is formed using rapidity (y) instead of pseudorapidity (η), as shown in the equation $\mathcal{R} = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$.

- If the proto-jet's energy is greater than 3 GeV, it is stored.

The third stage of the jet energy clustering involves more precisely determining whether the proto-jet's energy is actually only a part of a larger jet, or whether its energy is actually multiple jets that have been formed into a single jet. To account for these cases, the technique of using midpoints has been developed using the list of found proto-jets from above and it is detailed as follows:

- For each proto-jet, a list of neighboring proto-jets is formed within a distance $0.5 < \Delta\mathcal{R} < 1.0$.
- If there are proto-jets in this list, the clustering procedure is redone, this time without a $\Delta\mathcal{R}$ distance requirement.
- Once the new proto-jet is formed, its tower energy is checked to see if its towers are shared among other proto-jets. If so, the following "Split/Merge" prescription is applied:
    1. If more than 50% of the proto-jet's energy is shared with another proto-jet, then the two proto-jets are merged into one.
    2. If less than 50% of the proto-jet's energy is shared with another proto-jet, then the energy is assigned to the closer of the two proto-jets and removed from the other.
- The final list of proto-jets is subjected to a 6 GeV threshold.

The proto-jets that pass all of these criteria are stored as calorimeter jets.

Jet Energy Scale

The energy resolution ($\frac{\sigma(E)}{E}$) of the calorimeter is given by the equation

$$\frac{\sigma(E)}{E} = \frac{A}{\sqrt{E}} + \frac{B}{E} + C. \tag{4.5}$$

In this equation, there are three parts that describe the energy resolution: the stochastic response ($\frac{A}{\sqrt{E}}$), instrumental effects ($\frac{B}{E}$), and the constant term (C).

The first of these is the stochastic response, which accounts for the fluctuations with a particle's shower development. Due to its energy dependence, this effect significantly contributes over the entire accessible energy range at the Tevatron. The second term in Equation 4.5 is dictated by instrumental effects, such as electronics noise and pedestal fluctuations. Because this term goes as $1/E$, it mainly contributes to the overall energy resolution at low energies. The third term is the constant term, and the largest contributions to its uncertainty come from calibration errors, including higher-order effects such as cell response linearity fitting and gain-switching, non-uniformities in the absorber materials, and the amount of material upstream of the calorimeter. The constant term limits the calorimeter performance at high E_T.

From the above discussion, it is apparent that there are limitations to the calorimeter energy resolution that are inherent to the detector itself. Even when considering particle showers within the calorimeter, there are many effects that must be understood. These include particle energy loss before reaching the calorimeter, the difference in energy response based on particle type, energy from other $p\bar{p}$ interactions convoluted with that of the measured particle, or showering effects of real particles that cannot be easily simulated with the current energy reconstruction algorithms. However, using techniques that parameterize these effects, the true energy of the original jet can be ascertained. Thus, the goal of the jet energy scale is to provide the best estimate of the original particle energy based on the deposition of its energy in the calorimeter [39]. Its true energy $E_{\mathrm{jet}}^{\mathrm{Part}}$ can be parameterized in the following way:

$$E_{\mathrm{jet}}^{\mathrm{Part}} = \frac{E_{\mathrm{jet}}^{\mathrm{Measured}} - E_0(R, \eta, L)}{R_{\mathrm{jet}}(E_{\mathrm{jet}}^{\mathrm{Measured}}, \eta) \times R_{\mathrm{cone}}(R, E_{\mathrm{jet}}^{\mathrm{Measured}}, \eta)} \tag{4.6}$$

where the response (R) is broken down into the calorimeter response to hadronic jets(R_{jet}) based on the amount of measured energy $(E_{\mathrm{jet}}^{\mathrm{Measured}})$ and the fraction of the particle's jet energy within the reconstruction cone (R_{cone}). These effects are also parameterized with respect to η and instantaneous luminosity (L) as this response depends on these quantities.

The jet energy scale can be divided into three separate parts: the fraction of the jet's energy inside the reconstruction cone (R_{cone}), the offset energy [7] (E_0), and the calorimeter response to hadronic jets (R_{jet}). The fraction of the jet's energy inside the reconstruction

---

[7] This is energy that does not come from the jet itself, but is still present within in the jet's reconstruction cone.

cone can be estimated from varying the reconstruction cone size. There are two standard cone sizes used at DØ, $\mathcal{R} = 0.5$ and $\mathcal{R} = 0.7$, and the stability of the energy measurement is studied against pseudorapidity, luminosity and the amount of hadronic energy deposited. The offset energy comes mainly from the underlying event (including overlapping minimum bias events), with additional energy coming from noise effects and pile-up from previous events. To estimate the size of E_0, triggers were set up to select only events without a hard scattering process found to obtain a data sample of pure minimum bias events.

The calorimeter response to hadronic jets can be estimated using the balance of energy found within "γ+jet" events. Because the EM energy scale can be determined using the higher precision energy measurements of electrons, the subsequent response of photons is well known. Thus, the balanced momentum of the photon-jet system can be transformed as:

$$\vec{p}_T^{\,\gamma} + \vec{p}_T^{\,\mathrm{Had}} = 0 \rightarrow R_{EM}\vec{p}_T^{\,\gamma} + R_{\mathrm{Had}}\vec{p}_T^{\,\mathrm{Had}} = \vec{E}_T^{\mathrm{miss}} \tag{4.7}$$

where R_{EM} is the photon energy response, R_{Had} is the hadronic energy response, and $\vec{E}_T^{\mathrm{miss}}$ is the missing transverse energy vector in the event (Equation 4.1). The missing transverse energy is corrected to account for the photon energy response:

$$\vec{E}_{T_{\mathrm{CORR}}}^{\mathrm{miss}} = \vec{E}_T^{\mathrm{miss}} + (R_{EM} - 1)\vec{p}_T^{\,\gamma} \tag{4.8}$$

The calorimeter response to hadronic jets can be written solely in term of the corrected missing transverse energy, the photon momentum, and the direction of the photon's measured calorimeter energy ($\vec{\eta}^{\,\gamma}$):

$$R_{\mathrm{Had}} = \frac{1 + \vec{E}_{T_{\mathrm{CORR}}}^{\mathrm{miss}} \cdot \vec{\eta}_T^{\,\gamma}}{p_T^{\gamma}}. \tag{4.9}$$

The final response is determined from simulated and data "γ+jet" events where the photon and jet are required to be back-to-back for a proper energy balance. From this response, the size of correction to hadronic jet energy is given.

4.2.6 Heavy Flavor Jet Identification

Once the primary vertex and jets are fully reconstructed in an event, their information is used to identify those jets coming from hadronized heavy flavor quarks. There are many algorithms that have been developed within the DØ collaboration to optimize the selection of these jets. Some of the resulting discriminants rely on track-based techniques, and others rely on variables coming directly from reconstructed displaced vertices. Before diving into these algorithms, the general method of secondary vertex reconstruction is overviewed, setting the foundation for both the track- and vertex-based discrimination techniques[8] used at DØ.

Secondary Vertex Reconstruction

Bottom and charm hadrons have lifetimes long enough such that their decays occur measurable distances from the primary interaction vertex. The tracks from their decays originate from a secondary vertex that is displaced with respect to the primary one. These tracks contribute a large χ^2 contribution to the fitting of the primary vertex due to the fact that they naturally have a larger IP than those tracks coming directly from the hard scatter. Thus a similar technique is employed to reconstruct additional vertices within jets from the event [40]. In this procedure, tracks are preselected to be found within the reconstructed jet and each track must have a $|dca_{xy}| < 0.15$ cm, $|dca_z| < 0.4$ cm, $p_T > 0.5$ GeV, and at least 2 SMT hits. From this pool of tracks, the reconstruction of the secondary vertices begins:

- Select tracks with large IP significance ($|S_{IP}| > 3.0$).
- From these selected tracks, find all 2-track seed vertices.
- Attach additional tracks pointing to seeds based on the χ^2 of the vertex.
- Select vertices based on decay length, collinearity angle, and vertex χ^2.

It should be noted that this procedure produces two sets of lists as an output, one of vertices formed only from tracks with $S_{IP} > 0$, and the other from tracks with $S_{IP} < 0$. Physically, a track with $S_{IP} > 0$ would point to a decay vertex on the same side of the primary vertex as the jet, whereas a track with $S_{IP} < 0$ would point to one on the opposite side.The vertices composed of tracks with negative impact parameter significance are called “negative tagged” vertices, and these reconstructed vertices are a result of mismeasurements due to the tracking resolution. Jets with these “negative tagged” vertices are mainly light flavor.

Once these lists of secondary vertices have been formed, the next and final step is to ensure that the position of each reconstructed vertex is found to be within the jet cone. This

---

[8] Because these techniques were optimized with respect to b jets, those jets passing the selection criteria of these algorithms are called “b-tagged” jets. This phrasing will be used throughout the text, and is synonymous with “heavy flavor jet selection”.

is accomplished by imposing the requirement of $\Delta\mathcal{R}(vtx, jet) < 0.5$. The secondary vertices that pass this criterion are associated to the jet and their information can be assessed to better select heavy flavor jets.

Secondary Vertex Tagger

The secondary vertex reconstruction technique provides multiple interesting quantities to enhance the fractions of heavy flavor jets in a given data sample. The ones providing the most discrimination are provided in various combinations from the Secondary Vertex Tagger (SVT) [41]. The individual discriminants comprising the SVT are briefly overviewed below.

The transverse decay length of the secondary vertex (DL_{XY}) is calculated as the difference of the primary vertex's and the secondary vertex's transverse positions. DL_{XY} and the uncertainty associated with it ($\sigma_{DL_{XY}}$) are used to compute its transverse decay length significance as $S_{DL_{XY}} = \frac{DL_{XY}}{\sigma_{DL_{XY}}}$. A minimal requirement on this quantity translates to a cut on the decay length significance such that $S_{DL_{XY}} > 2.5$. Of course there are other discriminants coming from each secondary vertex, including its fitted position's χ^2 value, the total number of tracks associated to it, and its reconstructed mass. Discrimination can also be found using the number of reconstructed secondary vertices (only from the list of vertices constructed from tracks with $S_{IP} > 0$). Using a combination of these variables, more stringent conditions on the jet can be imposed to further enhance the fraction of heavy flavor jets in the data sample.

Counting Signed Impact Parameter

The Counting Signed Impact Parameter (CSIP) [42] is the most basic algorithm used at DØ for heavy flavor jet identification. It uses, much like the secondary vertex reconstruction, tracks with large $|S_{IP}|$ values to form the basis of its discrimination. Tracks associated to the jet are subjected to a preliminary set of selection criteria, including a track p_T cut of 1.0 GeV, at least two SMT hits associated to the track, and an $|dca_{xy}| < 0.2$ cm. In the list, each track's S_{IP} is computed. To optimize the discrimination of this variable, a scale factor a was introduced: $S_{IP} \rightarrow RS_{IP} = S_{IP}/a$, where the optimal value of a was found to be 1.2. To select a jet as heavy flavor, it must have at least two tracks with $RS_{IP} > 3$, or at least three tracks with $RS_{IP} > 2$. Alternatively, a combined single discriminant can be formed using the CSIP information. This is achieved by summing the weighted number of tracks with $S_{IP} > 3$ with the weighted number of tracks with $S_{IP} > 2$ together [43]. The weights applied to each number were optimized to provide the best discrimination.

Jet Lifetime Probability

The Jet Lifetime Probability (JLIP) [44] is similar to CSIP in that they both are track-based b-tagging algorithms. Tracks associated to the jet of interest are subjected to the quality criteria of $p_T > 1$ GeV, an $|IP| < 0.15$ cm, and its $|S_{IP}| < 50$. These selected tracks are inserted into a list and are separated into 29 different track categories based on their qualities, such as the number of total detector hits they contain, their fitted χ^2 value to the primary vertex, and others. An impact resolution function $\mathcal{R}(S)$ was developed for each of these categories, and it is the sum of four separate Gaussian functions. A probability is determined based on its S_{IP}, its track category, and its resulting $\mathcal{R}(S)$. The probability is calculated using Equation 4.10.

$$P(S_{IP})_{track} = \frac{\int_{-50}^{-|S_{IP}|} \mathcal{R}(S)dS}{\int_{-50}^{0} \mathcal{R}(S)dS} \tag{4.10}$$

Using the probabilities calculated for each track, an overall probability of the jet can be determined using Equation 4.11. It is worth noting that two probabilities are actually formed here, one only from tracks with $S_{IP} > 0$, and the other only from tracks with $S_{IP} < 0$.

$$P^{\pm}_{Jet} = \overset{\pm}{\prod} \times \sum_{j=0}^{N^{\pm}_{tracks}-1} \frac{-\log(\prod^{\pm})^{j}}{j!}, \qquad \overset{\pm}{\prod} = \prod_{i=1}^{N^{\pm}_{tracks}} P^{i}_{track}(S^{IP>0}_{IP<0}) \tag{4.11}$$

The probability for a jet is defined such that the smaller the value of P_{Jet}, the less likely it is to have originated from the primary vertex. For light and "negative tagged" jets, this probability distribution is mainly flat; however, for charm and bottom jets their distributions tend toward zero.

As the number of interactions per bunch crossing is typically greater than one, it is very possible for reconstructed jets to contain tracks coming from other collisions. These tracks will appear displaced with respect to the primary interaction vertex of its associated jet, and can enhance the probability of a light flavored jet to be b-tagged, especially when using track-based algorithms. An additional probability was developed using JLIP, one that is by construction less sensitive to this effect. *Reduced* JLIP is calculated in the same fashion as JLIP with the exception that the lowest value of P_{track}, the track least likely to be associated to the primary vertex, is omitted from the calculation.

b Jet Neural Network

With the abundance of discriminants developed for b-tagging, an effort was put forth to combine these variables into a single discriminating quantity, namely the output of a neural network [43, 45]. The general neural network structure is discussed in Appendix D, and the one used in this case has 7 input nodes, 24 hidden nodes, and 1 output node. It was trained using *b* jet events for the signal class (set to one) and light jets for the background class (set to zero). Additionally, there was a stipulation in the training such that if an event did not pass a logical "OR" of the following three conditions, then the event's output was set to zero. These conditions are as follows:

- The transverse decay length significance is greater than 2.5: $S_{DL_{XY}} > 2.5$ (SVT)
- $P_{Jet} < 0.02$ (JLIP)
- Weighted combination of all tracks' S_{IP} to be greater than 8 (CSIP)

Listed below are the seven input variables used to feed the neural network, coming from the CSIP, JLIP and SVT tagging algorithms. They are listed below in decreasing order of their discrimination power as determined by the optimal weights for the b Jet Neural Network (bNN):

1. $S_{DL_{XY}}$ (SVT)
2. Weighted combination of all tracks' S_{IP} (CSIP)
3. P_{Jet} (JLIP)
4. Fitted secondary vertex's χ^2 per number of degrees of freedom (SVT)
5. N_{tracks} associated to secondary vertex (SVT)
6. Secondary vertex reconstructed mass (SVT)
7. Number of secondary vertices in jet (SVT)

The resulting bNN output, shown in Fig. 4.5, shows the *b* jet sample's output peaking sharply at one and that of the light jet sample largely collecting near zero. The output of *c* jets, whose hadrons typically have roughly half the lifetime as those of *b* jets, have a tendency to either mimic the *b* jet sample or the light jet sample in any given event.

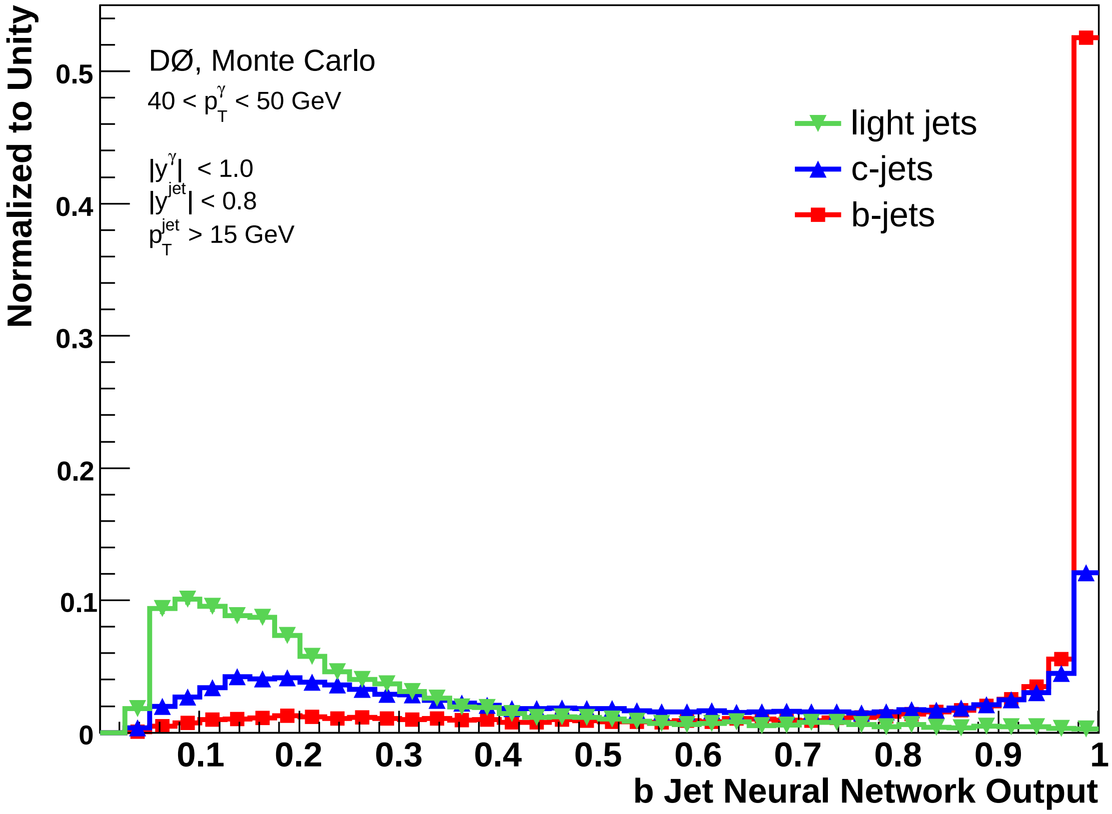


Figure 4.5: The output for the b Jet Neural Network (bNN) is shown for simulated light, c, and b jets. In this plot, the zero bin has been suppressed.

CHAPTER 5

ANALYSIS

Now the foundation has been laid for this measurement: there is a theory that predicts the measured cross section, a powerful accelerator providing the necessary collisions, and an intricate system to collect and translate data from the detector into physically relevant quantities. The next step is to utilize these physical objects and form a measurement of the production cross section.

This is accomplished in three basic steps:

- Selection of signal events
- Efficiency calculations for signal events to pass selection
- Signal purity estimation in data sample

The selection criteria are optimized to identify signal events and maximize the rejection against background events, and it is broken into three parts: event, photon, and jet selections. Their corresponding efficiencies are calculated for the individual selection requirements. The purity estimates measure the amount of signal photons and signal heavy flavor jets (both b and c jets separately) in the data sample that has passed all selection requirements.

The selection efficiencies and purity estimates were performed in each region of p_T^γ, with five bins corresponding to [30–40, 40–50, 50–70, 70–90, 90–150 GeV]. The rapidities of the photon and jet were each divided into two regions, $y > 0$ and $y < 0$. To increase the statistics and take advantage of the symmetry of the photon-jet system, two combinations of these rapidity regions were implemented. Region 1 is defined as when $y^\gamma \cdot y^{\mathrm{jet}} > 0$, and Region 2 is when $y^\gamma \cdot y^{\mathrm{jet}} < 0$.

5.1 Event Selection

Events are selected if they satisfy the following criteria:

- At least one EM trigger requirement from the list given in section 5.4.1 must have been satisfied in the event.
- The primary vertex has at least three associated tracks and $|PV_z| < 35$ cm.
- There is least one photon with $p_T^\gamma > 30$ GeV and $|y^\gamma| < 1$. DØ fiducial criteria are applied to avoid inter-calorimeter boundaries and cracks[1].
- The photon candidate satisfies the selection criteria described in Section 5.4.4.
- The leading jet has $p_T^{\text{jet}} > 15$ GeV and $|y^{\text{jet}}| < 0.8$. It must also satisfy the jet identification and b-tagging requirements described in Section 5.4.5.
- The photon candidate and the leading jet are spatially well separated $\Delta\mathcal{R} > 0.7$.
- To suppress background from $W \to e\nu$ and cosmic ray events, the missing transverse energy in the event is required to fulfill $E_T^{\text{miss}} < 0.7 \cdot p_T^\gamma$.

5.2 Cross Section Calculation

It is important to look how the physical observables that the signal selection criteria utilize translate into the measured cross section. This is achieved by combining acceptance ($\mathcal{A}$), integrated luminosity ($\mathcal{L}$), trigger efficiency (ϵ_t), photon and jet selection efficiencies ($\epsilon_s^\gamma, \epsilon_s^{\text{jet}}$), primary vertex efficiency (ϵ_{PV}), b-tagging efficiency (ϵ_b^{jet}), E_T^{miss} efficiency ($\epsilon_{E_T^{\text{miss}}}$), photon and b (c) jet purities ($\mathcal{P}_\gamma, \mathcal{P}_{b(c)}$) and the unsmearing factor ($\mathcal{U}$) to evaluate the triple differential cross sections as shown in Equation 5.1.

$$\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_T^\gamma \mathrm{d}y^\gamma \mathrm{d}y^{\text{jet}}} = \frac{\mathcal{U}\ N_{\text{Events}}\ \mathcal{P}_{b(c)}\ \mathcal{P}_\gamma}{\Delta y^{\text{jet}}\ \Delta y^\gamma\ \Delta p_T^\gamma\ \epsilon_t\ \epsilon_s^\gamma\ \epsilon_s^{\text{jet}}\ \epsilon_{pv}\ \epsilon_b^{\text{jet}}\ \epsilon_{E_T^{\text{miss}}}\ \mathcal{L}\ \mathcal{A}} \tag{5.1}$$

The following sections will explain each of these pieces in more detail to provide a complete picture of the measured cross section, starting at the beginning with the selection efficiency and their resulting purities.

---

[1] These criteria are accounted for in the photon acceptance.

5.3 Acceptance

An important factor that must be addressed when using data from DØ is the efficiency of the detector itself to measure signal events. This is especially true for photons as their showers are spatially small. To understand this impact, the amount of photon signal lost to the constraints imposed by the detector geometry has been thoroughly investigated. These studies test the efficiency of only fiducial areas of calorimeter cells to be used in EM cluster energy reconstruction. These fiducial requirements avoid low efficiency calorimeter inter-module areas as well as the outer edges of the calorimeter walls (in η) where this energy reconstruction is much less robust, and these constraints result in a much more stable and reliable energy measurement.

The acceptance correction is determined from simulated "$\gamma+$ jet" events. Simulated particles in this analysis are created from a leading order Monte Carlo program, PYTHIA [46], and then using the detector simulation program GEANT [47], these particles are fully reconstructed. To calculate the acceptance correction, the number of simulated photons is counted (N_{gen}). These photons are matched to their reconstructed calorimeter clusters, and the number of matched photons to clusters is counted (N_{match}). The fraction of matched clusters to the number of photons indicates the fraction of photons that are properly reconstructed, which is the acceptance. The acceptance correction (A) is calculated in every bin of p_T^γ (ΔP_T^γ), and is written as

$$\mathcal{A}(\Delta P_T^\gamma) = \frac{N_{match}}{N_{gen}}. \tag{5.2}$$

These studies were performed to assess the loss of photon candidates in ϕ and η separately in the "$\gamma+$ inclusive jet" analysis [48]. The resulting acceptances from that analysis are shown in Figs. 5.1.

There are other criteria that reconstructed objects must meet to ensure they contain enough detector information. Specifically, these consist of additional detector geometric constraints, such as a required number of SMT or CFT hits, or a minimum track p_T for track information. One example of this is *jet taggability*, as discussed in Chapter 4. In this case, the efficiency of this requirement also implicates the object's acceptance. For this analysis, these specific situations will be treated as object selection efficiencies and are presented as such in their appropriate sections.

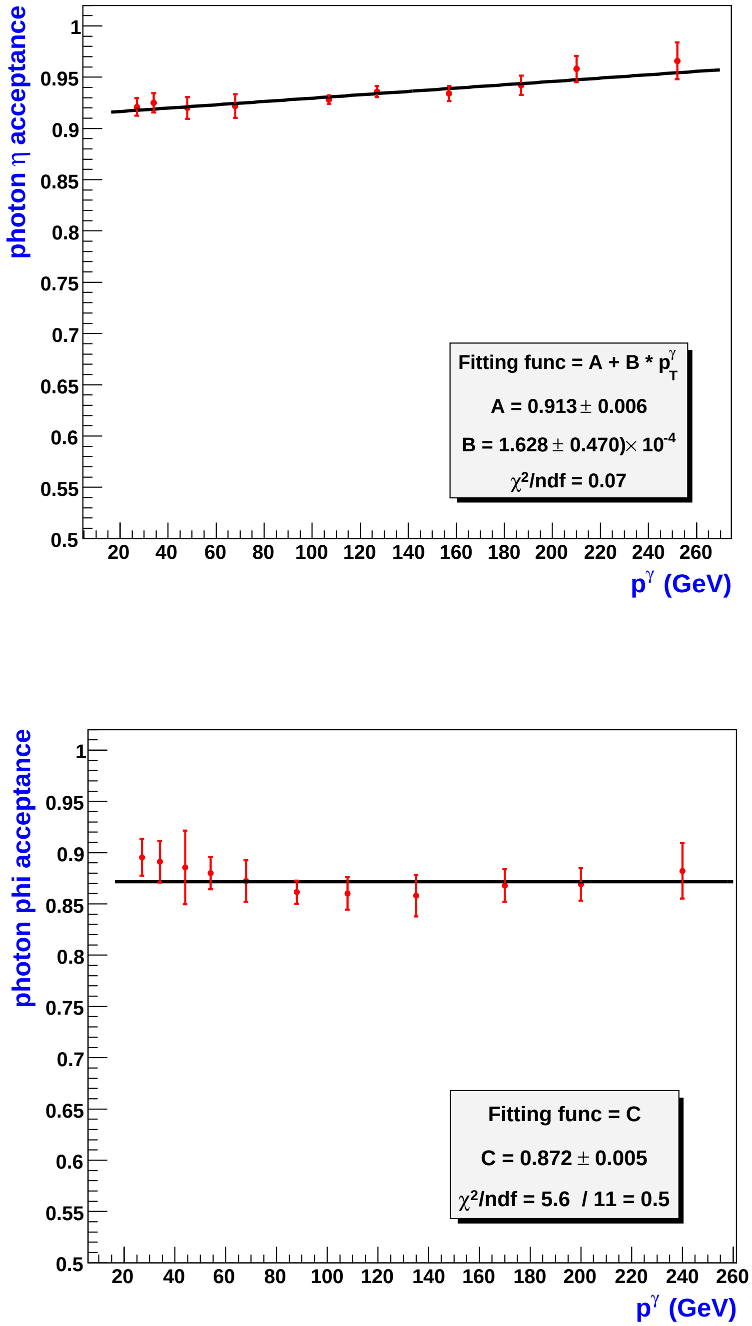


Figure 5.1: The acceptance of photon candidates to be reconstructed as EM objects in the calorimeter with respect to geometric constraints in both η (top) and ϕ (bottom).

5.4 Efficiencies

5.4.1 Trigger and Data Quality

Triggering is performed with respect to the photon candidate's electromagnetic (EM) calorimeter energy distribution, and the triggers themselves are contained in trigger lists. Trigger lists contain all of the triggers used in a given timespan when collecting data, and the triggers within them have been improved over time. The triggers used in this analysis required there to be an EM object in the event, and the specific list of triggers is documented in Ref. [49]. Only one trigger is required to have been satisfied, or fired, for the event to be considered; however, as the differences in trigger are mainly due to the object p_T threshold, many of these triggers overlap and these events will have multiple fired triggers. It is important to also note that the triggers used in this analysis are unprescaled, that is to say the rate at which they fire is not limited due to an overabundance of events, thus avoiding this potential bias.

Requiring this trigger selection is very efficient, $\sim$ 96% for photon candidates with $p_T^\gamma = 30$ GeV, and nearly 100% for photons with $p_T^\gamma > 40$ GeV. Because these efficiencies were derived using $Z \rightarrow e^+e^-$ events [50], an additional systematic uncertainty of 1–2% is assigned to account for possible efficiency differences between photons and electrons.

In addition to at least one fired EM trigger in the event, the quality of the data must be confirmed to be good. This is done using event "flags" that determine if the data are contaminated by either noise or detector malfunctions on an event by event basis. This flag-based data quality criteria used is common for all DØ analyses and it has a 3.2% inefficiency correction. Thus the entire dataset used for this analysis, after applying both trigger conditions and event quality flags, is found to be 1.02 ± 0.06 fb^{-1} [32].

5.4.2 Vertex Selection

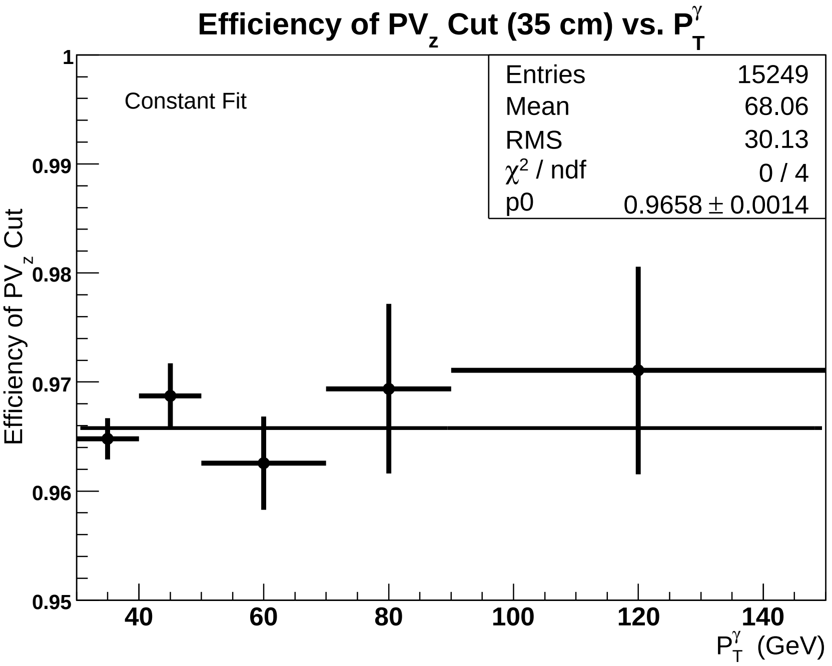
Events are required to have a z vertex position $|PV_z| < 35$ cm from the center of the detector. The efficiency is calculated after both jet and photon selection criteria have been imposed, and the efficiency of all events to pass this PV_z criterion is 96.6%. Fig. 5.2 shows that this efficiency is constant with respect to both p_T^γ and instantaneous luminosity. Due to the geometric constraints required for jet taggability, the distribution of the PV_z becomes more peaked about $PV_z = 0$ than the case without taggability as shown in Fig. 5.4, and results in a more efficient primary vertex selection than in the inclusive "γ+jet" analysis [48]. The primary vertex is also required to have at least three tracks associated to it, $N_{Tracks} \geq 3$. This efficiency has been calculated to be 99.8% and shows no visible dependence with respect to photon p_T (Fig. 5.4). The systematic uncertainties corresponding to the vertex selection vary as a function of p_T^γ and range between 1.4–1.6%.

5.4.3 Missing Transverse Energy

Missing transverse energy (E_T^{miss}) in photon and heavy flavor jet events are expected to be low, whereas two potentially large backgrounds are expected to have large E_T^{miss}. The first of these is the case where a W boson decays into an electron and neutrino (W$\rightarrow e\nu$), and the electron is mis-identified as a photon. Additional jets from this interaction could be heavy flavor, but the neutrino's presence in the event results in large E_T^{miss}. The second case is from a cosmic muon that, through bremsstrahlung, deposits a large amount of energy in the EM calorimeter. If this were to happen within the time-frame of a heavy flavor dijet event, then a photon and a heavy flavor jet would possibly be reconstructed. In this case, there will be an overall imbalance of energy (large E_T^{miss}) because the cosmic muon's energy does not arise from the interaction.

To reject events coming from either of these two cases, a E_T^{miss} limit is imposed. However, due to the increased E_T^{miss} in the semileptonic decays of the b and c jets (Fig. 5.6), the E_T^{miss} condition that was used for the inclusive "$\gamma +$ jet" analysis [48] would remove a significant fraction of signal events: $E_T^{\text{miss}} < 0.35 \cdot p_T^\gamma + 12.5$ GeV. Thus the condition to be used must be sufficiently loose to still retain a high signal efficiency. Applying $E_T^{\text{miss}} < 0.7 \cdot p_T^\gamma$, the signal efficiency is high , and the data can be parameterized with the function $a - b \cdot e^{c \cdot p_T^\gamma}$, as shown in Fig. 5.6. The signal efficiency in simulation for both "$\gamma + c$ jet" and "$\gamma + b$ jet"events is parameterized in the same manner, as shown in Fig. 5.7. If there were a large background present in data from either of these two sources, the efficiency of this cut would be lower in data than in the signal simulation sample. The agreement in efficiencies between data and simulation indicates that the contamination of this background in data is minimal. The maximal difference in efficiency from the data sample to either the b or c jet simulation samples is 1.5% and is taken as the systematic uncertainty for the efficiency of this selection cut.

Figures 5.8–5.10 show the effect of this cut as well as a tighter E_T^{miss} cut from the



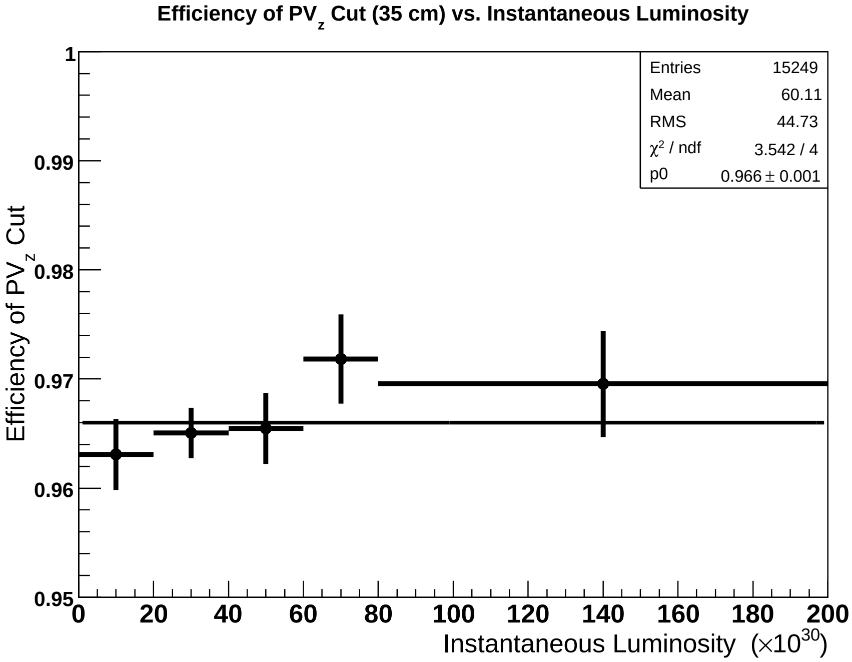


Figure 5.2: Efficiency to reconstruct the primary vertex z position within 35 cm of the center of the detector as a function of p_T^γ (top) and instantaneous luminosity (bottom). In each plot the parameterization of the efficiency (black line) is compared to the binned efficiencies for the data (black circles).

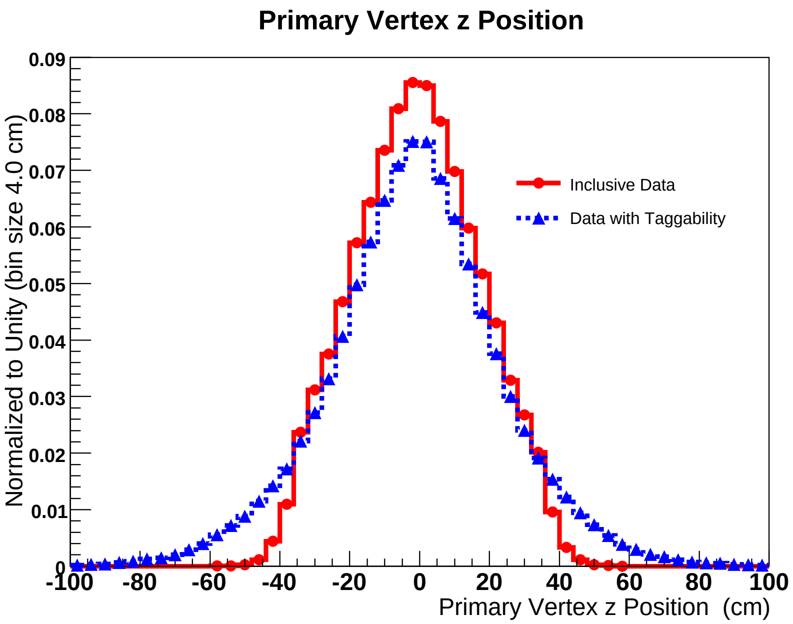


Figure 5.3: PV_z distribution, normalized to unity, for data events from the "$\gamma+$ inclusive jet" sample compared to the data sample in which jets are required to have enough detector information to be tagged as b jets. This jet requirement results in a more peaked distribution about $PV_z = 0$ due to the additional geometric constraints.

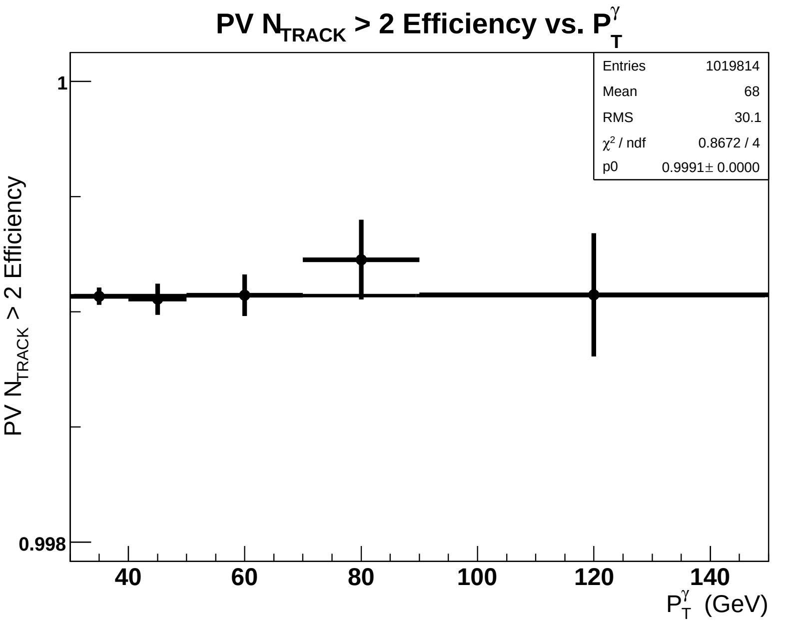


Figure 5.4: Efficiency in data of the event's primary vertex to have at least three tracks associated to it as a function of p_T^γ. The parameterization for this efficiency is a constant (black line), and shows good agreement to the binned data efficiencies (black dots).

inclusive γ + jet analysis on the missing energy distribution. Additional cross checks were performed based on previous studies to identify cosmic ray events [51]. One study showed that cosmic events typically deposit no energy in the third EM calorimeter layer, in contrast to signal photons, which deposit a substantial amount of energy there. The distribution of the fractional energy in the third EM calorimeter layer (Fig. 5.11) shows that no spike is seen at zero in the data. This is further verification that no contamination from cosmic events is expected in the final data sample.

5.4.4 Photon Selection

Photon candidates with transverse momenta greater than 30 GeV and with $|y| < 1.0$ are ordered in p_T and subjected to multiple criteria to reject jets that deposit their energy in the EM calorimeter (EM-like jets). Jets in general consist mainly of light mesons, and in EM-like jets there are high momenta mesons that decay to photons and shower in the EM layers of the calorimeter, imitating the signal of direct photons. These types of mesons are typically created through charge-exchange, such as $\pi^- p \rightarrow \pi^0 n$, where n is a neutron. In this case, the produced π^0 meson almost always decays into two photons, and if the π^0 meson carries enough of the original particle's momentum, then the two photons will be collinear and be reconstructed as a single photon candidate. The final state of two high p_T jets (dijets) has a much larger cross section than that of the direct photon and jet. Thus, even though only a small fraction of jets are EM-like, dijet events containing one EM-jet will greatly outnumber the signal events. The characteristics of EM-like jets, however, differ from direct photons in a few important ways. EM-like jets have a large particle multiplicity from the hadronization of the quark or gluon, compared to the individual direct photons. Thus the energy depositions of EM-like jets in the calorimeter tend to be wider spatially and are less constrained to the EM layers of the calorimeter. Also, charged particles within these jets interact with the CFT and SMT, and tracks are associated with these clusters. Single photons, on the other hand, do not leave tracks as they largely convert to electron-positron pairs in the calorimeter. These EM-like jet characteristics provide the basis for powerful discriminants, essential to the identification and removal of background from the signal photon sample. These criteria have been optimized to retain a very high signal efficiency, while maximizing the rejection of background jets. All photon candidates must meet the following conditions, which were previously discussed in Chapter 4.2.4:

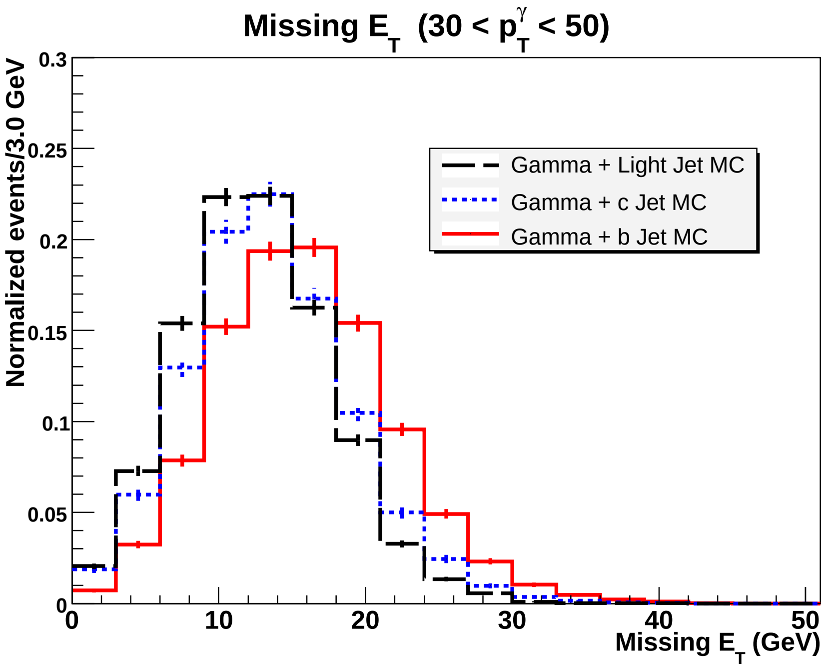


Figure 5.5: E_T^{miss} distribution in simulation for light, c and b jets. The large tail in this distribution for both c and especially b jets is due to the expected enhancement of semileptonic decays.

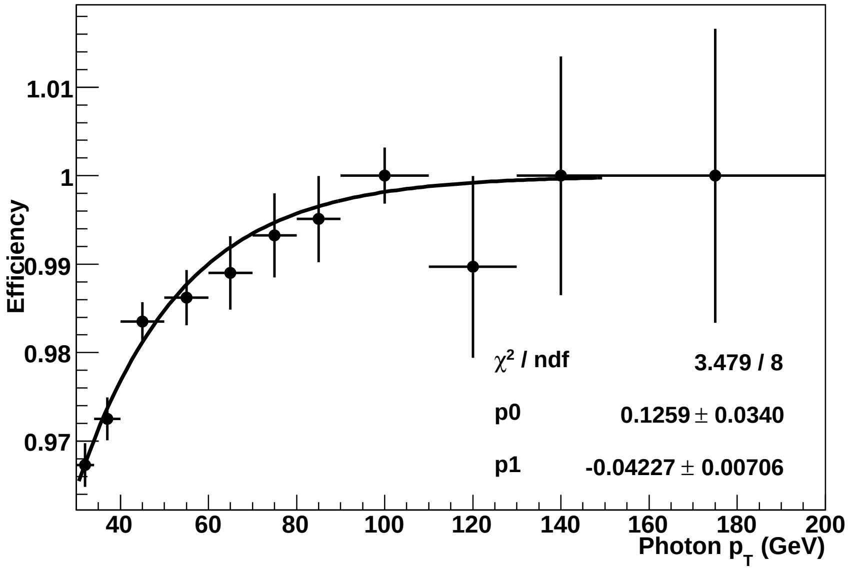


Figure 5.6: E_T^{miss} efficiency in the data sample with the fit parameterization of $1 - p_0 \cdot e^{p_1 \cdot p_T^{\gamma}}$.

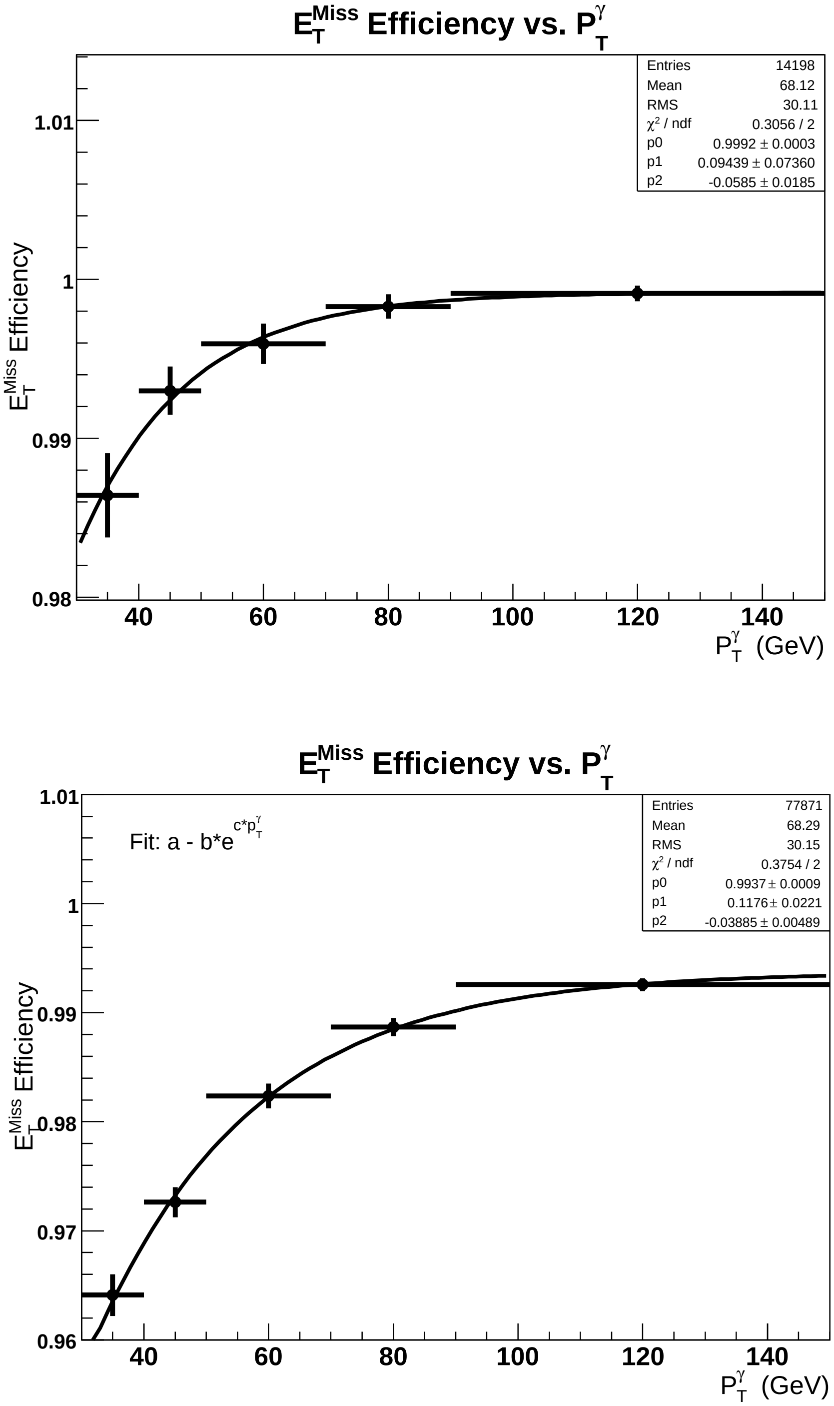


Figure 5.7: Efficiency of $E_T^{\text{miss}} < 0.7 \cdot p_T^\gamma$ in $\gamma + c$ jet simulation (top) and $\gamma + b$ jet simulation (bottom) with their corresponding parameterizations as a function of p_T^γ.

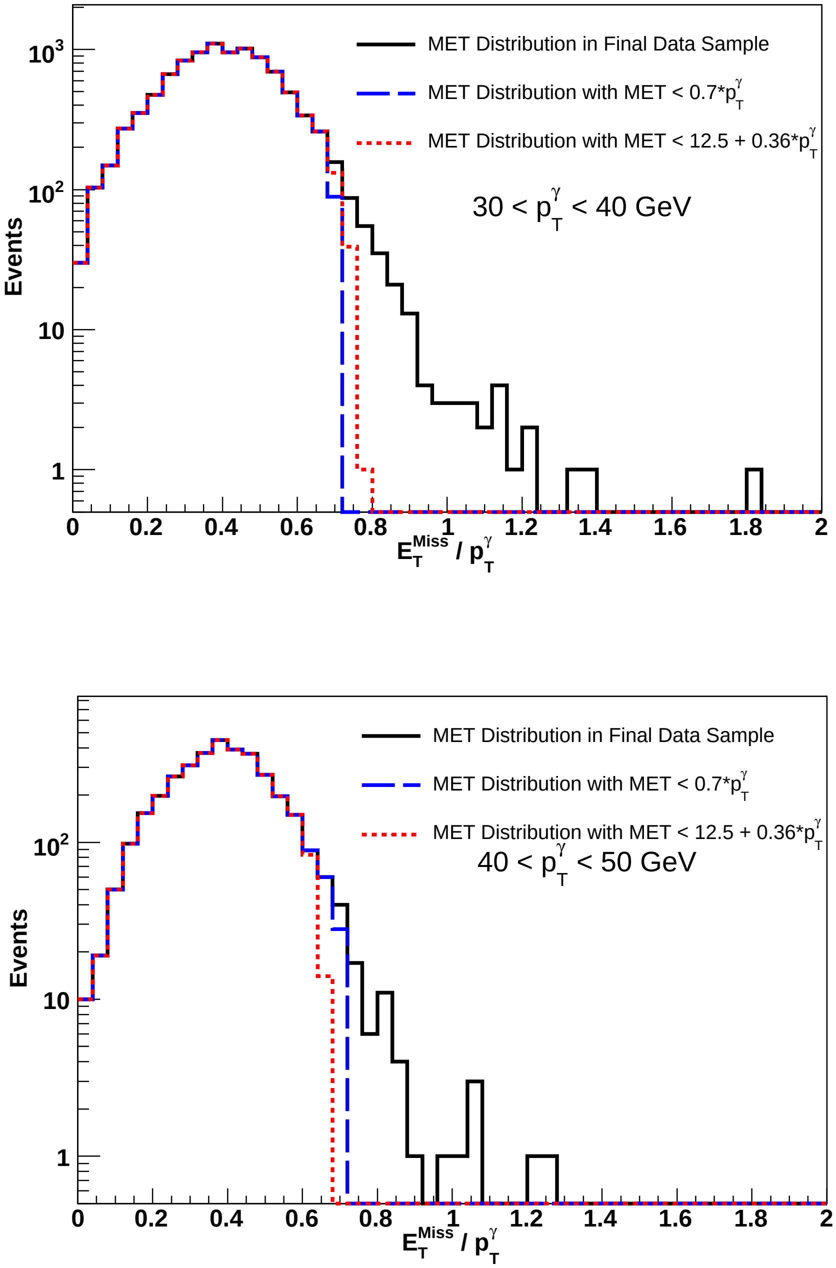


Figure 5.8: The $E_T^{\rm miss}$ divided by the p_T^γ with two separate missing energy requirements imposed for the p_T^γ intervals 30–40 GeV (top) and 40–50 GeV (bottom). Both are effective at removing spurious cosmic events, however the tighter requirement removes additional signal events.

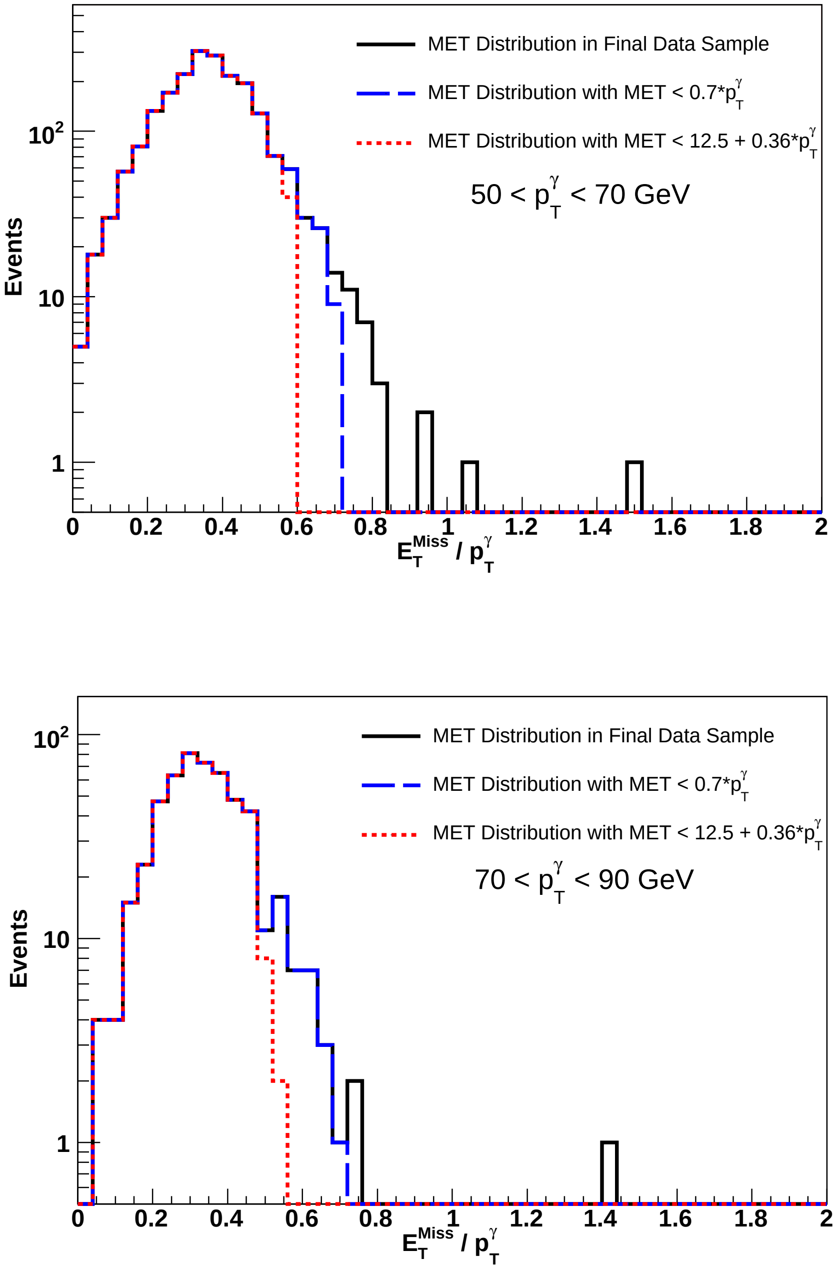


Figure 5.9: Same as in Fig. 5.8, except now the p_T^γ intervals shown are for the 50–70 GeV (top) and 70–90 GeV (bottom) bins.

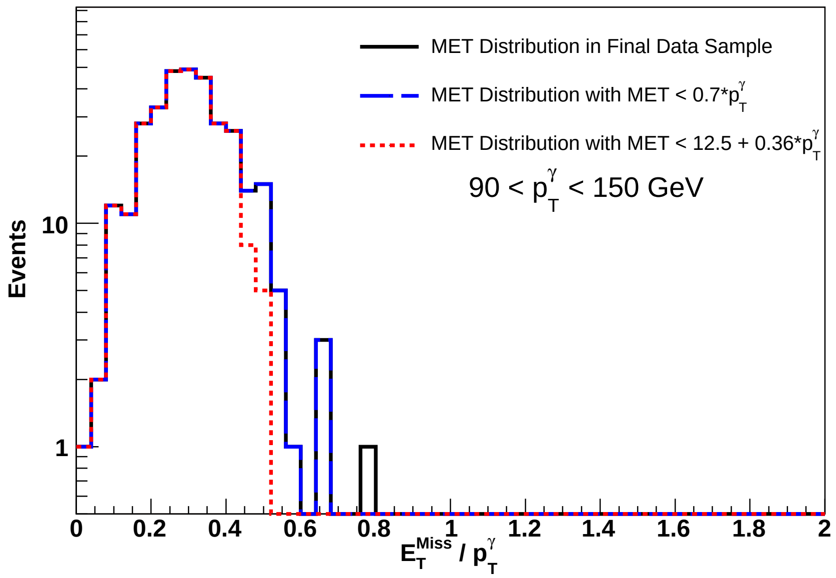


Figure 5.10: Same as in Fig. 5.8, except now the p_T^γ interval shown is for the 90–150 GeV bin.

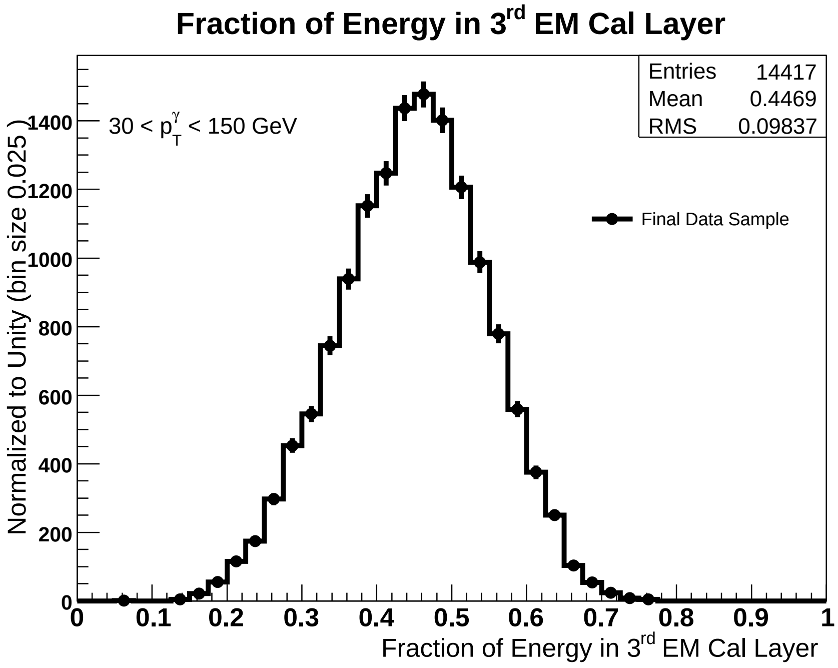


Figure 5.11: Fraction of energy deposited in the 3^{rd} EM calorimeter layer in data. This cross-check confirms that no spike at zero is seen, indicating that there are no remaining signs of cosmic muons in the final data sample.

- Greater than 96% of the photon candidate's energy must have been deposited in the EM layers of the calorimeter (Equation 4.2).
- The photon candidate must be well isolated, such that the surrounding energy must be less than 7% of the photon candidate's energy (Equation 4.3).
- The transverse shower width in the third layer of the EM calorimeter must be less than 14 cm^2 (Equation 4.4).
- There must be no track associated to the photon's energy cluster in the calorimeter, such that the probability of any spatially matched track to it is less than 0.1%.

Photon candidates typically have no tracks associated to them from the primary vertex, and because of this additional measures must be taken that the photon candidates in this sample originate from the chosen primary vertex of the event. In order to reject photons coming from interaction vertices other than the primary vertex, as well as cosmic muons that mimic the isolated photon signature, a "photon-pointing" algorithm was developed [48]. This technique uses the cluster's centroid position found in each layer of the calorimeter independently, and uses their combined positions to "point" back to an estimated primary vertex location (PV_z^{FIT}) for each photon candidate. Additionally, the reconstructed 3-D CPS clusters can be used as an extra layer in this pointing procedure, as the schematic in Fig. 5.12 illustrates. Using these clusters greatly improves the resolution of the found primary vertex z position, and ~90% of photon candidates have at least one associated CPS cluster ($N_{CPS} > 0$). Two cases are considered for the pointing technique as listed below:

- $N_{CPS} > 0 : |PV_z - PV_z^{FIT}| < 12$ cm
- $N_{CPS} \equiv 0 : |PV_z - PV_z^{FIT}| < 35$ cm

Unfortunately, after the application of these selection criteria, there still exists a large amount of background, stemming mainly from dijet event contributions. To both further reject this additional background and to estimate the amount of signal in the final data sample, a dedicated photon neural network (γ-ANN) was developed.

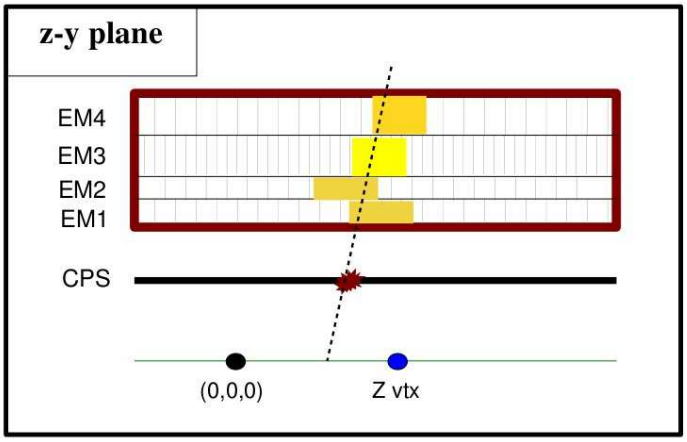


Figure 5.12: Schematic behind the "photon-pointing" technique, used to estimate a photon candidate's correct primary vertex z position.

Photon Neural Network

The γ-ANN was used in this analysis to provide a criterion for increased photon purity, and as a shape template for photon purity determination. Using the general basis for neural networks in Appendix D as a guide, the specific case of the γ-ANN [53] is quickly reviewed. This neural network uses three input nodes, two hidden nodes, and one output node. The signal is simulated photons and the output of these events is set to one. The background consists of EM-like jets, and their output is set to zero. To ensure the stability and convergence of the training procedure, the number of events used to train the neural network are taken to be 20–30$\times N_{ind}$[2]. This choice results in more than 100 patterns per weight, and the final product shows both stability and excellent separation between signal and background events. Before moving too quickly to the end of this process, the input variables should also be understood.

The three input variables that the γ-ANN employs are previously unused photon identification discriminants that have a clear distinction of signal photons to background EM-like jets. Two of these discriminants utilize information coming from the 1^{st} calorimeter EM layer. The first of these two variables takes advantage of the fact that EM showers from photons have a small transverse width compared to EM-like jets, a fact which directly correlates to having fewer energetic cells in the 1^{st} calorimeter EM layer. Fig. 5.13 shows

---

[2] N_{ind} is the number of independent weights and thresholds defined in Equation D.2.

this distribution for both the signal sample and the background EM-like jet sample after the application of 300 MeV cell energy threshold has been applied. This threshold was found to give the best separation between photons and EM-like jets. The second of these two variables uses the layer energy deposition of electromagnetic objects as compared to the energy deposition of EM-like jets. Typically, EM-like jets will have a larger fraction of energy in the first layer of the EM calorimeter than photons, as can be seen in Fig. 5.13.

The third variable provides further suppression of high p_T tracks within the same reconstruction cone as for photons. In this case, the same type of cone that was used for the cluster reconstruction energy in Section 4.2.4 is extended to the tracking volume. This cone has a radius of 0.05 (R = 0.05) in η–ϕ. A second cone of R = 0.4 is also made in the tracking volume, and a loop is made over the entire list of tracks for those falling in between the two cones ($0.05 < \Delta \mathrm{R}_{Track-Cluster} < 0.4$). For all tracks located in this annulus region, their momenta are summed and this total momentum is used as the discriminant. The technique is referred to as "hollow cone isolation", and takes advantage of the large track multiplicity for jets. Fig. 5.14 shows these distributions for both signal photons and for background EM-like jets, and a clear distinction between the shapes of signal and background is seen.

The resulting output of the γ-ANN is shown in Fig. 5.15 for simulated signal photons and background EM-like jets, and data after having passed all other photon identification criteria. The shape of the output for data indicates that a significant contribution of EM-like jets still contaminates the signal sample. To enhance the fraction of events coming from signal photons in the data sample, a photon candidate's γ-ANN output is required to be greater than 0.7 ($O_{\gamma-\mathrm{ANN}} > 0.7$). The efficiency of this selection is sensitive to the amount of material in front of the calorimeter, as more photons will convert to electron-positron pairs with more material present. To test the impact on the signal samples, the amount of additional material in front of the calorimeter in the Monte Carlo simulation was varied between $(0.4\text{–}0.7) \times X_0$. The resulting signal efficiency degraded by 0.5–1.1%. This fluctuation, along with the data to simulation differences from electrons for the efficiency of the γ-ANN cut, combine for a total systematic uncertainty for the γ-ANN of 2.4%.

The overall efficiency of the signal to pass all photon identification criteria is shown in Fig. 5.16 as a function of p_T^γ [49]. The total systematic uncertainty associated to this selection is 4.0–4.8%, mainly coming from the anti-track match χ^2 probability (3%), data to simulation differences (1.5–2%), and the γ-ANN (2.4%).

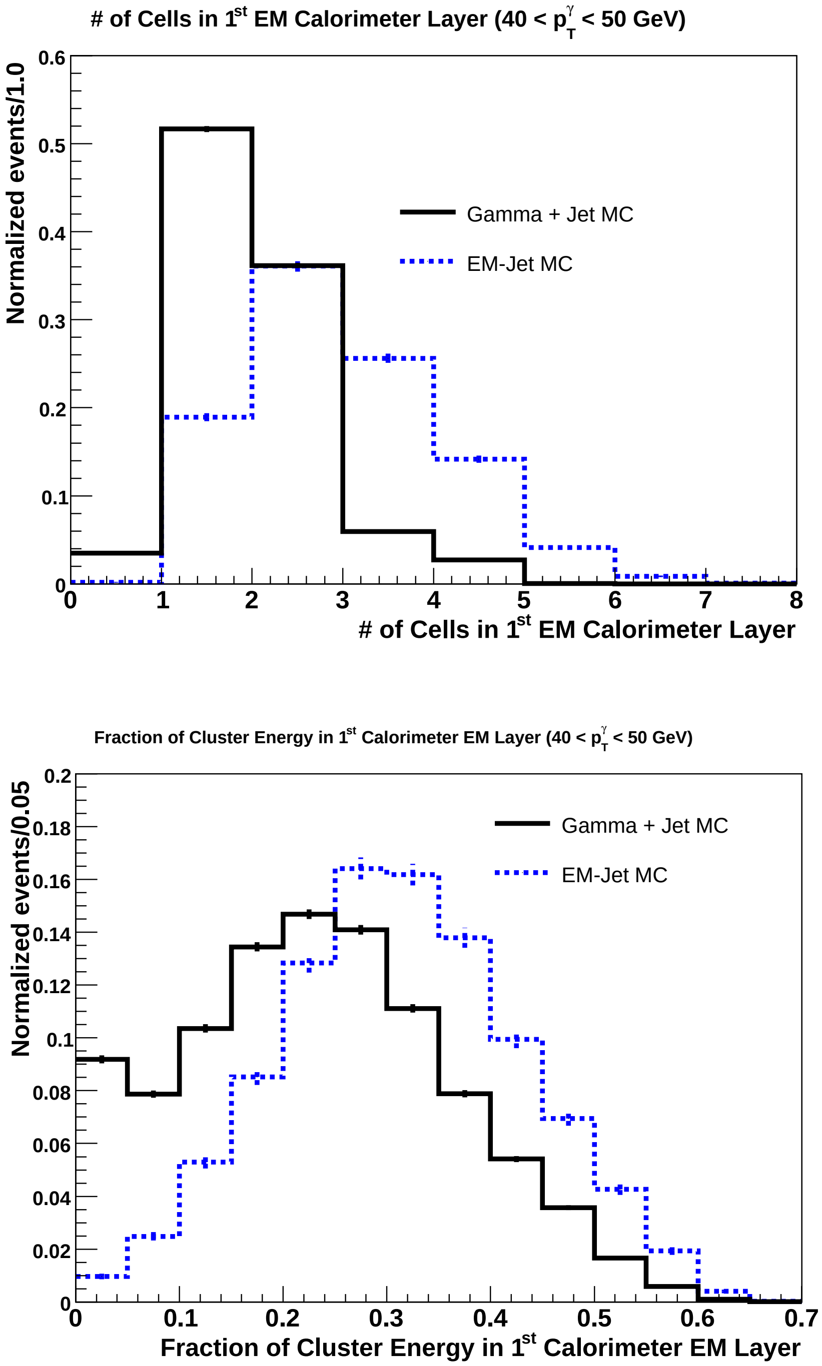


Figure 5.13: Distribution of the number of energetic cells (top) and the fraction of energy deposited (bottom) in the first layer of the EM calorimeter for both signal photons and background EM-like jets.

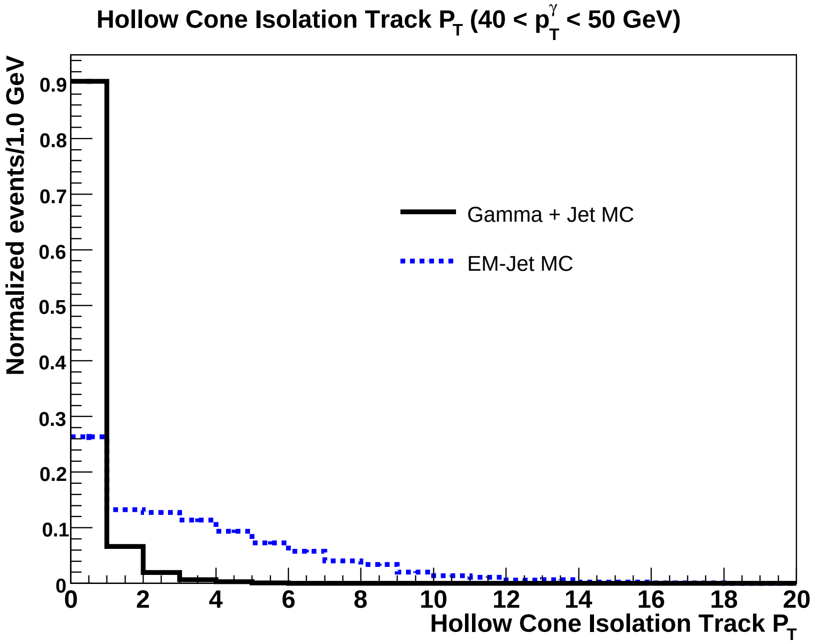


Figure 5.14: Distribution of the total sum of hollow cone track momentum for both signal photons and background EM-like jets.

5.4.5 Jet Selection

Jets are selected if they are located within a rapidity region of $|y| < 0.8$, they have transverse momenta greater than 15 GeV, and they are considered to be of *good* quality, which is to say that their associated calorimeter energy clusters satisfy the following [54]:

- No more than 40% of the jet cluster's energy must be from the calorimeter's coarse hadronic layers.
- At least 5% of the jet cluster's energy must be from the calorimeter's electromagnetic layers.
- The ratio of the jet's energy, as determined from the L1Cal trigger, to its reconstructed energy ($L1_{\text{Ratio}} = \frac{E_{\text{L1Cal}}}{E_{\text{Precision}}}$) must be greater than 0.5.

Dedicated studies were performed to estimate the efficiency of simulated jets to both pass the above criteria and to be reconstructed as calorimeter jets, and the efficiency was taken as the ratio of the number of reconstructed *good* jets over the number of simulated jets that

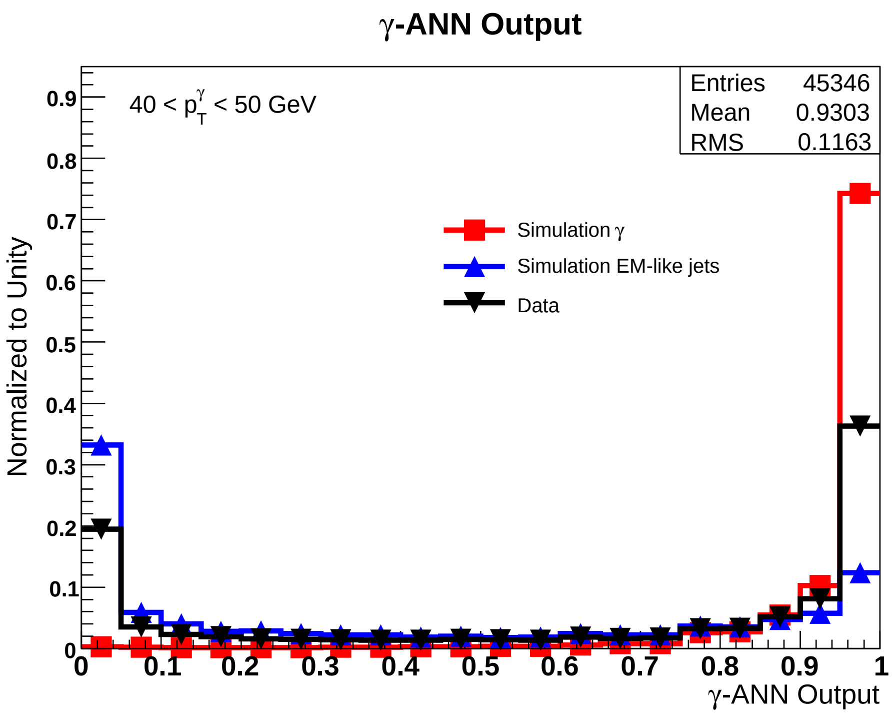


Figure 5.15: γ-ANN output for signal photons, EM-like jets, and the selected data sample. All histograms have been normalized to unity, so that the shape of each can be easily compared. The peak in data at zero is due to a significant dijet background contamination.

were generated. This efficiency also includes the p_T ordering effect on jets. This factor can be important if the highest p_T jet in an event, called the leading jet, has its energy under-measured to the point where it no longer is leading. Because the detector simulation did not reproduce the measured jet energy resolution very well, additional "smearing" [59] of the simulated jet's reconstructed energy was applied. The effect of smearing is an overall increased σ for the Gaussian resolution function of the jet's energy.

The overall efficiency of jet identification is given as its efficiency to be reconstructed (ϵ_{Reco}) multiplied by its efficiency to pass the jet selection criteria (ϵ_{JetID}), such that $\epsilon_s^{\text{jet}} = \epsilon_{\text{Reco}} \times \epsilon_{\text{JetID}}$. This efficiency was studied for inclusive jets (dominated by light flavor jets), c jets, and b jets separately. The resulting parameterized efficiencies for the cases of light and b jets are shown in Fig. 5.17.

Because the "smearing" parameterization used to correct simulated jets energies was tuned for an inclusive jet sample, it is not clear that the same parameterization is correct

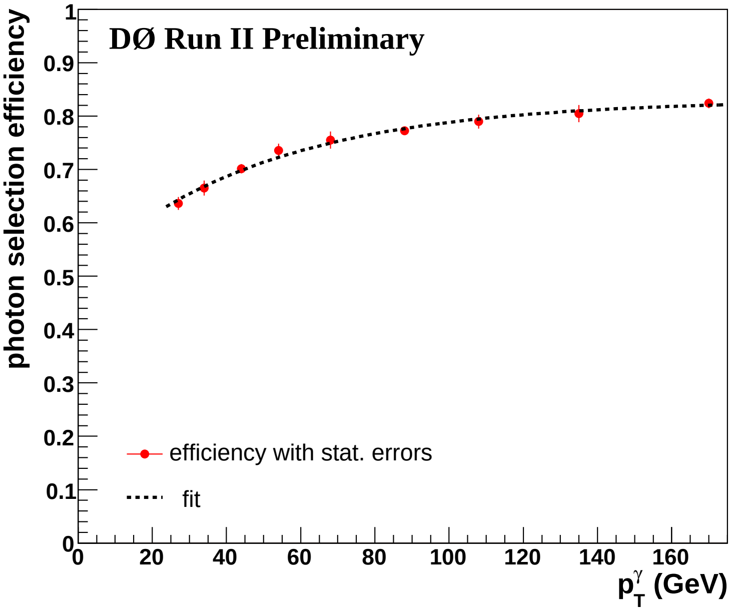


Figure 5.16: The signal efficiency to pass all photon selection criteria as a function of p_T^γ with parameterization.

for the case of b jets. To account for this fact, the b jet energy resolution was smeared by an additional 50% [3]. The resulting ϵ_{JetID} for b jets is lower than that of inclusive jets by 0.5% at low photon p_T^γ and 0.25% for p_T^γ ~150 GeV and is taken as a systematic uncertainty. Also, due to uncertainties in the modeling of the ϵ_s^{jet} efficiency from simulation, half of the difference of $(1 - \epsilon_s^{\text{jet}})/2$ is taken as an additional systematic uncertainty. This uncertainty accounts for possible variations in simulated jet production, such as the specific initial or final state radiation or fragmentation model used in PYTHIA.

The leading jet in the signal sample is required to have $p_T > 15.0$ GeV. The efficiency to pass this threshold was measured in the inclusive "$\gamma +$ jet" measurement, and as was done for that analysis, a 2σ uncertainty from the Jet Energy Scale (JES) is taken. This is a conservative estimate of the uncertainty as the JES correction is an extrapolation for the region of jet $p_T < 15$ GeV [39]. For the b-tagged sample there is also the additional factor of the b jet energy response. In simulation studies it has been shown that b jets have a lower

---

[3] The amount of this additional smearing was taken from the dedicated studies for jet energy corrections in Ref. [39] and this value is considered to be a conservative estimate.

energy response than light jets as seen in Fig. 5.18.

For the b jet case, the jet p_T is smeared by an additional 50% and varied by $\pm 19\% \times$ the jet p_T to determine the uncertainty caused by the jet p_T cut. The value of 19% is derived from adding in quadrature the ratio of energy response of b jets to light jets (17%) and the 2σ JES uncertainty (8%). A similar procedure is performed for c jets where the additional smearing applied is conservatively taken to be 25%. The jet p_T is varied by $\pm$ 13%× the jet p_T, which includes the c jet response ratio to light jets that is taken to be 10% conservatively.

The efficiency for b and c jets to pass this cut is parameterized as a function of p_T^γ and is shown in Fig. 5.19. We estimate the total uncertainties for b and c jet to pass this selection while taking into the uncertainties from the JES and energy response corrections, as noted above. We vary the jet p_T cut by the overall uncertainty on the b jet p_T expected from these effects (Fig. 5.20). The largest difference in efficiency between the central p_T cut and the varied p_T cuts is taken as the systematic error. From comparisons between the efficiency in the inclusive "γ+jet" data and simulation, we expect the actual differences to be much smaller.

5.4.6 Photon-jet distance cut

In events containing signal photons and jets, the photon and jet are spatially well separated. This is a consequence of the isolation requirements for the photon candidate that suppresses the diagrams in which the two would be close together. Thus, the distance d$\mathcal{R}$ between the photon and jet in $\eta - \phi$ is required to be $\mathrm{d}\mathcal{R}(\gamma, jet) > 0.7$. This requirement is highly efficient for the remaining events and is independent of p_T^γ, with $98.5 - 99\%$ of all events satisfying it. Fig. 5.21 shows a normalized distribution of data events in two photon p_T bins, $30 < p_T^\gamma < 40$ and $90 < p_T^\gamma < 150$ GeV.

5.4.7 b-Tagging

Heavy flavor jets are produced at a much smaller rate than light flavor jets (less than $\sim$ 10% of all jets are heavy flavor), and using techniques outlined in Section 4.2.6, the fraction of these heavy flavor jets can be enhanced. For this analysis, the tagging algorithm used is the b Jet Neural Network (bNN). In order for b-tagging algorithms to provide a consistent discriminant, a baseline of information is required for the jet's tracks. Taggability [55] is the requirement that a jet's track information is complete enough for the algorithm to be

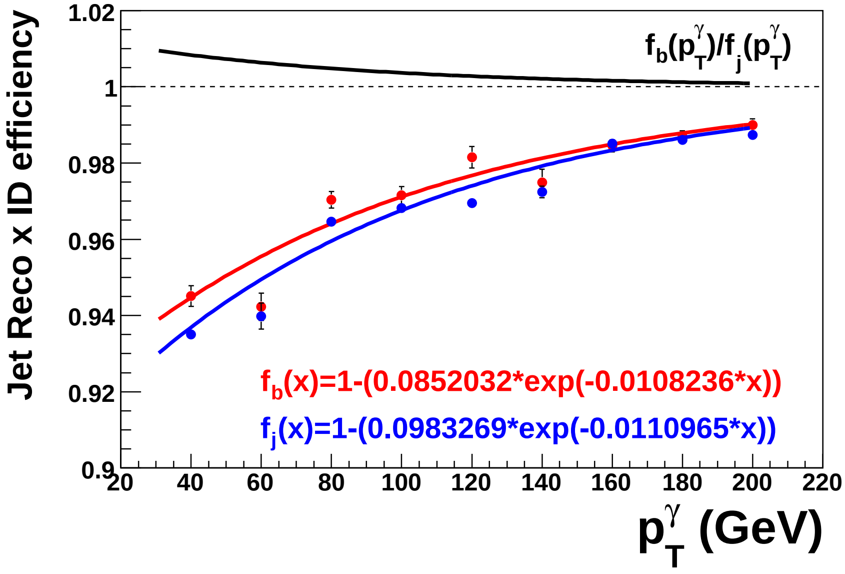


Figure 5.17: Efficiency of jet reconstruction × ID, along with their parameterizations, for inclusive jets (f_j) and jets originating from b quarks (f_b). The ratio of the two efficiencies is shown as the black line, and this difference is $< 1\%$ for all p_T^γ.

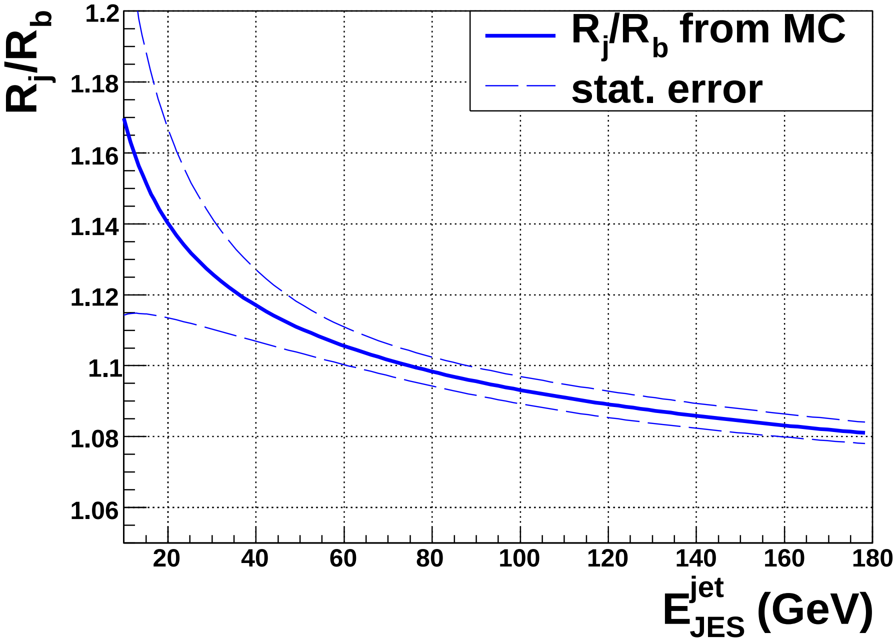


Figure 5.18: The ratio of the b jet response to the inclusive jet response. At 15 GeV, b jets have a $\sim 17\%$ lower response than light jets.

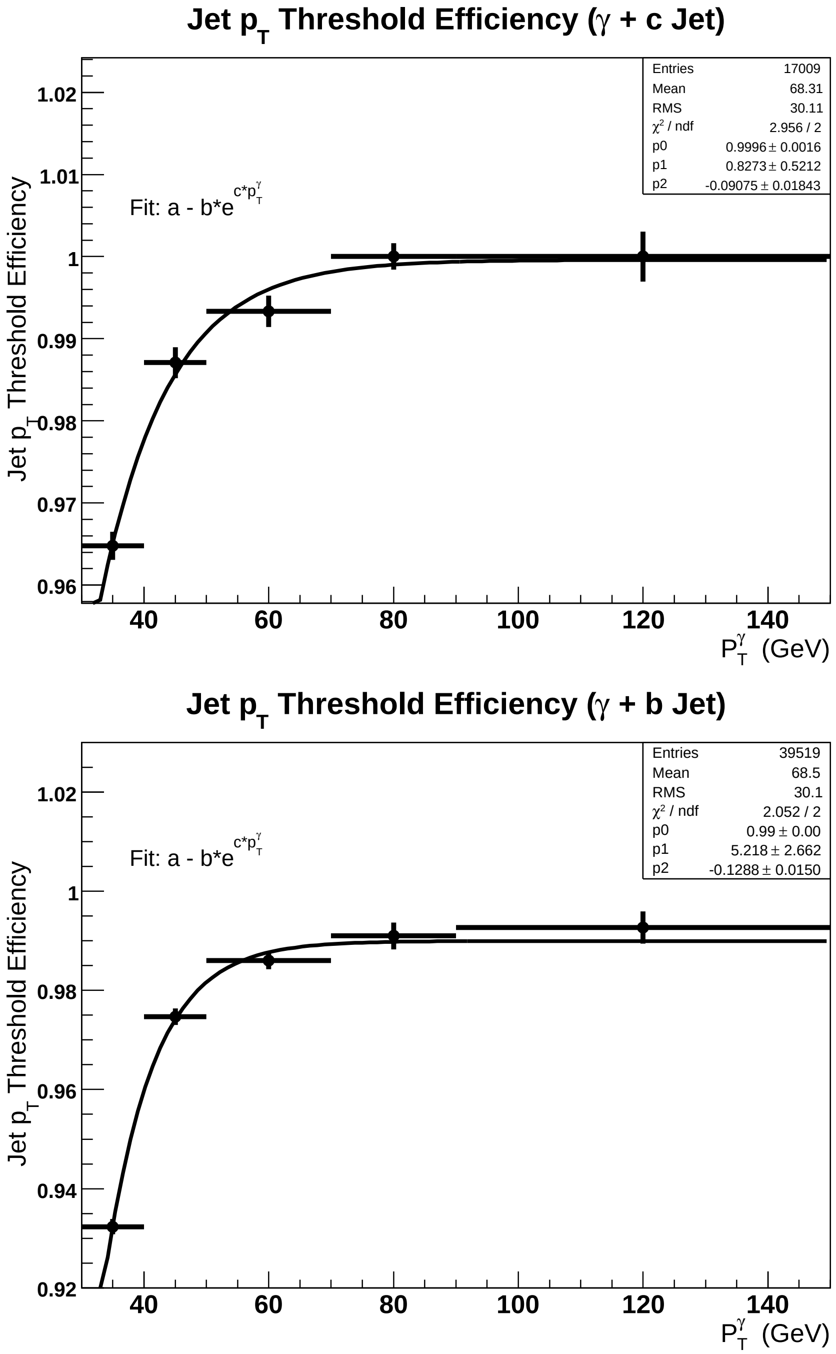
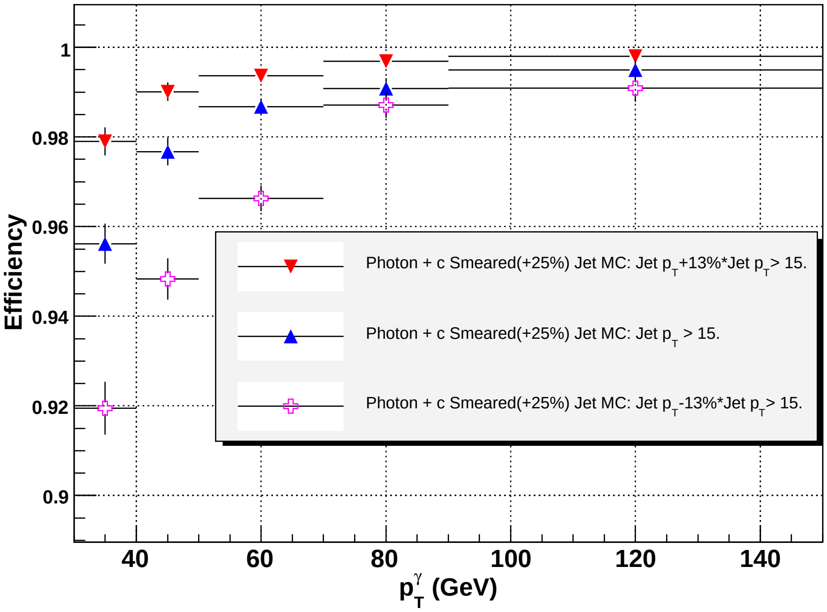


Figure 5.19: The efficiency of c jets (top) and b jets (bottom) in simulation to pass the jet p_T cut of 15 GeV, with parameterizations as a function of p_T^γ.



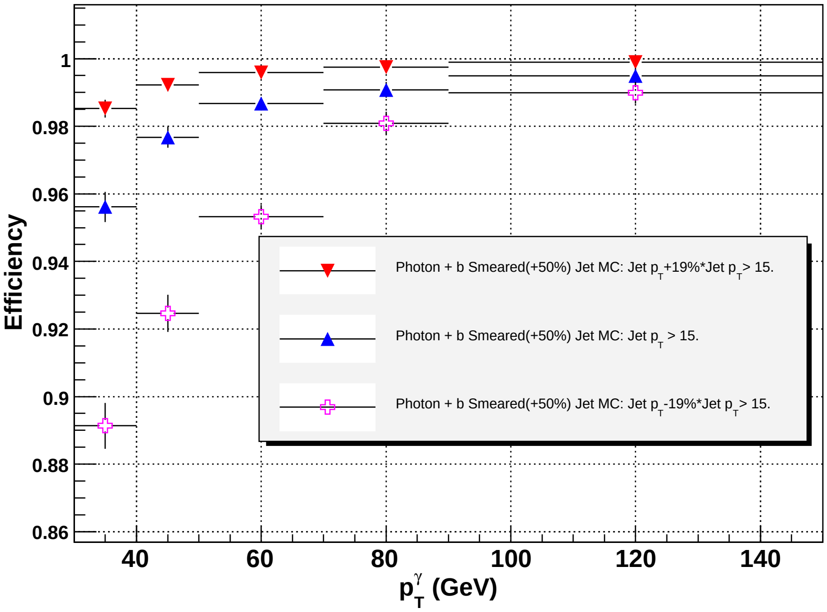
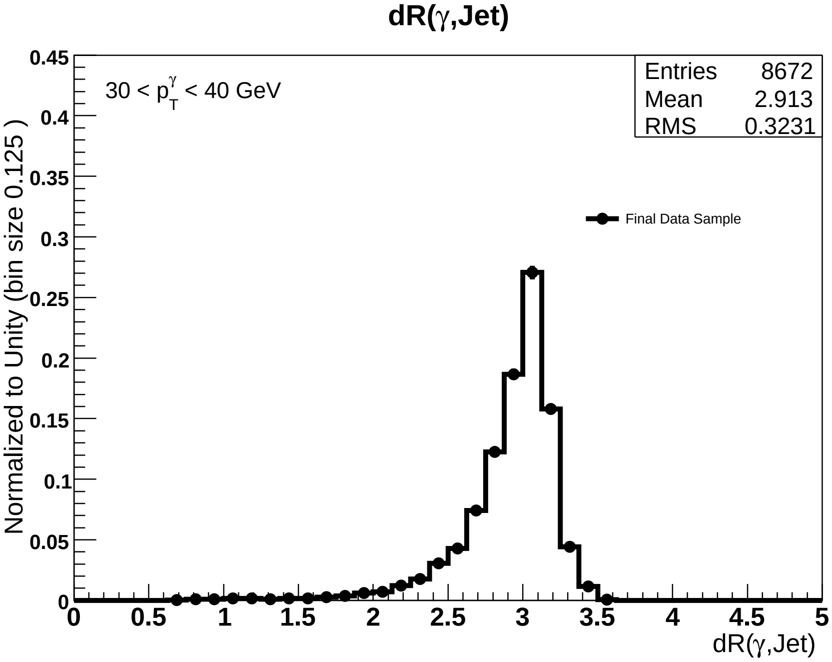


Figure 5.20: The effect of varying the jet p_T cut in simulation by the total uncertainties for both c jets (top) and b jets (bottom).



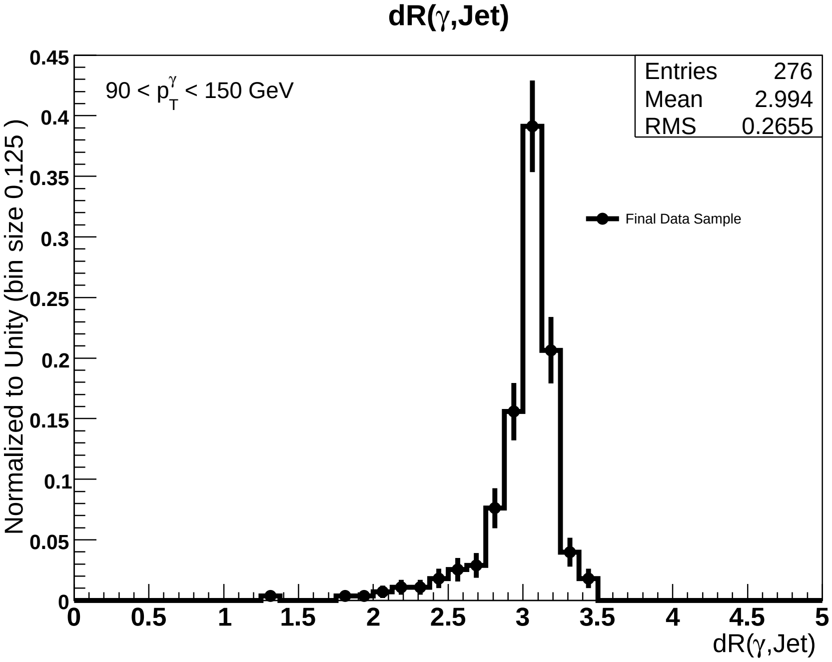


Figure 5.21: Normalized event distribution over distance d$\mathcal{R}(\gamma, jet)$ for two bins, $30 < p_T^\gamma < 40$ GeV (top) and $90 < p_T^\gamma < 150$ GeV (bottom).

effective, and it is discussed in further detail below. The third and final requirement for a jet to pass the b-tagging selection criteria is requiring that there is enough information such that the reduced Jet Lifetime Probability (from Section 4.2.6) can be calculated. Thus, the efficiency of the leading jet to be b-tagged in the data sample is a combination of these three efficiencies and can be expressed as:

$$\epsilon_{\text{b-tagged}} = \epsilon_{\text{Taggability}} \times \epsilon_{\text{bNN}} \times \epsilon_{\text{JLIP}}^{\text{Reduced}}. \tag{5.3}$$

The requirement that a jet be taggable imposes a set of geometric constraints to ensure that enough information is present in the jet to be able to properly identify it as a candidate heavy flavor jet. The most limiting of these conditions is that the jet's tracks contain at least some SMT information. The SMT barrel coverage extends to z positions of ± 37.8 cm, and for larger values of $|PV_z|$ combined with high particle rapidity, track reconstruction will be limited by the coverage of the SMT. The basic schematic that shows this effect is in Fig. 5.22. It makes sense that the efficiency of this requirement should have some dependence on the z position of the primary vertex. The taggability efficiency, therefore, has been separated into multiple regions of the primary vertex z position. Along with this treatment, it is advantageous to further divide these regions into a product of the jet cluster's pseudorapidity and the PV_z position, both positive ($\eta \times PV_z > 0$) and negative ($\eta \times PV_z < 0$). On average, the negative products have much more tracking information of the jet based on the geometry of their paths, as shown in the schematic of Fig. 5.22. Hence these jets also will have a higher taggability efficiency. For the region where $\eta \times PV_z > 0$, the primary vertex z position has been divided into three bins: $|PV_Z| < 20$ cm, $20 < |PV_Z| < 36$ cm, and $36 < |PV_Z| < 60$ cm. For the region where $\eta \times PV_z < 0$, there are only two bins: $|PV_Z| < 38$ cm, $38 < |PV_Z| < 60$ cm. In each of these bins, the taggability efficiency has been determined and parameterized as a function of the jet p_T and the jet rapidity. These given taggability efficiencies were studied, and it was found that for primary vertex positions $PV_z \gtrsim 36$ cm, these efficiencies became both very low and unstable. In order to maintain confidence in the efficiency of this requirement, the primary vertex z position was conservatively required to be within 35 cm of the center of the detector, and the resulting efficiency can seen in Fig. 5.23.

To both maximize the rejection of light jets and, at the same time, maintain high signal efficiencies for both c and b jets, the jet's b Jet Neural Network output (bNN) was required

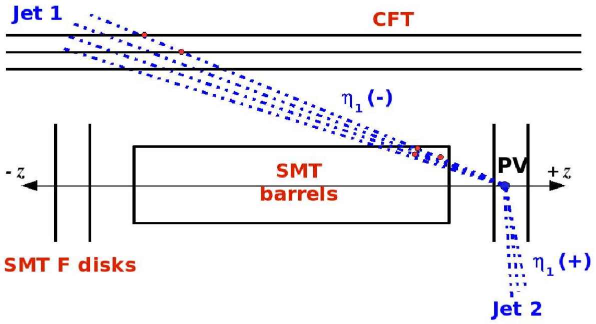


Figure 5.22: The schematic view of two jets with the same reconstructed primary vertex z positions, but with very different paths in the detector. In this diagram, jet 1 has a much better chance of satisfying the jet taggability requirements than jet 2.

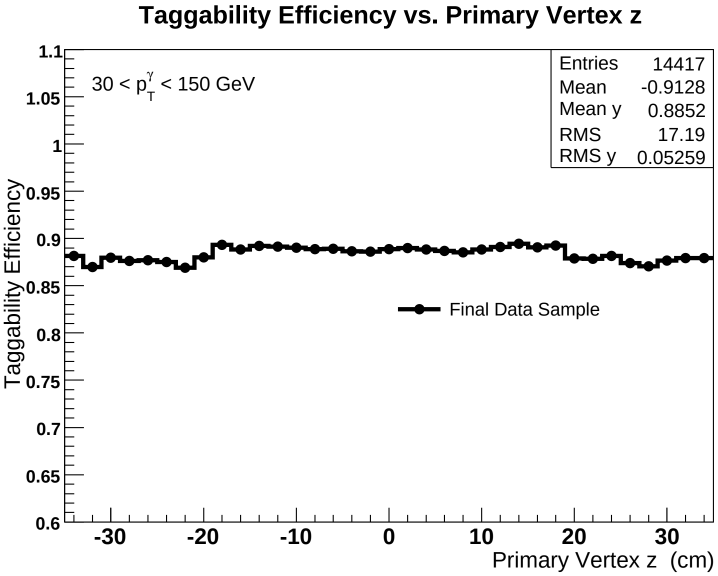


Figure 5.23: Taggability efficiency of the leading jet in the final data sample as a function of the primary vertex z position. The sharper edges of this efficiency at $PV_z = \pm 20$ cm are due to the binning used for the efficiency parameterization.

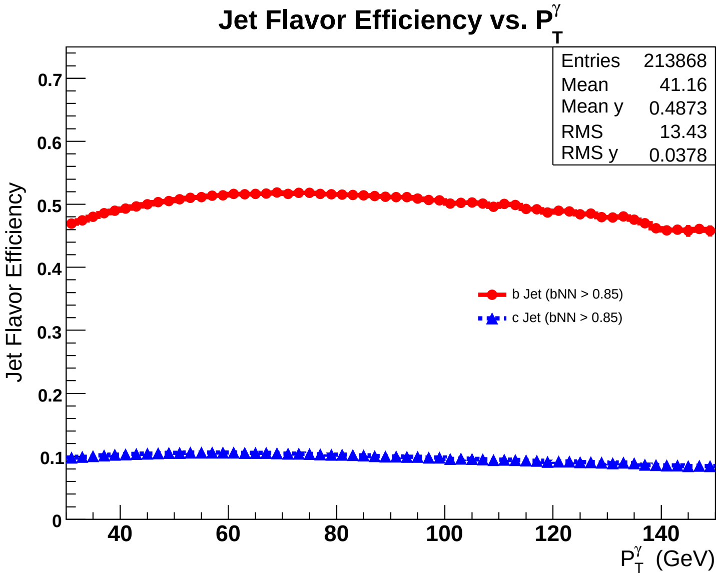


Figure 5.24: The efficiency for both c and b jets to the satisfy the bNN > 0.85 selection criterion as a function of p_T^γ. Neither c jets nor b jets show a strong dependence on p_T^γ.

to be greater than 0.85 (bNN > 0.85). The efficiency corresponding to this operating point was studied in-depth by a DØ group dedicated to identifying b jets. The efficiencies were determined separately for light, c, and b jets, and they are found to be weakly dependent on p_T as shown in Fig. 5.24. The efficiencies for each flavor are measured in simulation and then corrected by p_T dependent scale factors derived from data to simulation comparisons, which is documented in [45].

After both taggability and the bNN criteria have been imposed, we also require that the jet's reduced JLIP reduced value ([44]) is calculated. The efficiency for b-tagged b jets and c jets in simulation to meet this requirement is parameterized as a function of p_T^γ and is shown in Fig. 5.25. This efficiency in data is also parameterized in the same way, and its lower efficiency is due to the remaining light jet fraction of events in the data sample, and all these efficiencies are shown in Fig. 5.26. The difference in efficiency from the data sample to either the b, c, or light jet simulation sample is $\sim$3.8%. Because the data efficiency is a combination of light, c, and b jet efficiencies, the simulation efficiencies are weighted and summed to make a direct data to simulation efficiency comparison. There is at most a 1.5% difference between the two and this value is taken as the estimate of the systematic uncertainty.

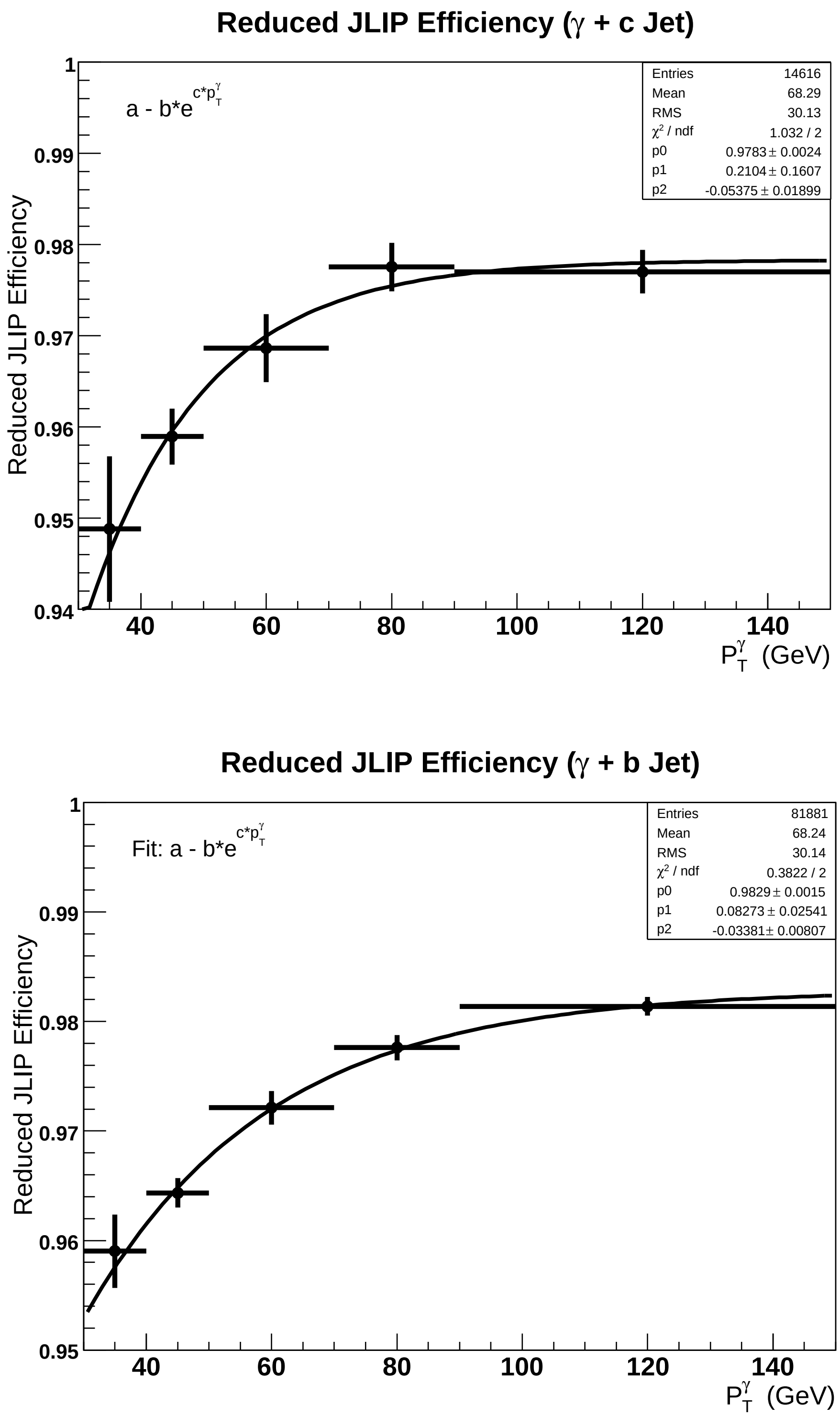


Figure 5.25: Efficiency for there to be enough information such that the reduced JLIP value can be calculated in the b-tagged $\gamma + c$ jet (top) and $\gamma + b$ jet (bottom) simulation samples. The parameterization of this efficiency is also shown.

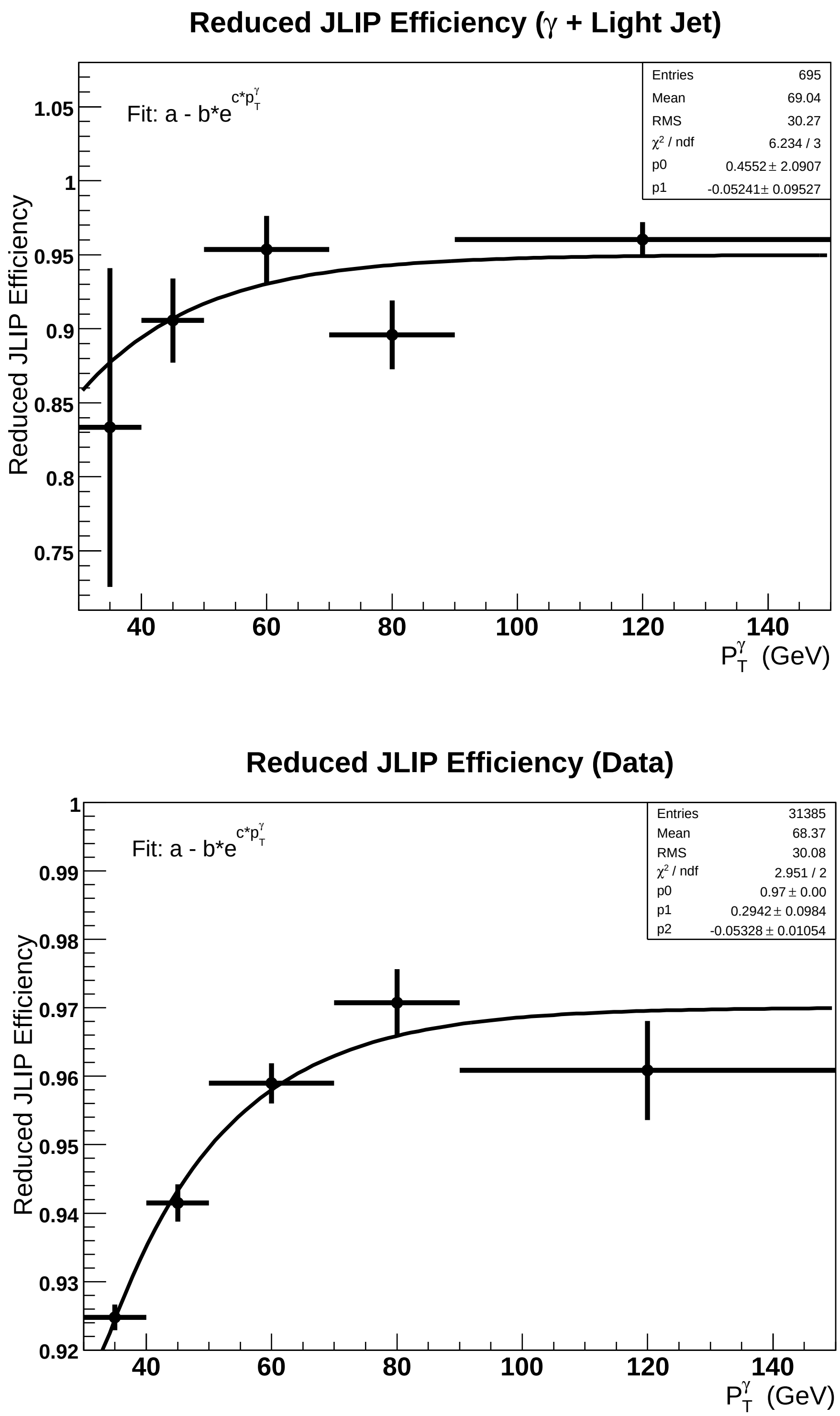


Figure 5.26: Efficiency for there to be enough information such that the reduced JLIP value can be calculated in the b-tagged light jet simulation (top) and in the data sample (bottom). The parameterization of this efficiency is also shown.

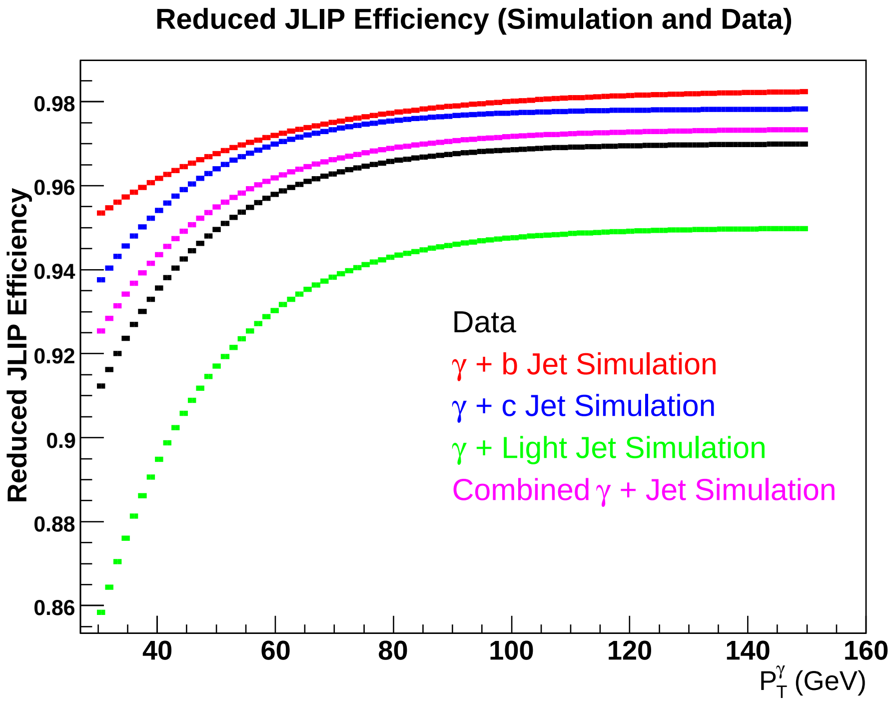


Figure 5.27: Parameterized efficiency for there to be enough information such that the reduced JLIP value can be calculated in each jet flavor simulation sample and in data. The green curve is the efficiency from simulated light jets and the magenta curve is the simulation efficiency weighted by the relative contribution from each of the three jet flavors. This weighted efficiency agrees with the found efficiency in data within 1.5% for all points in the curve.

5.4.8 Photon Purity Estimation

As mentioned before when selecting photons, the data contain a non-negligible fraction of background events. These contributions come most often from jets containing high momentum mesons, such as π^0, η, K^0_S, and ω mesons that mimic the signature of photons coming directly from the interaction. The first step to combating this background was detailed in Section 5.4.4, and involved the implementation of photon selection criteria, including the γ-ANN to retain signal and reject background events. The second step is to determine the photon purity in the final data sample, which is the subset of data satisfying all event selection criteria.

The γ-ANN was used as the discriminant to determine the fraction of signal in the final data sample. The data, as well as both signal and background simulation samples, were preselected using all of the selection criteria from the analysis. The photon purity for the data sample can be written as

$$\mathcal{P}_\gamma = \frac{N_\gamma}{N_\gamma + N_{\text{Jet}}}, \tag{5.4}$$

where N_γ is the number of photon events and N_{Jet} is the number of background EM-like jet events.

The shape of the γ-ANN in data can be used to calculate the photon purity because signal and background have different shapes in the region of $\gamma - \text{ANN} > 0.7$, and the shape in data is a linear combination of the two samples. To accurately determine the signal contributions to the data, a statistical/probabilistic method called TFractionFitter was used, based on the technique from HBOOK [56]. TFractionFitter incorporates the statistical uncertainty of the data and simulation histograms when fitting the simulation shapes to the data. The fit is simultaneously performed for both signal and background, such that their found fractions are not constrained to be within a certain range, nor are any relational assumptions made about the two. The photon purity fit is performed in each photon p_T range and separately in each photon-jet rapidity region. The results of this fitting procedure to the data can be seen in Fig. 5.28 for Region 1 and in Fig. 5.29 for Region 2. Additionally, the χ^2 fit per number of degrees of freedom for the normalized sum of simulation shapes to that of the data varies between 0.2–1.6 (depending on the p_T^γ bin), indicating good agreement between the two. The corresponding uncertainties from the fitting procedure are due to limited data statistics at high p_T^γ, and due to the low statistics of EM-like jets from simulation after all photon identification criteria have been applied.

The found photon purities are next fit to the following function:

$$\mathcal{P}_f^0 = 1 - \exp[-(a_1 + a_2 p_T^\gamma)], \tag{5.5}$$

where a_1 and a_2 are adjustable parameters. The resulting purities from the fitting technique with this parameterization are shown in Fig. 5.30.

The overall systematic uncertainty for the photon purity fitting technique can be broken down into three separate effects: the uncertainty from the fitting function used, the number of γ-ANN bins used in the fitting procedure, and the uncertainty coming from the fragmentation model used for the simulation samples.

To ensure against any possible bias from the specific form of the fitting function, it was

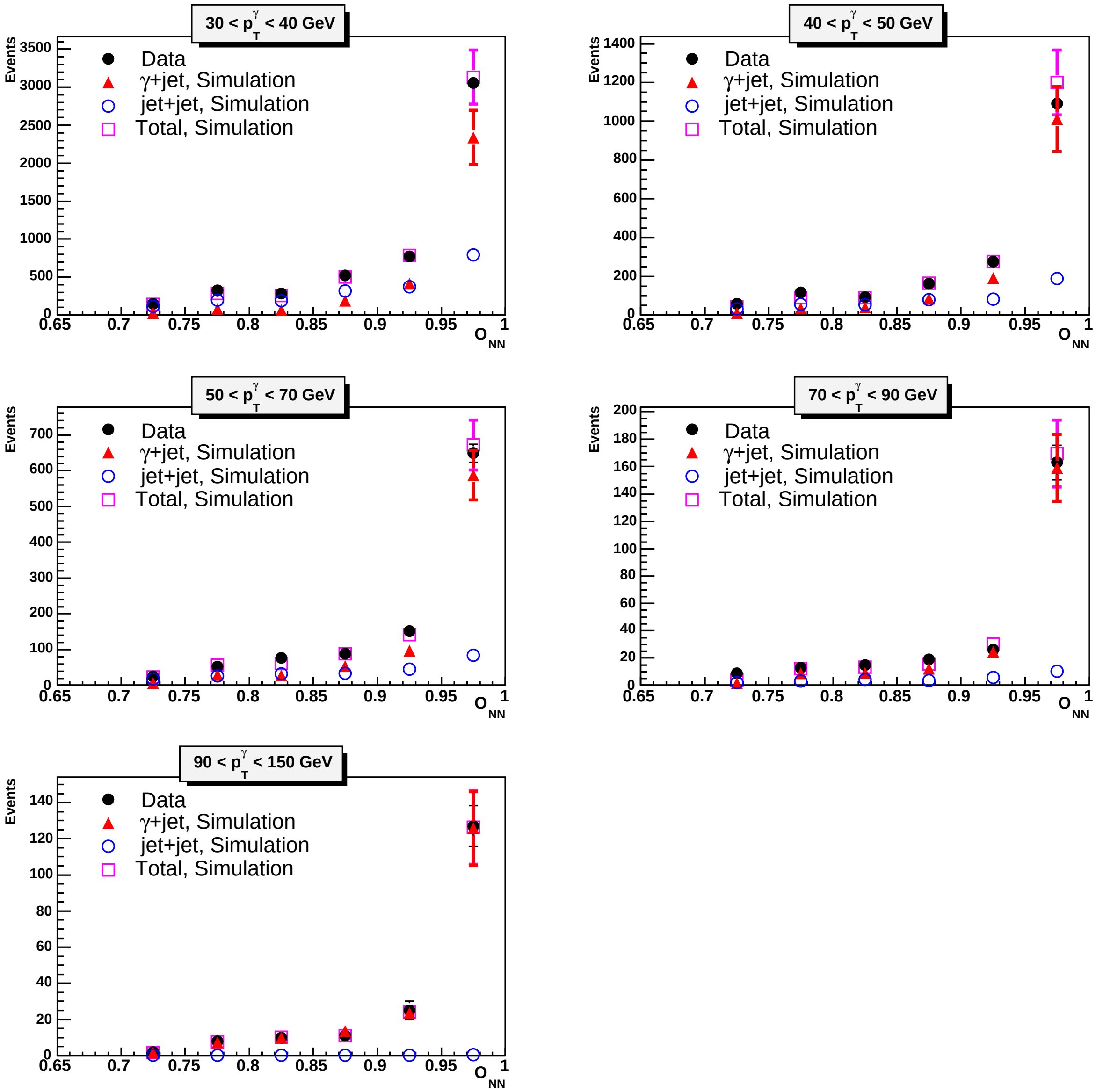


Figure 5.28: Distribution of the γ-ANN output for the five p_T^γ regions for data, simulation signal, and simulation background in Region 1. The simulation templates are weighted by their measured fractions, summed and then normalized to the number of events in data for a shape comparison.

Table 5.1: Results of photon purity fit using different fitting functions (Region 1).

Parameters	Fitting functions		
	$1/(1+a_1\ (p_T^\gamma)^{a_2})$	$1-\exp[-(a_1+a_2\ p_T^\gamma)]$	$a_1+a_2\ \log(p_T^\gamma)+a_3\ \log^2(p_T^\gamma)$
a_1	43.68±7.1	0.416±0.110	0.507±0.421
a_2	-1.19±0.04	0.016±0.023	-0.130±0.208
a_3	—	—	0.045±0.026
χ^2/ndf	0.48	0.42	0.37

compared with two other fitting function forms:

$$\mathcal{P}_f^1 = \frac{1}{1+a_1(p_T^\gamma)^{a_2}} \tag{5.6}$$

and

$$\mathcal{P}_f^2 = a_1 + a_2\log(p_T^\gamma) + a_3\log^2(p_T^\gamma). \tag{5.7}$$

The χ^2_{NDOF} results of all three fitting functions, as well as the parameter values used, are shown in Table 5.1. The systematic uncertainty derived from the differences in these photon purity fitting functions can be assessed by building a matrix as shown in Equation 5.8.

$$V'_{ij} = \frac{1}{N-1}\sum_{k=1}^{N}(x_k(i)-\mu(i))(x_k(j)-\mu(j)) \tag{5.8}$$

Here, i and j represent individual p_T^γ bins, μ is the value of the default fit, and N is the number of fitting functions used as cross-checks. The diagonal elements of the matrix give rise to the resulting systematic uncertainty, $(\delta\mathcal{P}_f^{\mathrm{Syst(1)}})^2 = V'_{ii}$.

The number of bins used in the template histograms of the γ-ANN output in the photon fitting procedure was varied between 6–14 (default of 10). Based on the stability of the results of the found photon purities, the size of this systematic was determined to be $\delta\mathcal{P}_f^{\mathrm{Syst(2)}} = 3.5\%$. The third source of uncertainty comes from the simulation modeling of the fragmentation model used in PYTHIA, which was studied in detail in Ref. [21]. This uncertainty is found to be $\delta f_{\mathrm{Frag}}^{\mathrm{Syst(3)}} = 0.20\cdot\exp[-0.0428\cdot p_T^\gamma]$. This corresponds to a $\sim 5\%$ uncertainty for $p_T^\gamma \approx 30$ GeV and $\sim 1\%$ uncertainty for $p_T^\gamma \approx 70$ GeV.

The total uncertainties from these sources are shown in the 'γ-Purity' column of Table 5.4 in each p_T^γ bin. These calculated uncertainties are also shown in the systematic band in Fig. 5.30.

Table 5.2: Estimate of W+1 and 2 jet background to be in the b-tagged data sample from simulation estimations. The ϵ is the efficiency for the sample to pass the entire event selection criteria. The cross sections (σ) are given from PYTHIA. The ratio in the final column gives the expected contribution of the W+jets sample to the b and c jet cross sections.

p_T^γ bin (GeV)	$(\epsilon \times \sigma)^{1jet}_{W+jet}$	$(\epsilon \times \sigma)^{2jet}_{W+jet}$	$\frac{(\epsilon\times\sigma)^{1+2jet}_{W+jet}}{(\epsilon\times\sigma)_{\gamma+\text{jets}}}$
30-40	1.34×10^{-7}	6.46×10^{-8}	1.18×10^{-3}
40-50	1.43×10^{-7}	7.30×10^{-8}	1.25×10^{-3}
50-70	2.41×10^{-7}	8.86×10^{-8}	1.91×10^{-3}
70-90	2.95×10^{-7}	2.11×10^{-7}	2.68×10^{-3}
90-150	2.73×10^{-7}	1.62×10^{-7}	2.86×10^{-3}

An additional test of the photon purity fitting technique was performed using a method of solving for the purity from each sample's efficiency to pass a certain cut. The photon purity $\mathcal{P}$ is calculated by combining the efficiency of data ε_d, signal ε_s, and background ε_b (determined with respect to the cut $\gamma - \text{ANN} > 0.9$). The derived photon purity is shown in Equation 5.9, and a more detailed explanation of this technique can be found in Ref. [49].

$$\mathcal{P} = \frac{\varepsilon_d - \varepsilon_b}{\varepsilon_s - \varepsilon_b}. \tag{5.9}$$

The purities derived by this method are found to be in good agreement with those from the template fitting technique.

The background contribution caused by W + jet events to the measurement of inclusive "γ+jet" cross section has previously been found to be $\leq 0.5\%$[48]. For the b jet case, we study the expected contribution from W+jet events that pass the additional b tagging requirements. First, the efficiency for the W+1 jet exclusive simulation sample to pass the entire selection criteria was determined and it was then multiplied by the cross section given in PYTHIA. This is repeated for the W+2 jet exclusive simulation sample as well as for the inclusive "γ+jet" simulation sample. Finally the ratio of the two combined W+jet samples' efficiency × cross section values are taken with respect to the "γ+jet" efficiency × cross section as shown in Table 5.2. The maximum contributions from this background to the final sample of events is estimated to be less than 0.3%.

5.4.9 Jet Flavor Fractions

After requiring jets to pass the b-tagging selection criteria outlined in Section 5.4.7, the fractional contributions of light, c, and b jets to the final data sample must be determined. Unfortunately, after the initial selection of heavy flavor jets using the criterion of bNN > 0.85, the shape of the bNN no longer contains enough discrimination to adequately distinguish individual jet flavors. This means that using a template fitting technique with the bNN output shapes, as was performed for the photon purity determination in Section 5.4.8, will result in large uncertainties for the found flavor fractions and will not produce robust results.

However reduced JLIP, discussed in Section 4.2.6, still shows significant discrimination among light, c, and b jets. Even more, by making a transformation of this variable as rJLIP $= -\ln(P_{\text{Jet}}^{\text{Reduced}})$, its shape can be used to distinguish individual jet flavors, as can be seen in Fig. 5.31. The fitting procedure used to estimate the fractions of jet flavors that exist in the data sample is the same as in the photon purity technique, with the exception that now a three-template fit is performed using the rJLIP variable. The fit is simultaneous for all three fractions, with the only limitations being that each fraction was constrained to be between 0 and 1, and that the sum of all three fractions must equal one. To reduce the dependence of simulation to data disagreements, as well as to increase statistics for light jets, an enhanced light jet data sample of “negative tagged” jets was used to replace the simulated light jets.

The fitting procedure was performed in each region of p_T^{γ}, and the result for b jets is shown for Region 1 in Fig. 5.32. Because the measured flavor fractions for both c and b jets were in agreement within uncertainties between regions in each p_T^{γ} bin and no systematic fluctuations were seen, the regions were combined and the flavor fractions were determined for the two rapidity regions together. The results are shown for all five bins of p_T^{γ} in Fig. 5.33 for both c and b jets. To test the ability of these combined flavor fractions to model the data, the jet flavor templates were weighted by their measured fractions, summed, and then normalized to the data to compare the respective rJLIP distributions. The agreement between the two shapes is found to be quite good and can be seen in Fig. 5.34 for all bins of p_T^{γ} , and again in Fig. 5.35 on a log scale for the y-axis. The agreement of the shape from simulated jets with that of the “negative tagged” data sample is confirmed by using the simulation shape in the overall shape comparison in Fig. 5.36.

Parallels can again be made with the techniques to determine photon purities by using the efficiency method, as shown in Equation 5.9. In this case, there are three distinct jet types that contribute to the data and so the equation becomes more complicated.

$$\begin{aligned} N_d &= N_b + N_c + N_l \\ \text{After } 1^{st} \text{ Cut}: \quad \varepsilon_d N_d &= \varepsilon_b N_b + \varepsilon_c N_c + \varepsilon_l N_l \\ \text{After } 2^{nd} \text{ Cut}: \quad \varepsilon'_d N_d &= \varepsilon'_b N_b + \varepsilon'_c N_c + \varepsilon'_l N_l \end{aligned} \tag{5.10}$$

In these equations, ε_d (ε'_d), ε_b (ε'_b), ε_c (ε'_c) and ε_l (ε'_l) are the efficiencies to pass the first (second) cut in data and in simulated b, c and light jet samples, respectively. Because we have three unknowns (the fractions of light, c, and b jets), we require all samples to pass two selection requirements, each using a separate operating point, to obtain three equations to solve for them. Then the b jet flavor fraction (f_b) can be expressed as

$$f_b = \frac{(\varepsilon_d - \varepsilon_l) - (\varepsilon'_d - \varepsilon'_l) \cdot k}{(\varepsilon_b - \varepsilon_l) - (\varepsilon'_b - \varepsilon'_l) \cdot k}, \tag{5.11}$$

where

$$k = (\varepsilon_c - \varepsilon_l)/(\varepsilon'_c - \varepsilon'_l). \tag{5.12}$$

The flavor fractions for both c and b jets are found using three separate pairs of operating points for rJLIP such that rJLIP $>$ $Cut\ 1$ ($Cut\ 2$), corresponding to $(Cut\ 1, Cut\ 2)$ =[(3,7),(2.5,6),(3,8)]. The results for all of these cut combinations are compared to the flavor fractions found from the template fitting technique and is shown in Fig. 5.37. The resulting fractions are found to be both stable with choice of operating point and flavor fraction determination technique.

The measured flavor fractions for both c and b jets seem to have a distinct and more pronounced shape in the first two regions of p_T^γ. This behavior was verified by splitting the p_T^γ 30–40 GeV region into two separate regions: 30–34 GeV and 34–40 GeV. The flavor fractions were rederived and confirm that the structure of this shape is a real effect in the heavy flavor jets' found fractions, as can be seen in Fig. 5.38. These p_T^γ regions were recombined as the uncertainties were smaller in this case.

5.4.10 Unsmearing corrections

In order to correct for the finite resolution of the calorimeter, an "unsmearing correction" is applied. For the DØ calorimeter, the parameters A, B, and C in Equation 4.5 have been

Table 5.3: The measured fitting parameters from Equation 5.14 for both the "$\gamma + b$ jet" and "$\gamma + c$ jet" cases.

Parameter	"$\gamma + b$ jet"	"$\gamma + c$ jet"
a_0	$(6.23 \pm 0.87) \times 10^5$	$(3.83 \pm 0.62) \times 10^7$
a_1	4.12 ± 0.43	4.78 ± 0.42
a_2	6.31 ± 2.09	6.64 ± 4.39

determined to be:

$$A = 0.29 \text{ GeV}, B = 0.224 \text{ GeV}^{1/2}, \text{ and } C = 0.0439. \tag{5.13}$$

The Run I calorimeter resolution was such that these effects could be ignored for EM objects [58], but it has degraded with the additional material added for the Run II DØ upgrade. These effects now must be estimated given that this cross section is steeply falling with respect to p_T^γ.

The unsmearing procedure was performed in the following way:

- An ansatz is made to parameterize the initial p_T^γ spectrum, and for this case a three parameter function was used:

$$\sigma(p_T) = a_0 \cdot p_T^{-a_1} \cdot (1 - 2p_T/\sqrt{s})^{a_2}. \tag{5.14}$$

- The ansatz is smeared using the calorimeter energy resolution for EM objects given in Equation 4.5 with the coefficients listed above in Equation 5.13.
- The smearing correction is taken as the ratio of the unsmeared ansatz over the smeared ansatz for the given p_T^γ intervals of the cross section measurement.

The resulting parameters determined after fitting the smeared ansatz to the data can be found in Table 5.3 for both the "$\gamma + b$ jet" and "$\gamma + c$ jet" cross sections.

The ratio of the unsmeared to the smeared "$\gamma + b$ jet" and "$\gamma + c$ jet" cross sections are shown in Fig. 5.39. As can be seen in these plots, the unsmeared cross section is generally 1–2% smaller than the smeared one. The parameterization of this correction as a function of p_T^γ can be seen in these plots. The corrections for the "$\gamma + b$ jet" and "$\gamma + c$ jet" cross sections are within 1% for all p_T^γ intervals.

The relative uncertainty from the ansatz fitting is less than 0.5% in all p_T^γ intervals, but to be conservative, a 1% systematic uncertainty is taken to account for potential differences between the true and the fitted EM energy resolutions. As a cross check, the ratio of the smeared ansatz to the data was made and is found to be within uncertainties (the full experimental uncertainties are used for the data), shown in Fig. 5.40. The unsmearing corrections are applied to the cross sections to arrive at their final values.

5.5 Systematic Uncertainties

The systematic uncertainties are shown in Table 5.4 for the "$\gamma + b$ jet" cross section and in Table 5.5 for the "$\gamma + c$ jet" cross section, both in Region 1. The uncertainties for the "$\gamma + c$ jet" are very similar, differing only with respect to the uncertainties on the flavor fractions and to the c jet selection efficiency. The uncertainties corresponding to Region 2 only differ from Region 1 due to the uncertainty of the photon purity fitting, and the total change in error is less than 2%.

The largest uncertainties are from the jet flavor fractions (at high p_T^γ), the photon purity fitting (at low p_T^γ), the photon and jet selection efficiencies and the luminosity. Also, uncertainties from the EM energy scale and the photon energy scale correction ($C_{p_T^\gamma}$) are included. All of these contributions to the overall uncertainties can be seen for Region 1 of the "$\gamma + b$ jet" cross section in Fig. 5.41. The uncertainties at low p_T^γ are correlated due to the parameterization of the photon purity; however, the bin-to-bin correlations are small for large p_T^γ due to the independent determination of flavor fractions in every photon p_T interval.

5.6 Results and Theoretical Comparisons

The total number of $\gamma +$ heavy flavor jet candidate events remaining in Regions 1 and 2 after application of all the selection criteria is 14,417 ($\sim$51.9% in Region 1 and $\sim$48.1% in Region 2). These events are used to calculate the cross sections in five p_T^γ intervals (30–150 GeV). The leading jet and photon p_T spectra can be seen in Fig. 5.42 and show good agreement when compared to simulation as seen in Fig. 5.43.

The final "$\gamma + b$ jet" cross sections are shown in Table 5.6 for Region 1 and Table 5.7 for Region 2. Similarly, the cross section results for Region 1 are found in Table 5.8 and in Table 5.9. These results have been plotted in Figs. 5.44 and 5.45 and include next-to-

Table 5.4: Systematic uncertainties (in %) for b jets in Region 1 due to photon purity ($\mathcal{P}_\gamma$), jet flavor fractions ($\mathcal{P}_b$), trigger efficiencies (ϵ_t), photon (ϵ_s^γ) and jet (ϵ_s^{jet}) selection efficiencies, the EM energy scale (ES), p_T^γ correction ($C_{p_T^\gamma}$), PV_z selection efficiency (PV_z), photon acceptance ($\mathcal{A}_\gamma$), luminosity ($\mathcal{L}$), unsmearing ($\mathcal{U}$), E_T^{miss} cut (E_T^{miss}) and b-tagging efficiency ($\epsilon_{\text{b-tag}}$). The uncertainties shown for Region 1 are within 2% for Region 2.

p_T^γ (GeV)	$\mathcal{P}_\gamma$	$\mathcal{P}_b$	ϵ_t	ϵ_s^γ	ϵ_s^{jet}	ES$_\gamma$	$C_{p_T^\gamma}$	PV_z	$\mathcal{A}_\gamma$	$\mathcal{L}$	$\mathcal{U}$	E_T^{miss}	$\epsilon_{\text{b-tag}}$	$\mathbf{E_{Syst}}$
30 – 40	10.5	8.6	1.8	5.2	7.8	3.3	1.9	1.4	1.4	6.0	1.0	1.5	3.9	18.5
40 – 50	8.6	6.6	1.2	5.1	4.4	3.1	2.3	1.4	1.4	6.0	1.0	1.5	4.0	15.5
50 – 70	7.1	8.3	1.0	5.0	3.2	2.8	2.4	1.4	1.4	6.0	1.0	1.5	4.3	15.2
70 – 90	5.8	17.0	1.0	4.9	2.3	2.7	2.8	1.5	1.4	6.0	1.0	1.5	4.7	20.8
90 – 150	7.1	22.7	1.0	4.9	1.9	2.5	4.0	1.6	1.4	6.0	1.0	1.5	5.5	26.2

Table 5.5: Systematic uncertainties as shown in Table 5.4, but now for c jets in Region 1.

p_T^γ (GeV)	$\mathcal{P}_\gamma$	$\mathcal{P}_c$	ϵ_t	ϵ_s^γ	ϵ_s^{jet}	ES$_\gamma$	$C_{p_T^\gamma}$	PV_z	$\mathcal{A}_\gamma$	$\mathcal{L}$	$\mathcal{U}$	E_T^{miss}	$\epsilon_{\text{b-tag}}$	$\mathbf{E_{Syst}}$
30 – 40	10.5	8.3	1.8	5.2	5.8	3.3	1.9	1.4	1.4	6.0	1.0	1.0	5.2	18.1
40 – 50	8.6	10.6	1.2	5.1	3.7	3.1	2.3	1.4	1.4	6.0	1.0	1.0	5.2	17.7
50 – 70	7.1	12.3	1.0	5.0	3.0	2.8	2.4	1.4	1.4	6.0	1.0	1.0	5.2	17.9
70 – 90	5.8	17.5	1.0	4.9	2.3	2.7	2.8	1.5	1.4	6.0	1.0	1.0	5.2	21.3
90 – 150	7.1	24.2	1.0	4.9	1.9	2.5	4.0	1.6	1.4	6.0	1.0	1.0	5.2	27.5

leading order theoretical predictions from Ref. [24]. For these predictions, the choice of renormalization, factorization and fragmentation scales were chosen as $\mu_R = \mu_F = \mu_f = p_T^\gamma$. One can see that in the range of $30 < p_T^\gamma < 150$ GeV for Regions 1 and 2, for both b and c jets, the cross sections fall more than two orders of magnitude. Statistical errors vary from 0.2% in the first p_T^γ bin to 8–9% in the last bin while systematic errors vary between 17–30% (depending on the p_T^γ region).

To make a more direct comparison between the results of the measurements and theoretical predictions, the data to theory ratio is taken.

In these ratios, a result of one represents perfect agreement between the measured and calculated cross sections. Deviation from one in these ratios indicate the size of discrepancies between the two. Also included in these plots are the uncertainties from the PDF set [26] used in the theoretical calculations, as well as the theoretical scale uncertainties. The scale uncertainties are obtained from a simultaneous variation of all the three scales by a factor of two up and down, $\mu_{R,F,f} = 0.5 \cdot p_T^\gamma$, and $\mu_{R,F,f} = 2 \cdot p_T^\gamma$. The "$\gamma + b$ jet" cross section ratios are presented in Fig. 5.46 for Region 1 and Fig. 5.47 for Region 2, and agreement is seen between data and theory within uncertainties for both Region 1 and 2 and in the entire p_T^γ range. Two possible models of intrinsic charm (IC) [27] are also present in these ratios as a ratio of the theoretical calculations with enhanced charm to the standard next-to-leading order results. For the "$\gamma + b$ jet" case, no enhancement of the cross section is expected, and none is seen.

The "$\gamma + c$ jet" cross section ratios are presented in Fig. 5.48 for Region 1 and in Fig. 5.49 for Region 2. Agreement is seen in both Regions for $p_T^\gamma < 50$ GeV in both rapidity regions; however, the more striking effect is the rising disagreement between data and theory as p_T^γ increases. This effect is seen in Region 1 and 2 for $p_T^\gamma > 70$ GeV, and represents a larger measured cross section than was predicted from theoretical calculations. The two IC models both show enhanced cross section results in these plots, and although at least one IC model (BHPS) describes the data better than the standard theoretical predictions, there is no model that describes the data over the entire p_T^γ range in both rapidity regions. There are other effects that could play a part in this discrepancy. One potentially major contribution could be the underestimation of gluon to charm anticharm pair fraction, as measured from LEP [19]. The contribution of the quark-antiquark annihilation subprocess rises as a function of p_T^γ, and this results in an increased sensitivity to this splitting fraction.

Table 5.6: "$\gamma + b$ jet" cross section for *Region 1* for each p_T^γ bin, along with its event-weighted p_T^γ value ($< p_T^\gamma >$). The total uncertainties ($\mathrm{E_{Tot}}$) are composed of statistical ($\mathrm{E_{Stat}}$) and systematic ($\mathrm{E_{Syst}}$) parts.

p_T^γ bin (GeV)	$< p_T^\gamma >$ (GeV)	Cross section (pb/GeV)	$\mathrm{E_{Stat}}$(%)	$\mathrm{E_{Syst}}$(%)	$\mathrm{E_{Tot}}$(%)
30 – 40	34.1	2.73×10^{-1}	1.5	18.5	18.6
40 – 50	44.3	1.09×10^{-1}	2.5	15.5	15.7
50 – 70	57.6	2.72×10^{-2}	3.3	15.2	15.6
70 – 90	78.7	6.21×10^{-3}	6.6	20.8	21.8
90 – 150	108.3	1.23×10^{-3}	8.2	26.2	27.5

Table 5.7: "$\gamma + b$ jet" cross section for *Region 2*.

p_T^γ bin (GeV)	$< p_T^\gamma >$ (GeV)	Cross section (pb/GeV)	$\mathrm{E_{Stat}}$(%)	$\mathrm{E_{Syst}}$(%)	$\mathrm{E_{Tot}}$(%)
30 – 40	34.1	2.23×10^{-1}	1.6	19.1	19.2
40 – 50	44.2	9.53×10^{-2}	2.6	16.0	16.2
50 – 70	57.4	2.67×10^{-2}	3.3	15.3	15.7
70 – 90	78.3	6.10×10^{-3}	6.7	20.8	21.9
90 – 150	110.0	1.09×10^{-3}	8.9	25.7	27.2

Table 5.8: "$\gamma + c$ jet" cross section for *Region 1*.

p_T^γ bin (GeV)	$< p_T^\gamma >$ (GeV)	Cross section (pb/GeV)	$\mathrm{E_{Stat}}$(%)	$\mathrm{E_{Syst}}$(%)	$\mathrm{E_{Tot}}$(%)
30 – 40	34.1	$1.90\times10^{+0}$	1.5	18.1	18.2
40 – 50	44.3	5.14×10^{-1}	2.5	17.7	17.9
50 – 70	57.6	1.53×10^{-1}	3.3	17.9	18.2
70 – 90	78.7	4.45×10^{-2}	6.6	21.3	22.3
90 – 150	108.3	9.63×10^{-3}	8.2	27.5	28.7

Table 5.9: "$\gamma + c$ jet" cross section for *Region 2*.

p_T^γ bin (GeV)	$< p_T^\gamma >$ (GeV)	Cross section (pb/GeV)	$\mathrm{E_{Stat}}$(%)	$\mathrm{E_{Syst}}$(%)	$\mathrm{E_{Tot}}$(%)
30 – 40	34.1	$1.56\times10^{+0}$	1.6	18.7	18.8
40 – 50	44.2	4.51×10^{-1}	2.6	18.1	18.3
50 – 70	57.4	1.50×10^{-1}	3.3	18.0	18.3
70 – 90	78.3	4.39×10^{-2}	6.7	21.3	22.3
90 – 150	110.0	8.57×10^{-3}	8.9	27.0	28.4

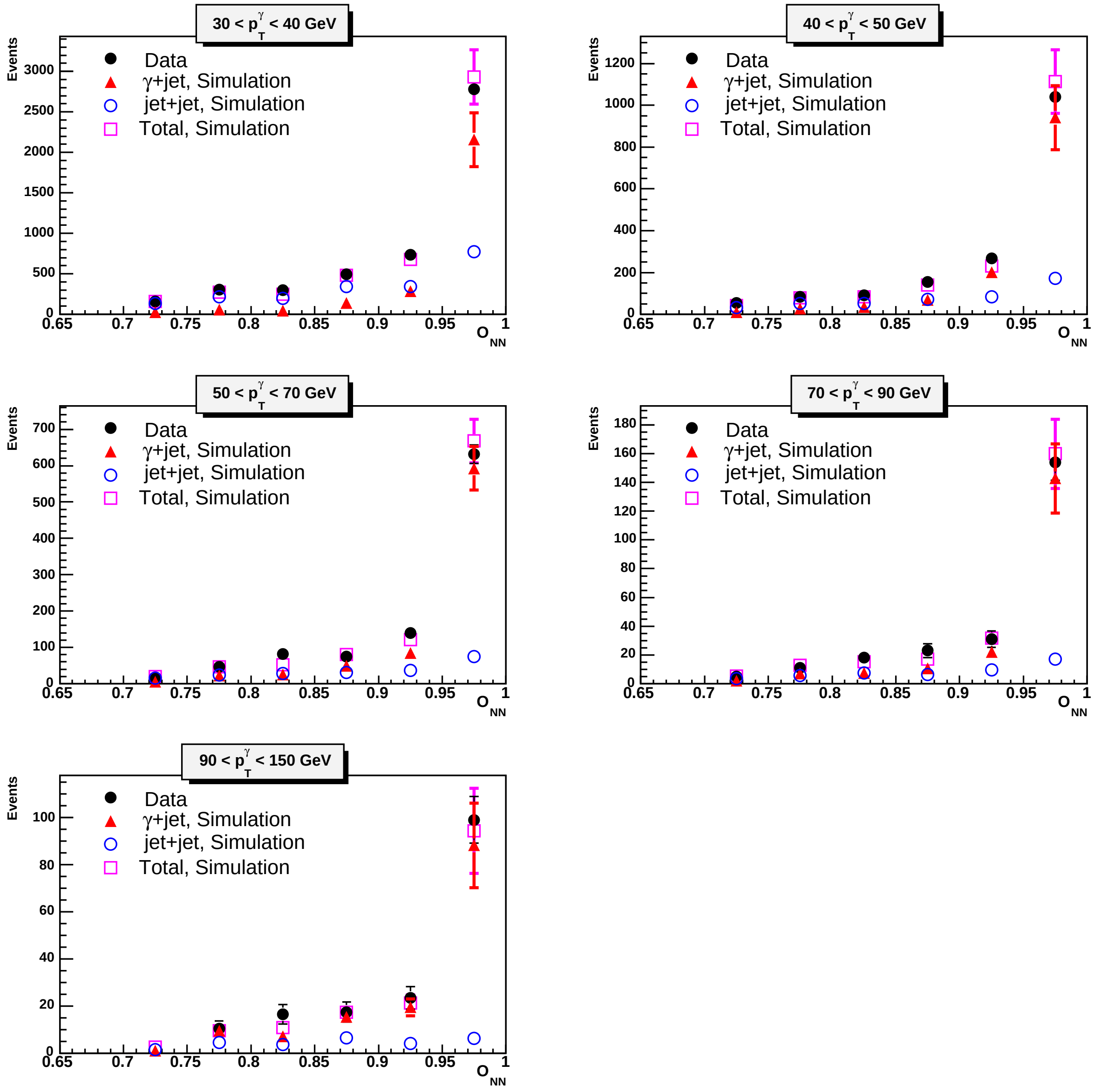


Figure 5.29: Same as in Fig. 5.28, except now Region 2 is shown.

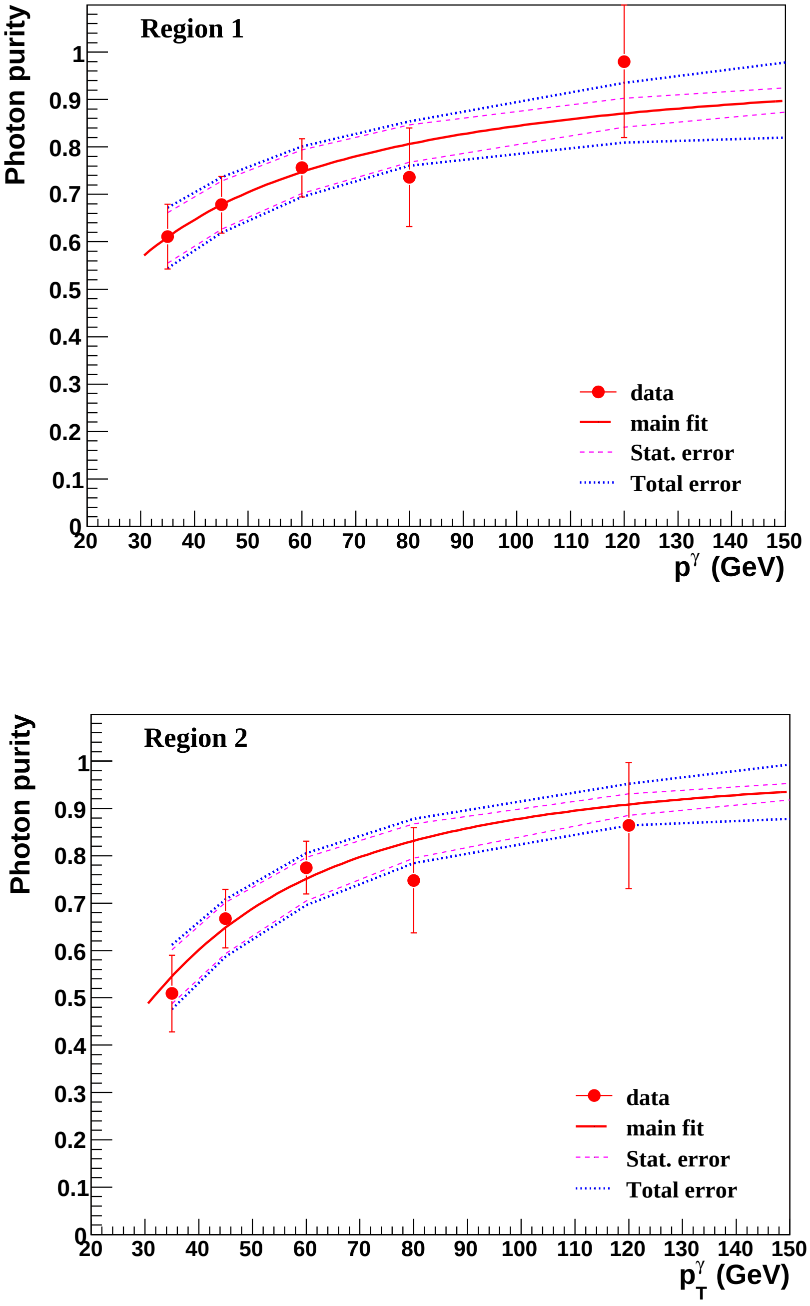


Figure 5.30: The photon purity in the selected "$\gamma^{dir} + b$ jet" events as a function of p_T^{γ} for the two measured configuration of photon and jet rapidities: the top plot is for Region 1 and the bottom plot for Region 2. The result of the functional fit $(1 - \exp(a_1 + a_2 \cdot p_T^{\gamma}))$ is shown by the full lines, together with the statistical uncertainty in the default fit (dashed lines), the systematic uncertainty (dotted lines) and the total uncertainty (dash-dotted lines).

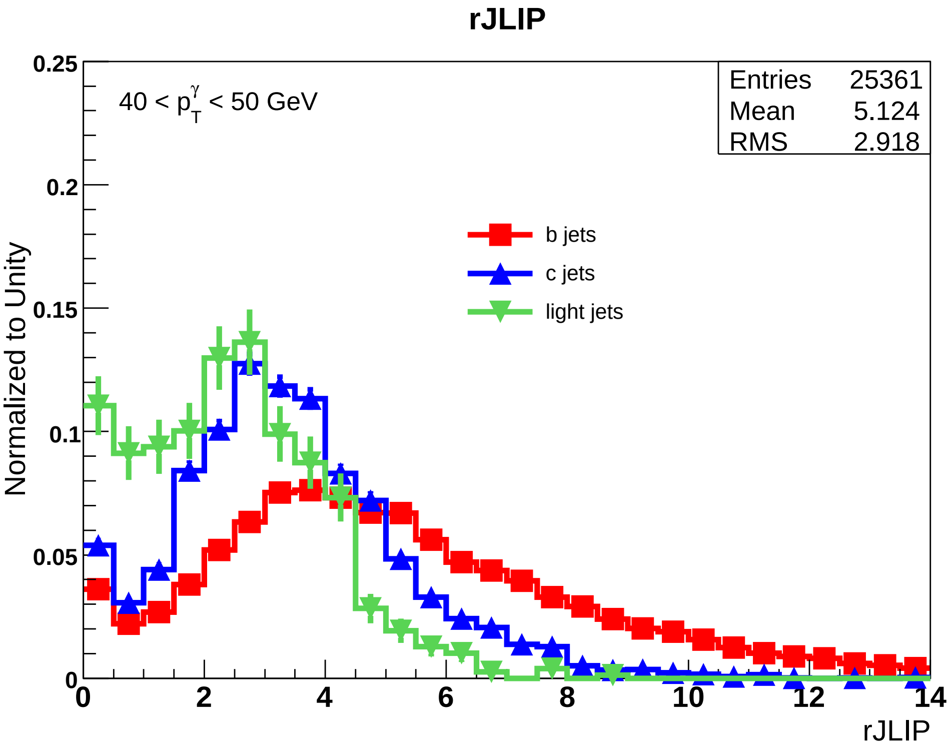


Figure 5.31: Distribution of the rJLIP variable for simulated light, c, and b flavored jets.

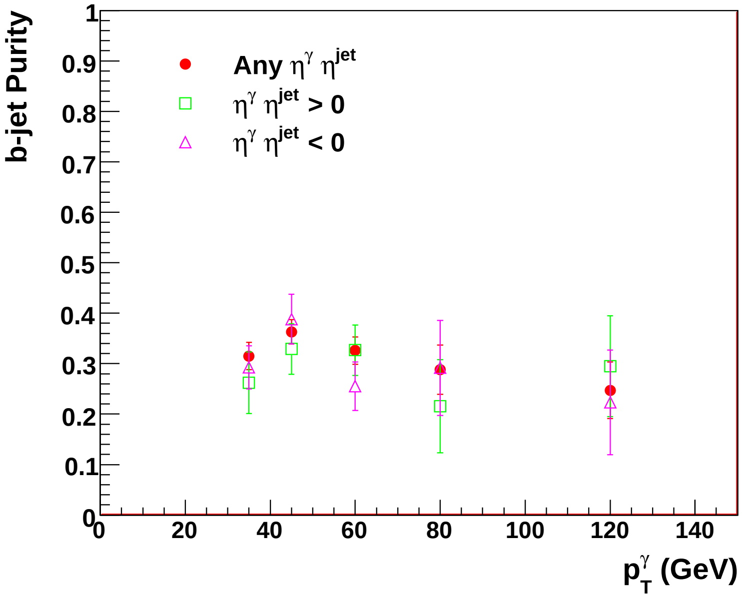


Figure 5.32: Fractions of b jets in the final data sample, shown for separately for Region 1 and Region 2, and for the single combined region. The shown errors are statistical, and the results from the two regions agree within uncertainties for the entire p_T^γ range.

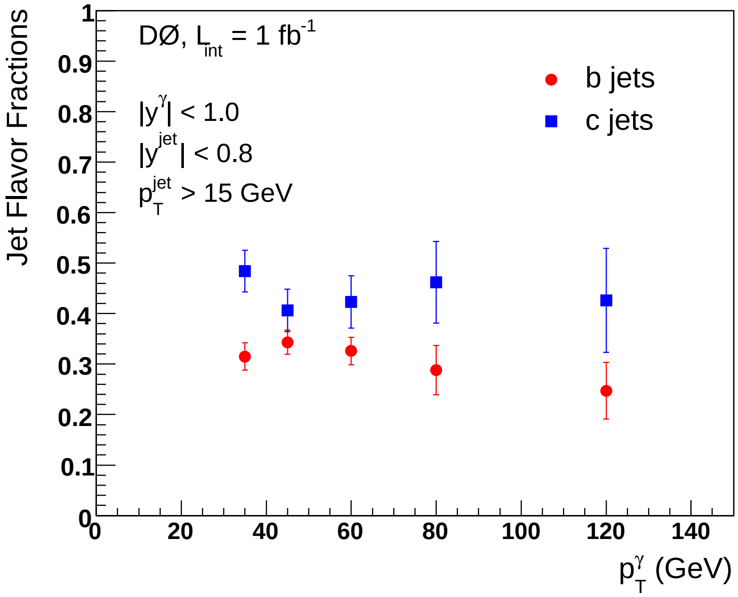


Figure 5.33: Measured fractions of c and b jets vs. p_T^γ in the final data sample.

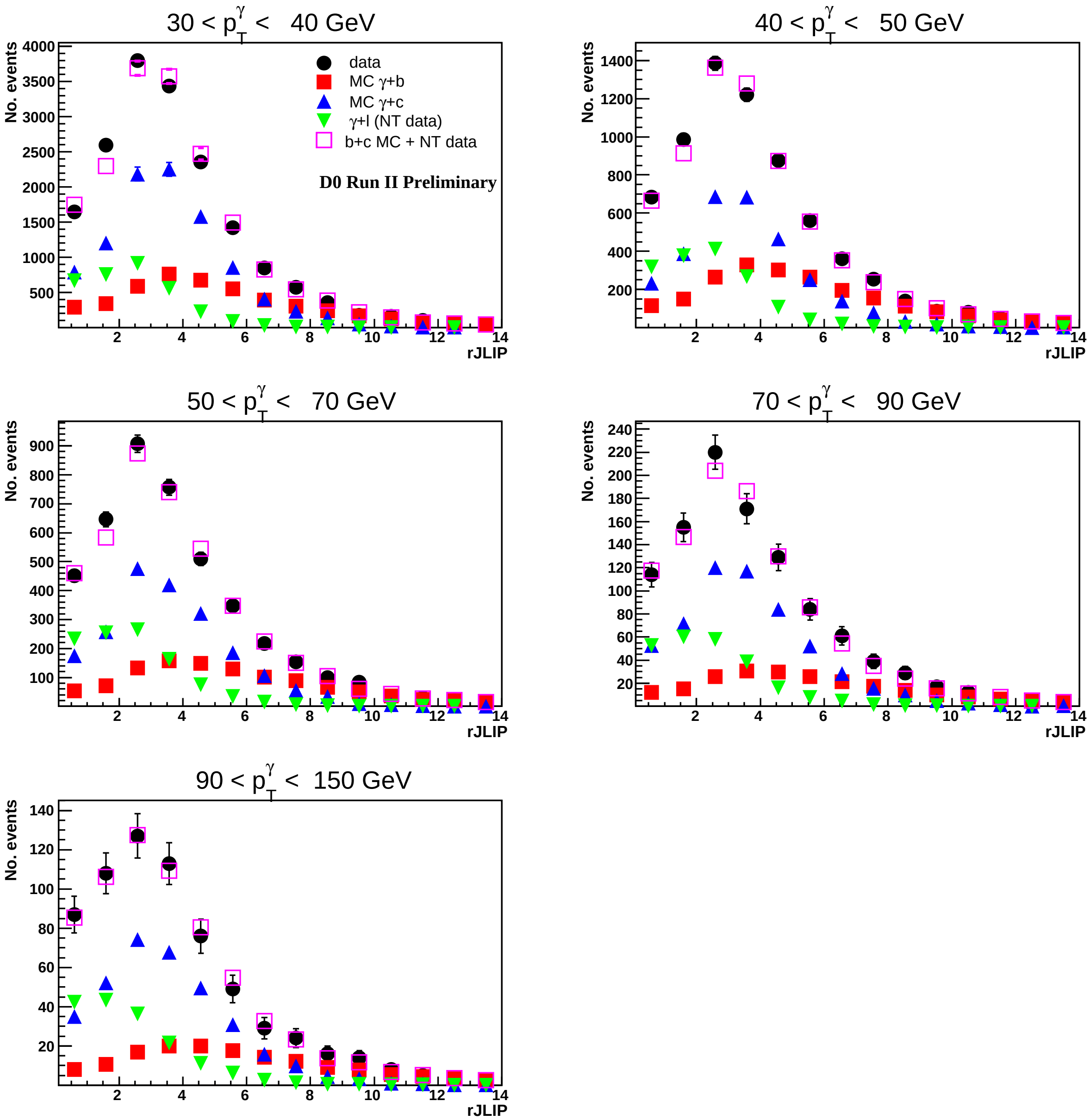


Figure 5.34: Distribution of the number of events in the final data sample for the rJLIP$\equiv$ $-\ln$(reduced JLIP) for five p_T^γ intervals from 30 to 150 GeV. The jet flavor templates for light, c, and b jets have been normalized to the data and weighted by their measured flavor fractions. The sum of these templates directly compares the shape derived from the fitting technique to that of the data, and good agreements is seen between the two.

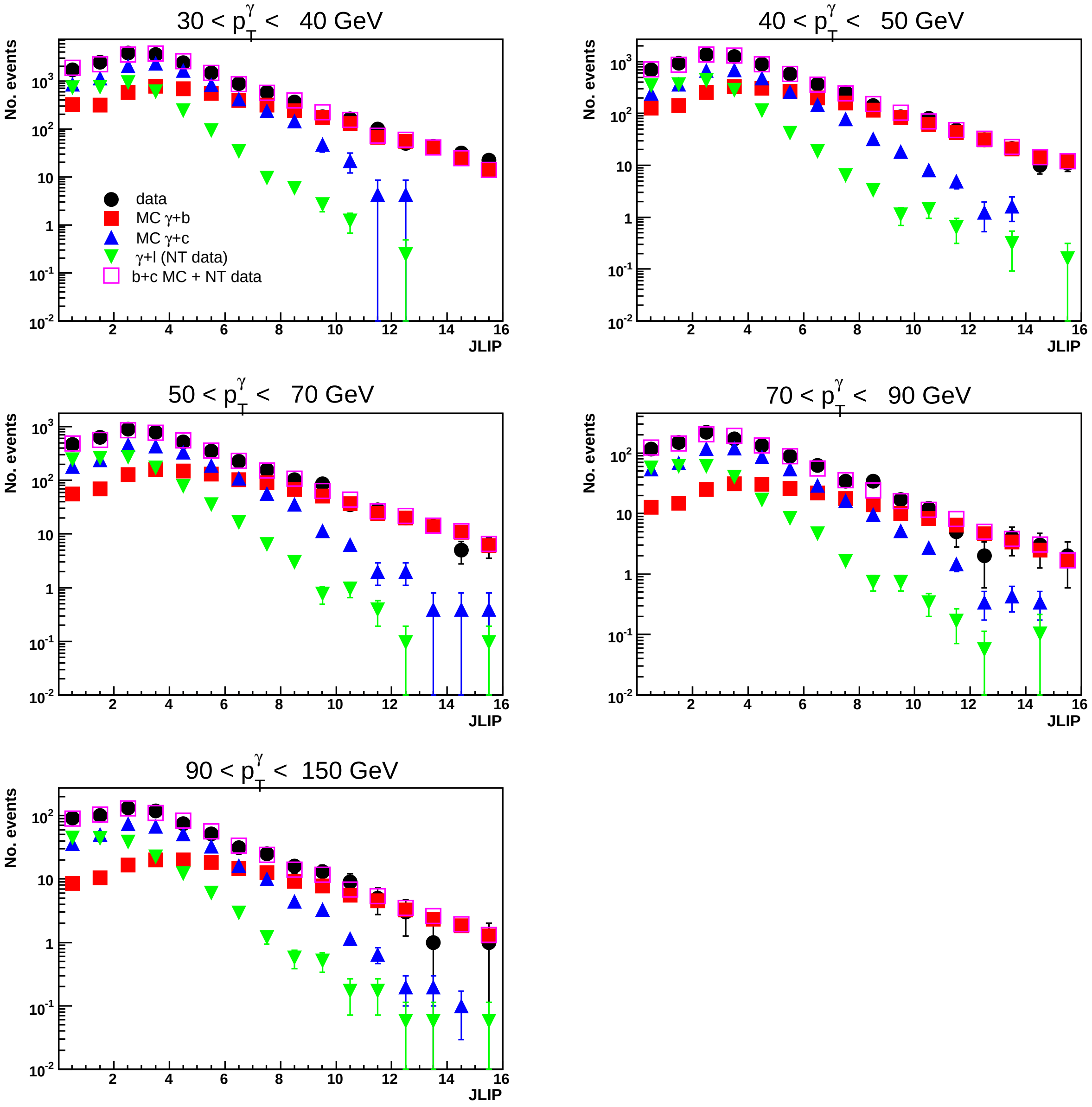


Figure 5.35: Identical to Fig. 5.34 but with a log scale for the event numbers (y-axis).

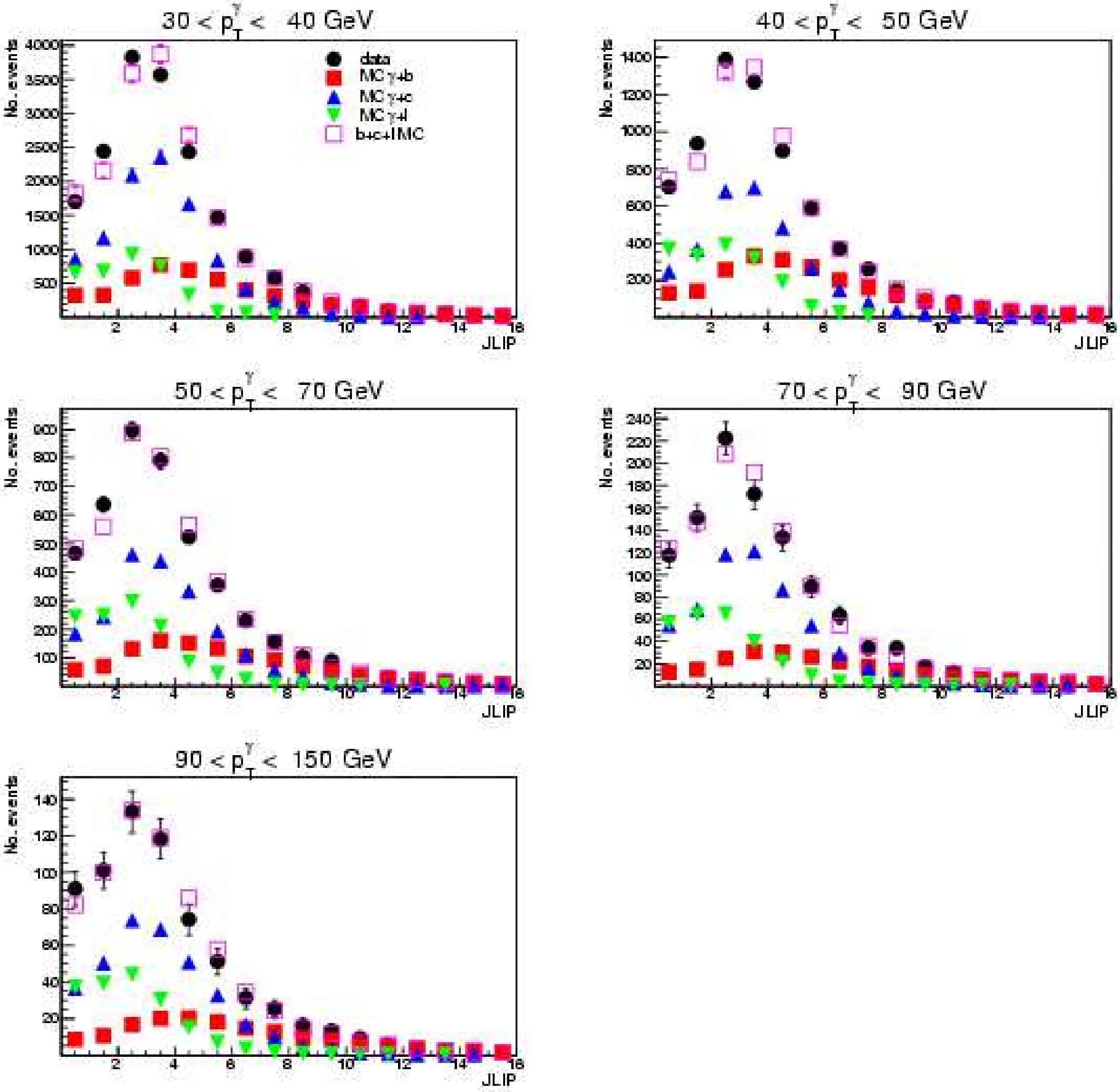


Figure 5.36: Identical to Fig. 5.34, except that the light jet template is taken from simulation. The flavor fractions that the template shapes are weighted by are taken from the results measured with the "negative tagged" light jet templates.

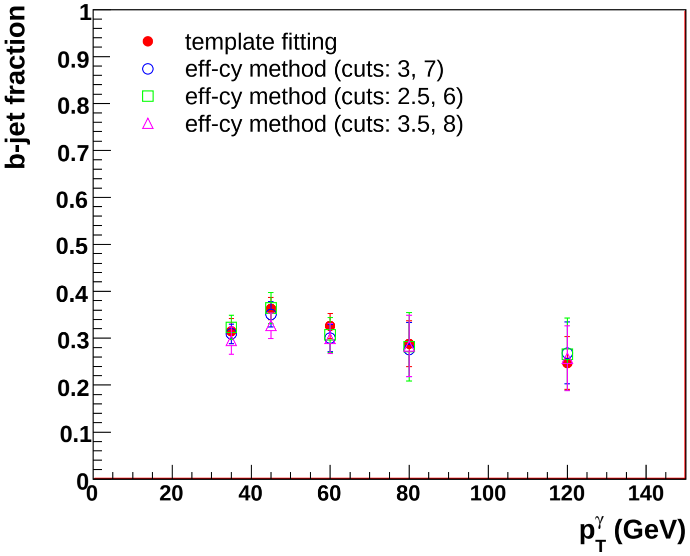


Figure 5.37: The b flavor fractions determined from the efficiency method using three sets of operating point pairs: [3,7], [2.5,6], [3,8]. The found fractions are compared to those found by the template fitting technique.

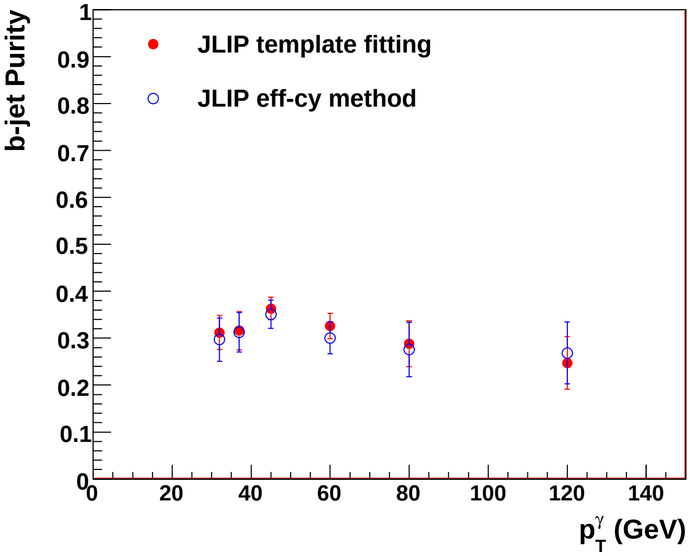


Figure 5.38: The measured flavor fractions for b jets using six p_T^γ intervals. Using this interval choice, the structure in the flavor fraction shape at low values of p_T^γ is confirmed to be a real effect.

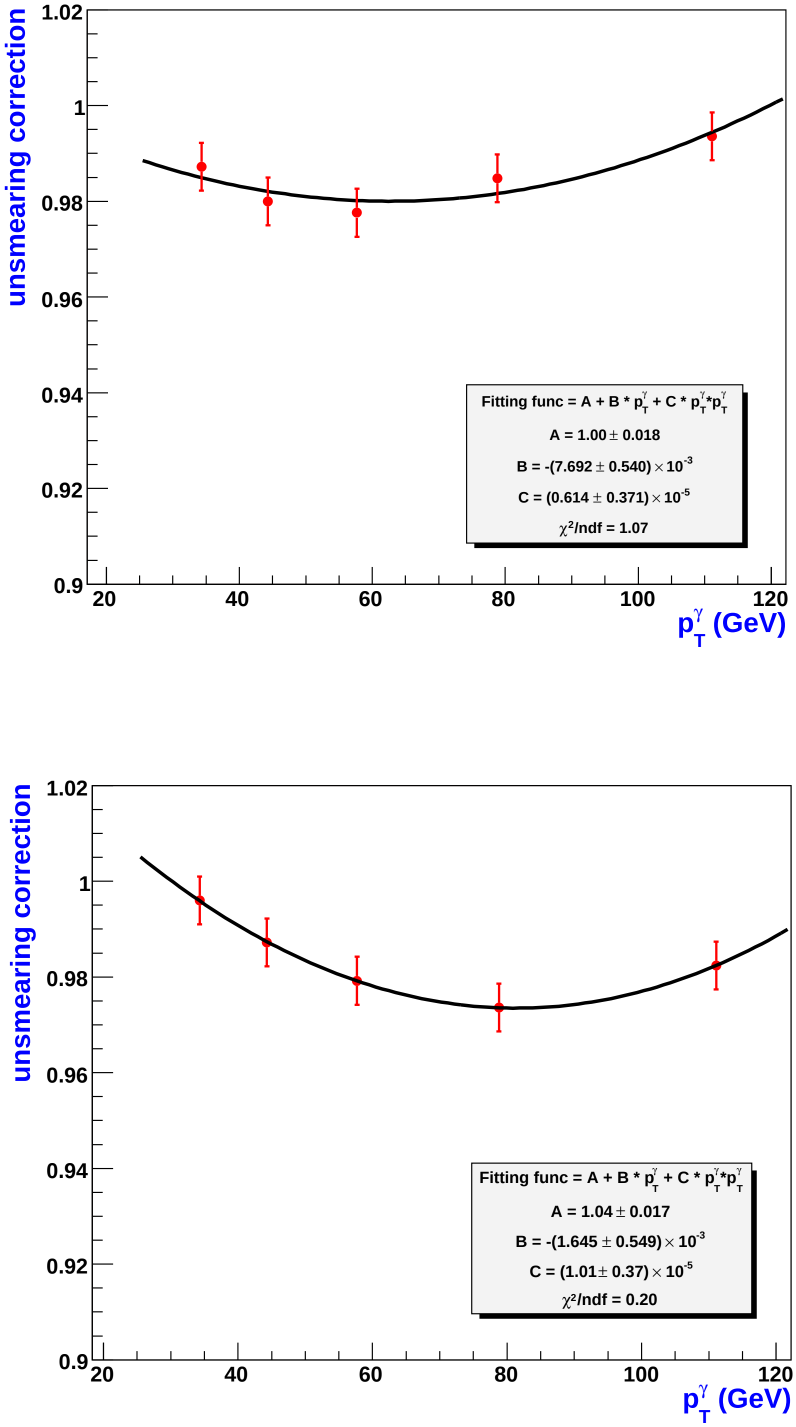


Figure 5.39: The ratio of the unsmeared ansatz to the smeared one calculated for the "$\gamma + c$ jet" (top) and "$\gamma + b$ jet" (bottom) cross sections. Their parameterizations are determined separately for each case and are shown as the red line in each plot.

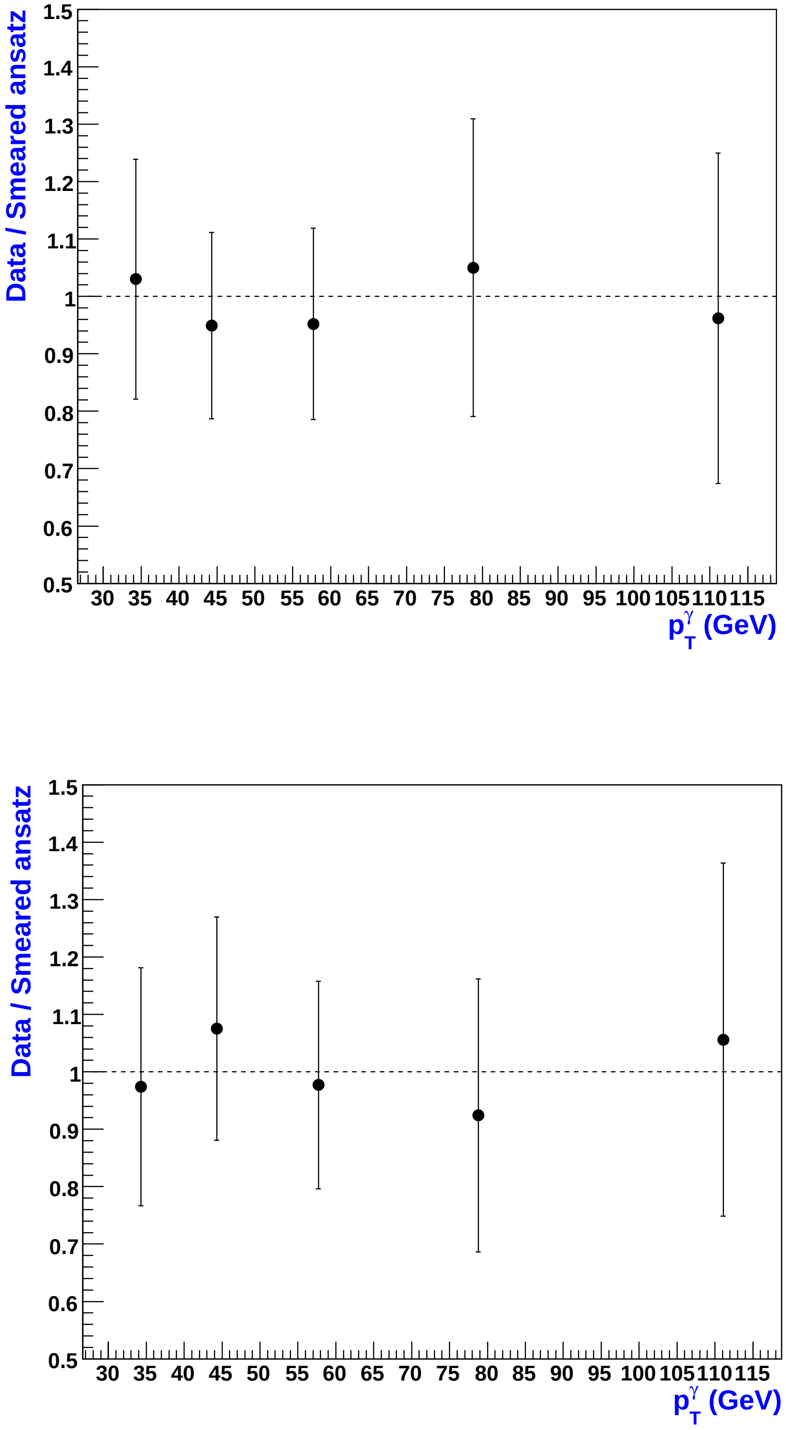


Figure 5.40: The ratio of the smeared ansatz function to the data for the "$\gamma + c$ jet" case (top) and the "$\gamma + b$ jet" case (bottom). The uncertainties shown correspond to the full systematic uncertainties from the measurement.

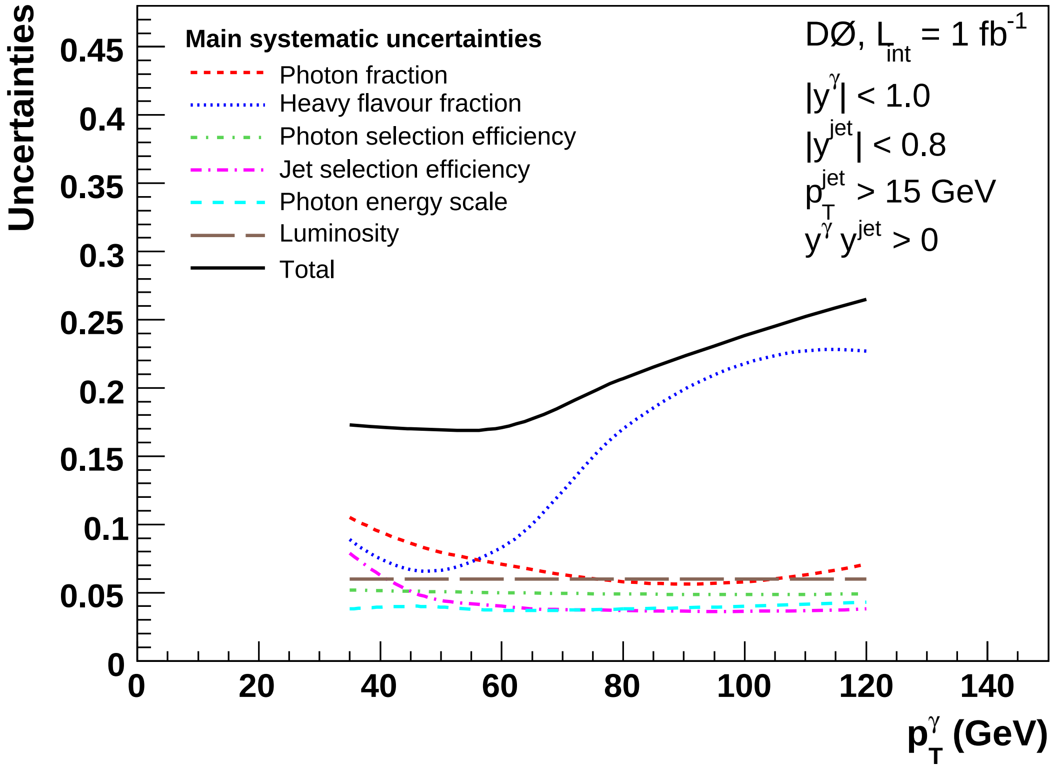


Figure 5.41: The main systematic uncertainties for the differential cross section measured in Region 1 for $\gamma+b$ jet production. Uncertainties for Region 2 differ by $\leq 2\%$. The uncertainties for $\gamma + c$ jet production are of the same size.

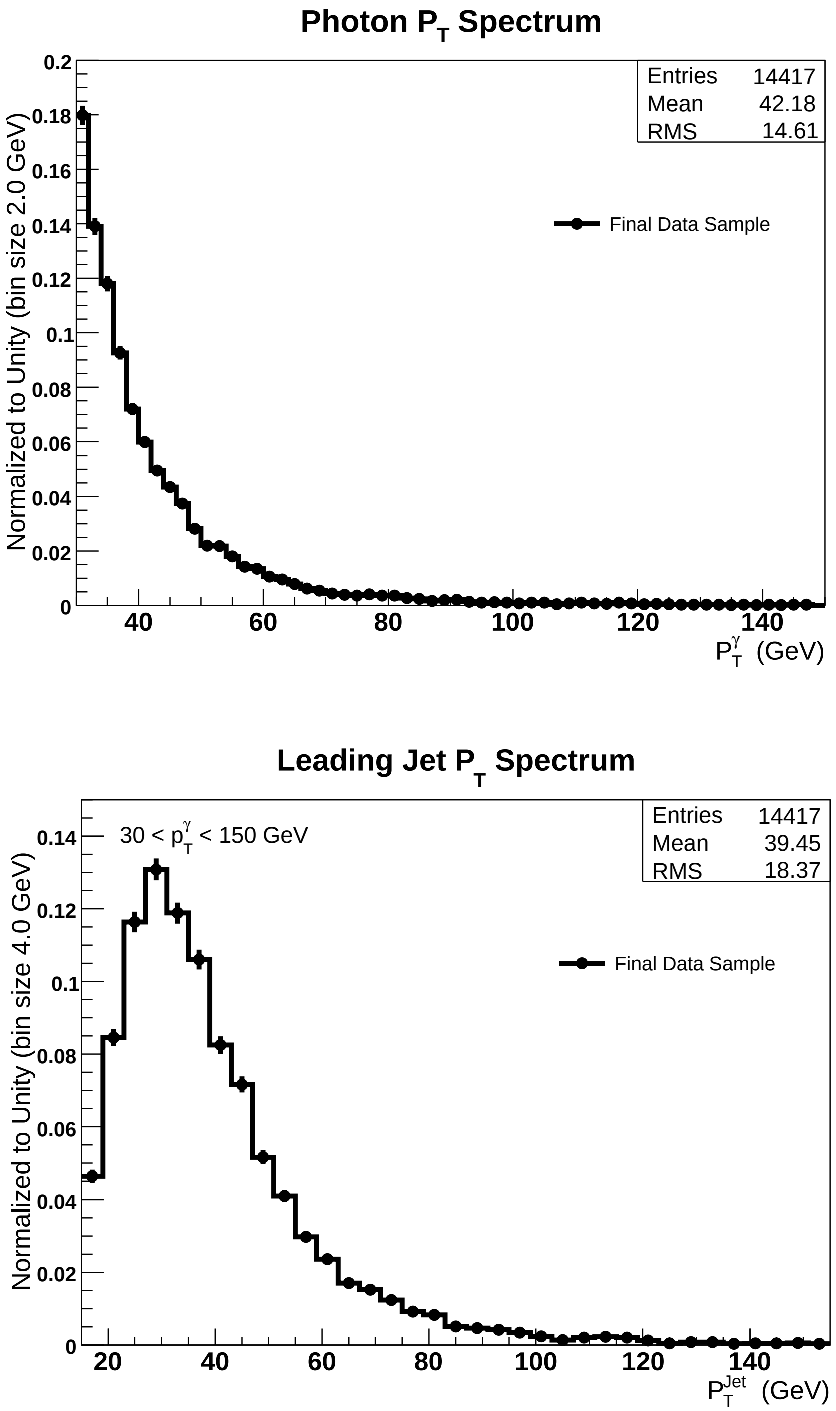


Figure 5.42: The p_T spectra are shown for the leading photon (top) and leading jet (bottom) candidates in the final data sample.

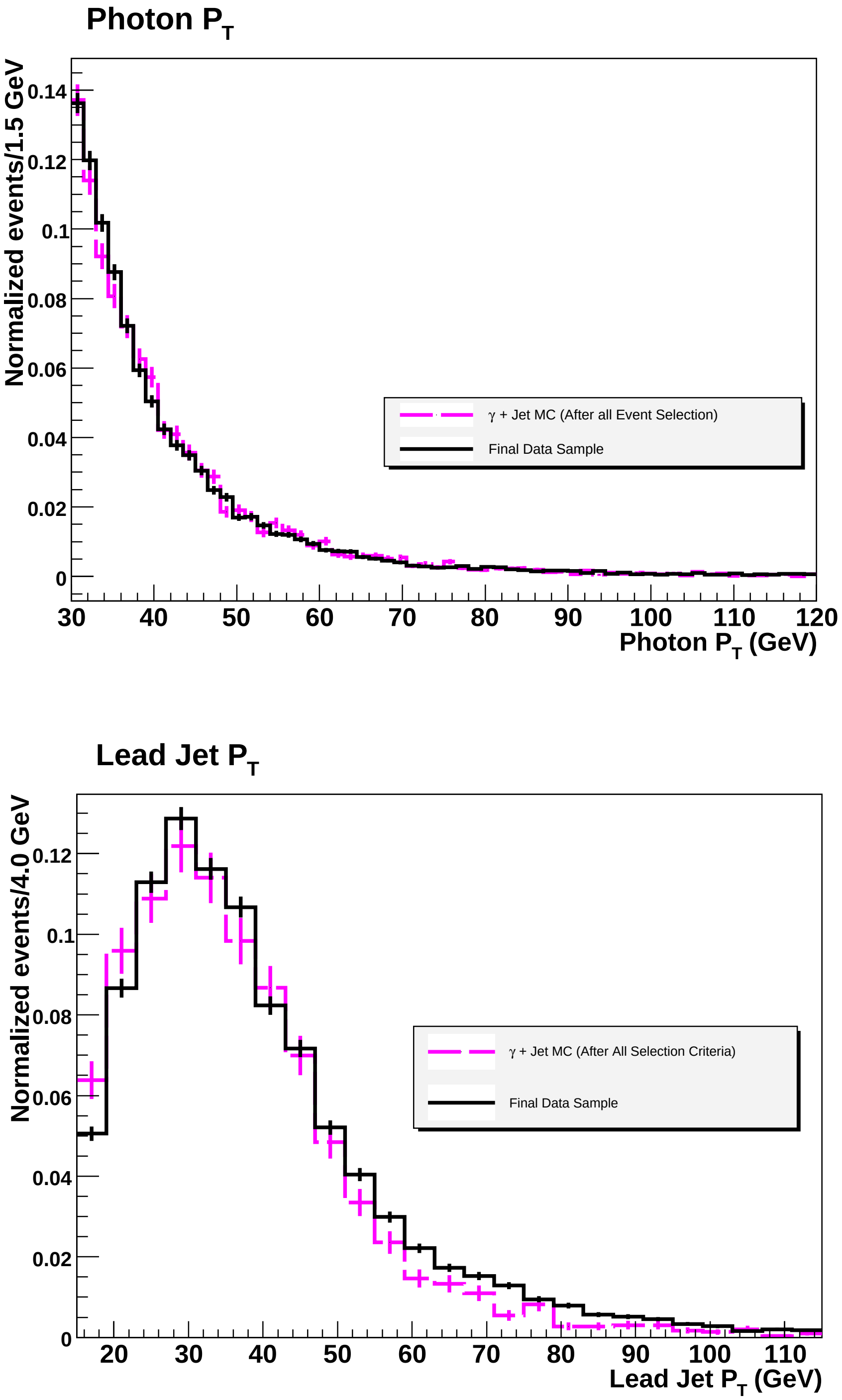


Figure 5.43: The p_T spectra from data are compared to simulation for the leading photon (top) and leading jet (bottom) candidates, and both show good agreement.

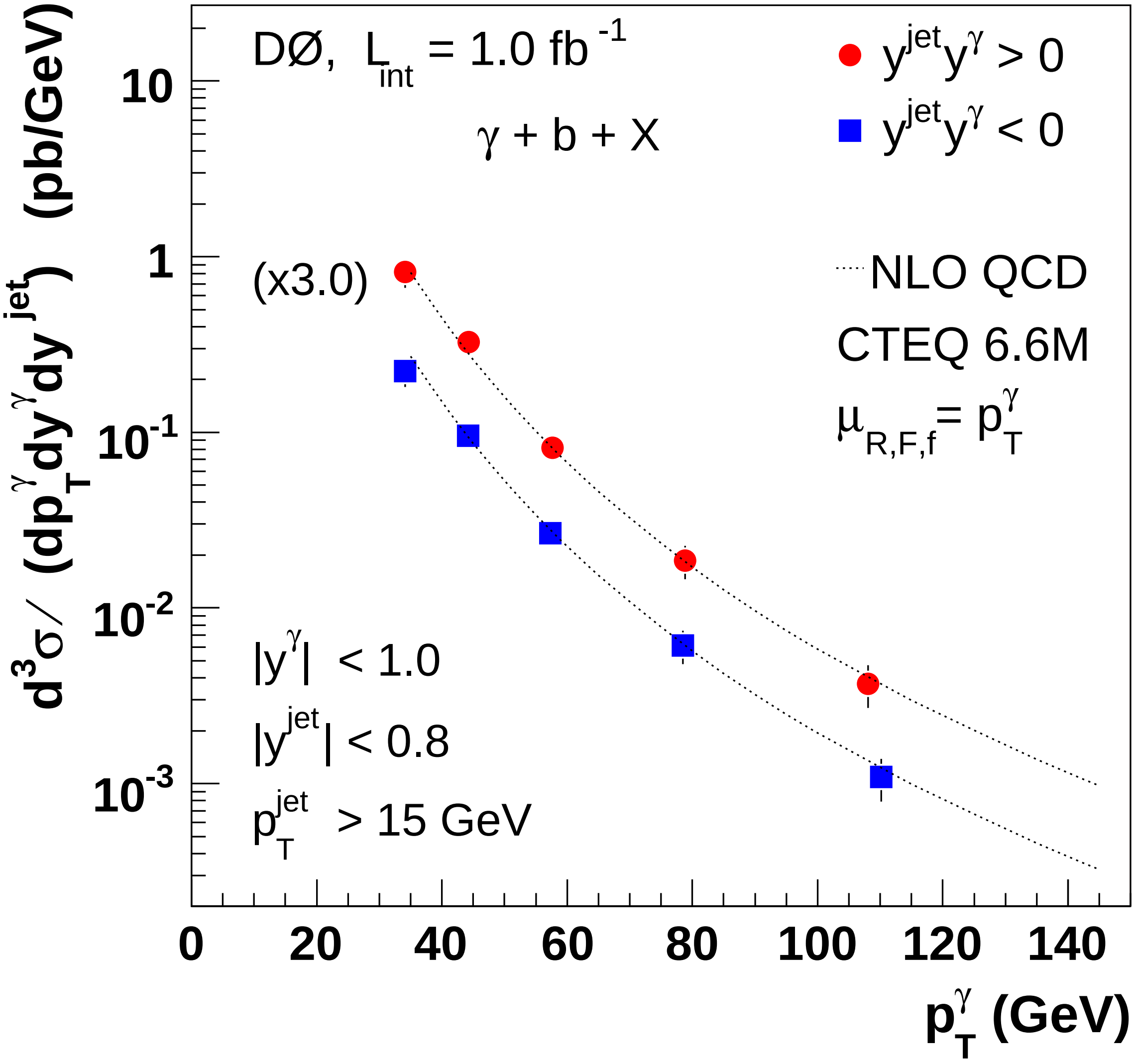


Figure 5.44: The "$\gamma + b$ jet" cross section as a function of p_T^{γ} in Region 1 (multiplied by a factor of three) and in Region 2. The uncertainties on the points in data are the full uncertainties.

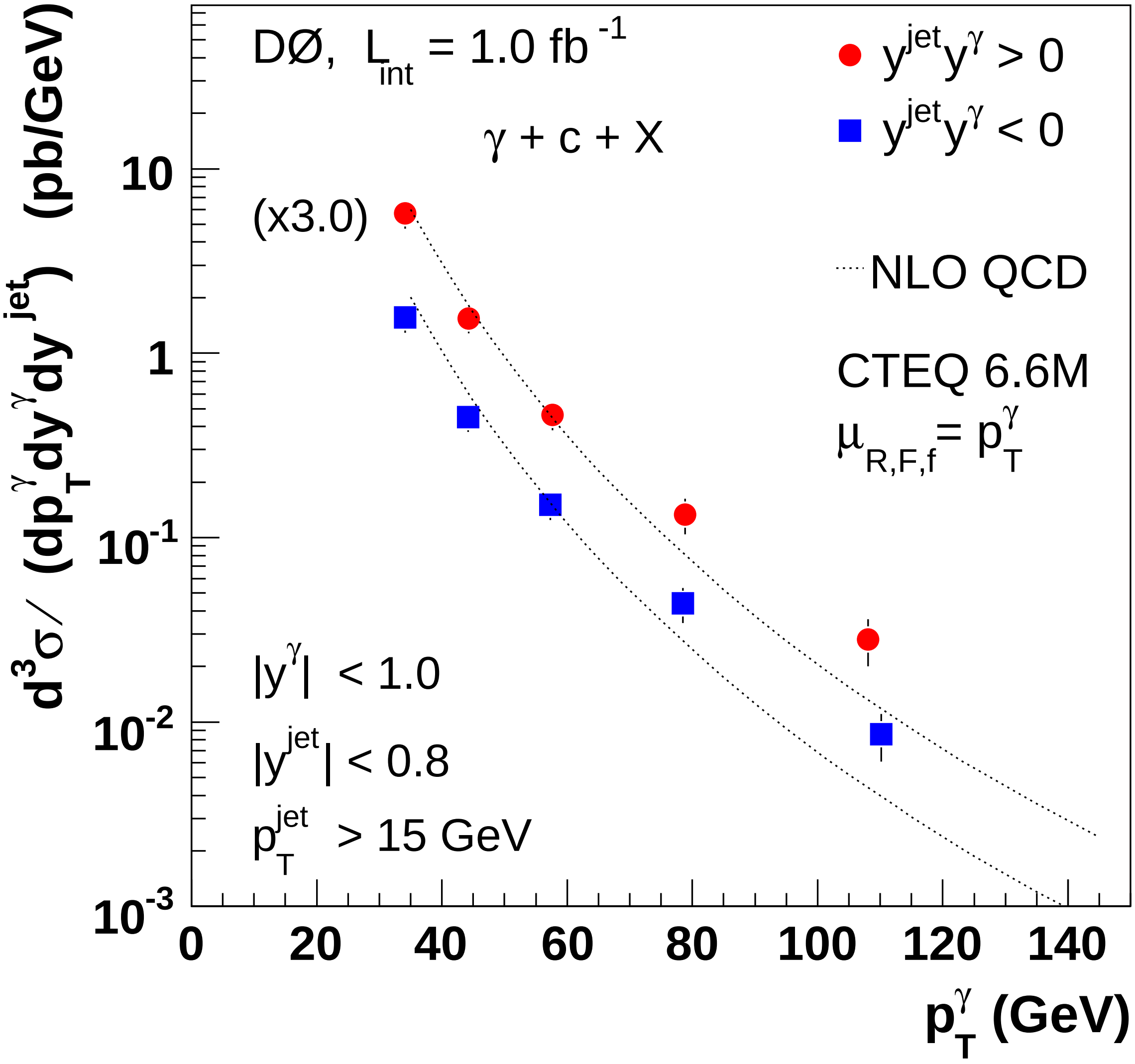


Figure 5.45: The "$\gamma + c$ jet" cross section as a function of p_T^{γ} in Region 1 (multiplied by a factor of three) and in Region 2. The uncertainties on the points in data are the full uncertainties.

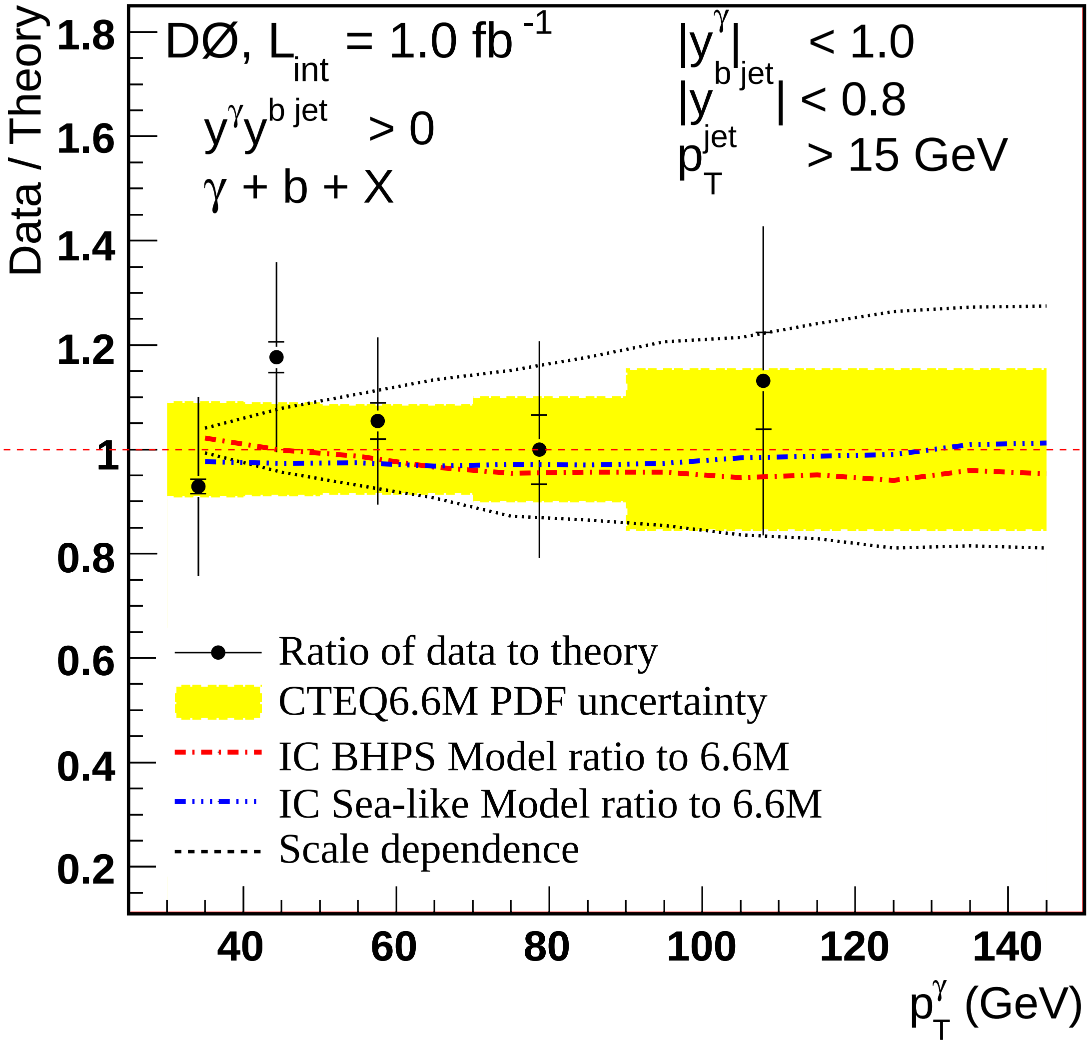


Figure 5.46: The "$\gamma + b$ jet" cross section ratio of data to theory as a function of p_T^{γ} in Region 1. This includes the theoretical scale uncertainties as well as the CTEQ6.6M PDF uncertainties. The uncertainties on the points in data include both statistical (inner line) and the full uncertainties (the entire line).

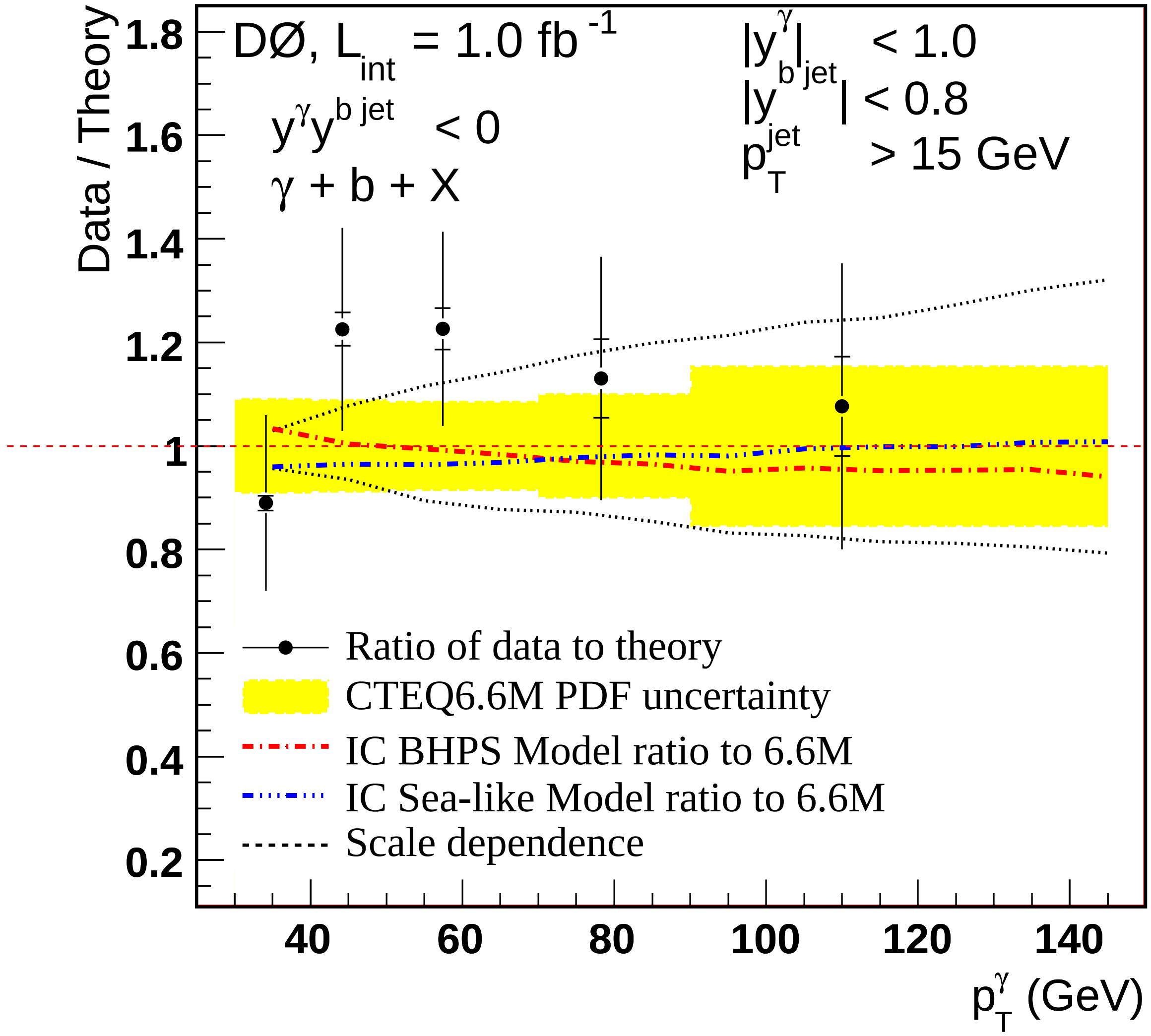


Figure 5.47: The "$\gamma + b$ jet" cross section ratio of data to theory as a function of p_T^{γ} in Region 2. This includes the theoretical scale uncertainties as well as the CTEQ6.6M PDF uncertainties. The uncertainties on the points in data include both statistical (inner line) and the full uncertainties (the entire line).

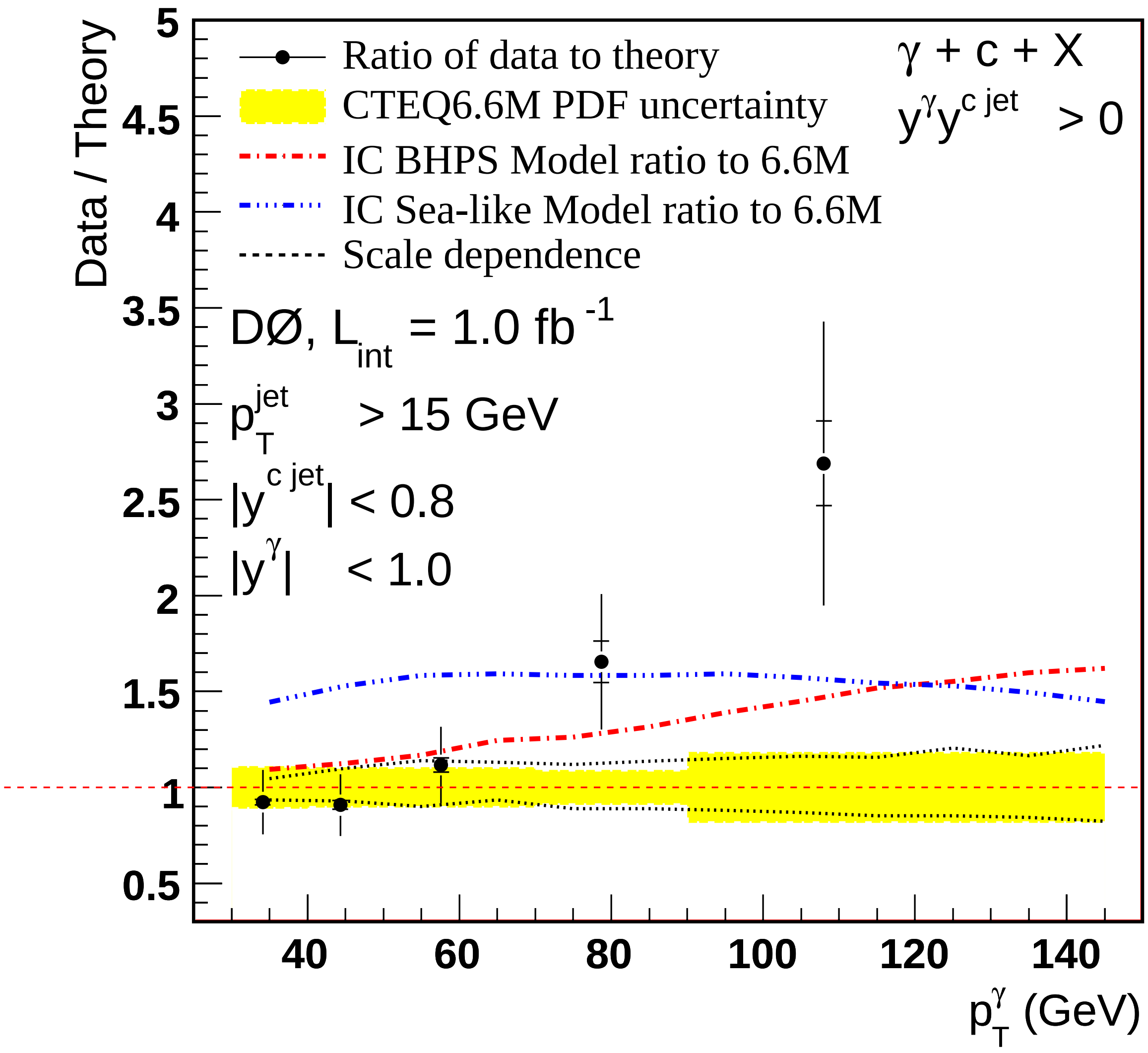


Figure 5.48: The "$\gamma + c$ jet" cross section ratio of data to theory as a function of p_T^γ in Region 1. This includes the theoretical scale uncertainties as well as the CTEQ6.6M PDF uncertainties. The uncertainties on the points in data include both statistical (inner line) and the full uncertainties (the entire line).

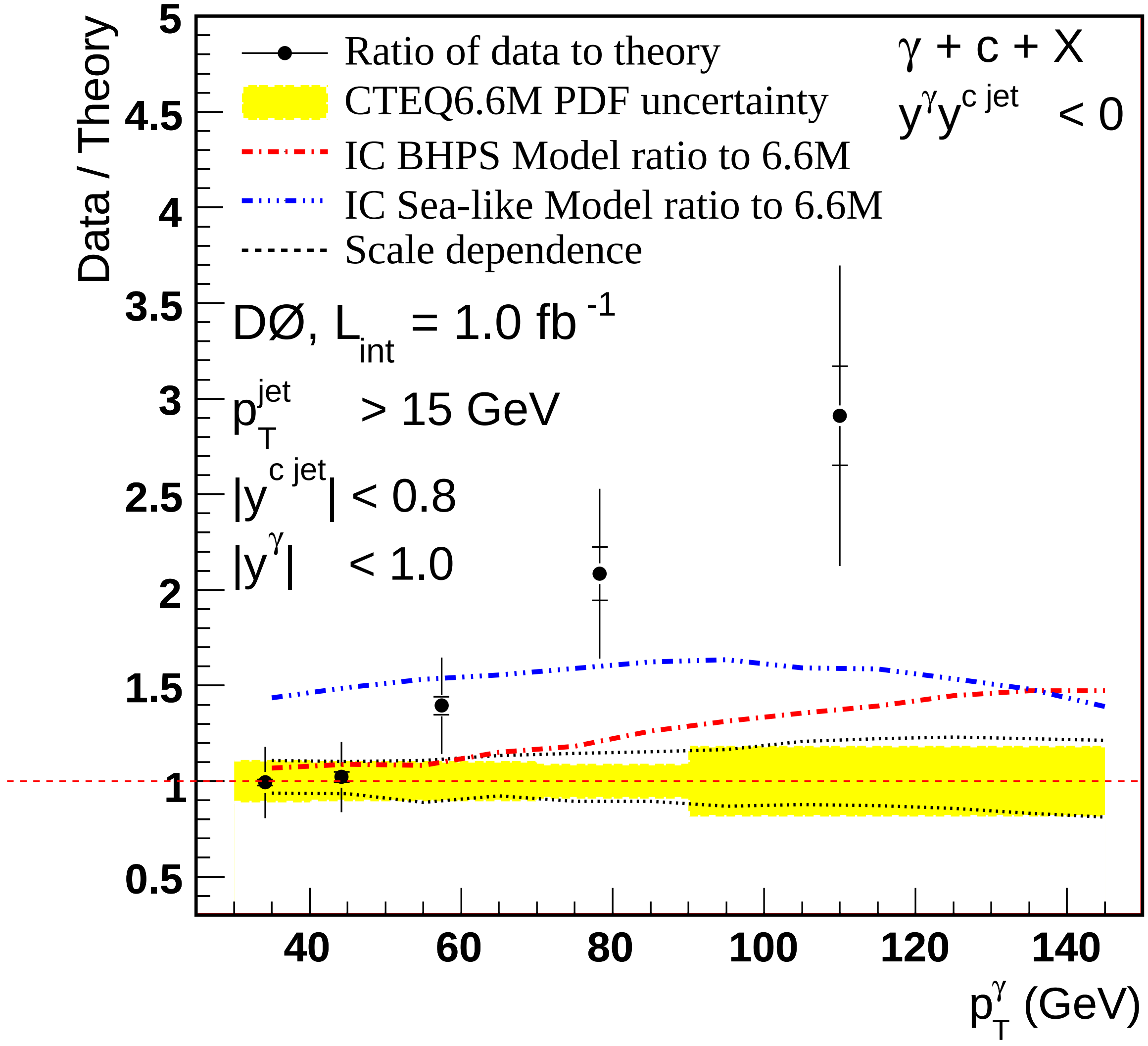


Figure 5.49: The "$\gamma + c$ jet" cross section ratio of data to theory as a function of p_T^γ in Region 2. This includes the theoretical scale uncertainties as well as the CTEQ6.6M PDF uncertainties. The uncertainties on the points in data include both statistical (inner line) and the full uncertainties (the entire line).

CHAPTER 6

CONCLUSION

The triple differential cross sections $\mathrm{d}^3\sigma/(\mathrm{d}p_T^\gamma \mathrm{d}y^\gamma \mathrm{d}y^{\mathrm{jet}})$ have been measured for the production of $\gamma+b$ jets and $\gamma+c$ jets from proton-antiproton collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. These measurements are presented for photon transverse momenta between 30 and 150 GeV, with photon rapidities of $|y^\gamma|<1.0$ and jet rapidities of $|y^{\mathrm{jet}}|<0.8$. This is the first measurement of the differential cross section for associated photon and heavy flavor production at any hadron-hadron collider. The measured cross sections cover the kinematic range of parton momentum fractions x of $0.01 < x < 0.35$ and with $0.9 \cdot 10^3 < Q^2 < 2 \cdot 10^4$ GeV^2 and provide information about b, c and gluon PDFs in this region.

The measured cross sections were compared to NLO QCD predictions with the CTEQ6.6M PDF set [24]. The results for "$\gamma + b$ jet" production show agreement with the theoretical predictions, within uncertainties, for the entire range of p_T^γ in both rapidity regions. However, there is a noticeable disagreement for "$\gamma + c$ jet" production starting from $p_T^\gamma > 70$ GeV for both rapidity regions that increases with p_T^γ.

As referenced in Chapter 2, these discrepancies may be explained by an underestimate of the fragmentation of gluons to $c\bar{c}$ pairs. Another explanation is that there is an intrinsic charm component inside the proton and antiproton, that, depending on its distribution, gives an additional contribution to the cross section.

APPENDIX A

CROSS SECTION COMPARISONS

A.1 bNN > 0.85 and bNN> 0.20 Regime Comparisons

In Section 5.6 we have presented results on the "$\gamma + c$ jet" and "$\gamma + b$ jet" cross sections calculated using criterion of bNN > 0.85. To verify the stability of those results, we have recalculated the cross section using the same technique, but imposing a criteria of bNN > 0.20. This is done to test the strength of the flavor fraction fitting technique because the loosening of this cut will allow a much larger fraction of light jets to enter in the sample. It also changes the shape of the rJLIP distribution from data and in the flavor jet simulation samples. Thus, both the results of the fitting technique itself and the jet flavor simulation shapes can be compared and verified. The results of such a comparison are shown in Figs. A.1 and A.2 for Regions 1 and 2, respectively. The plots show ratio of the "$\gamma + b$ jet" and "$\gamma + c$ jet" cross sections obtained in the two regimes with account of uncertainties on the $b(c)$ fractions, b-tagging efficiencies for $b(c)$ jets, and statistical uncertainties. All other uncertainties should cancel with the ratio as they have remained unchanged from the shift in bNN operating points. Agreement within uncertainties is found between the two regimes.

A.2 Inclusive Cross Section Comparisons

Studies were done to compare the "γ+jet" inclusive cross sections calculated with jet cone sizes of $\mathcal{R} = 0.7$ and $\mathcal{R} = 0.5$. To calculate the inclusive cone$_{\mathcal{R}=0.5}$ cross section, we removed the requirements of taggability and b-tagging on the final data sample, and we adopted the primary vertex cut of 50 cm and its efficiency parameterization used in the inclusive "γ+jet" measurement [48]. The results agree within 3% percent for all p_T^γ as shown in Fig. A.3.

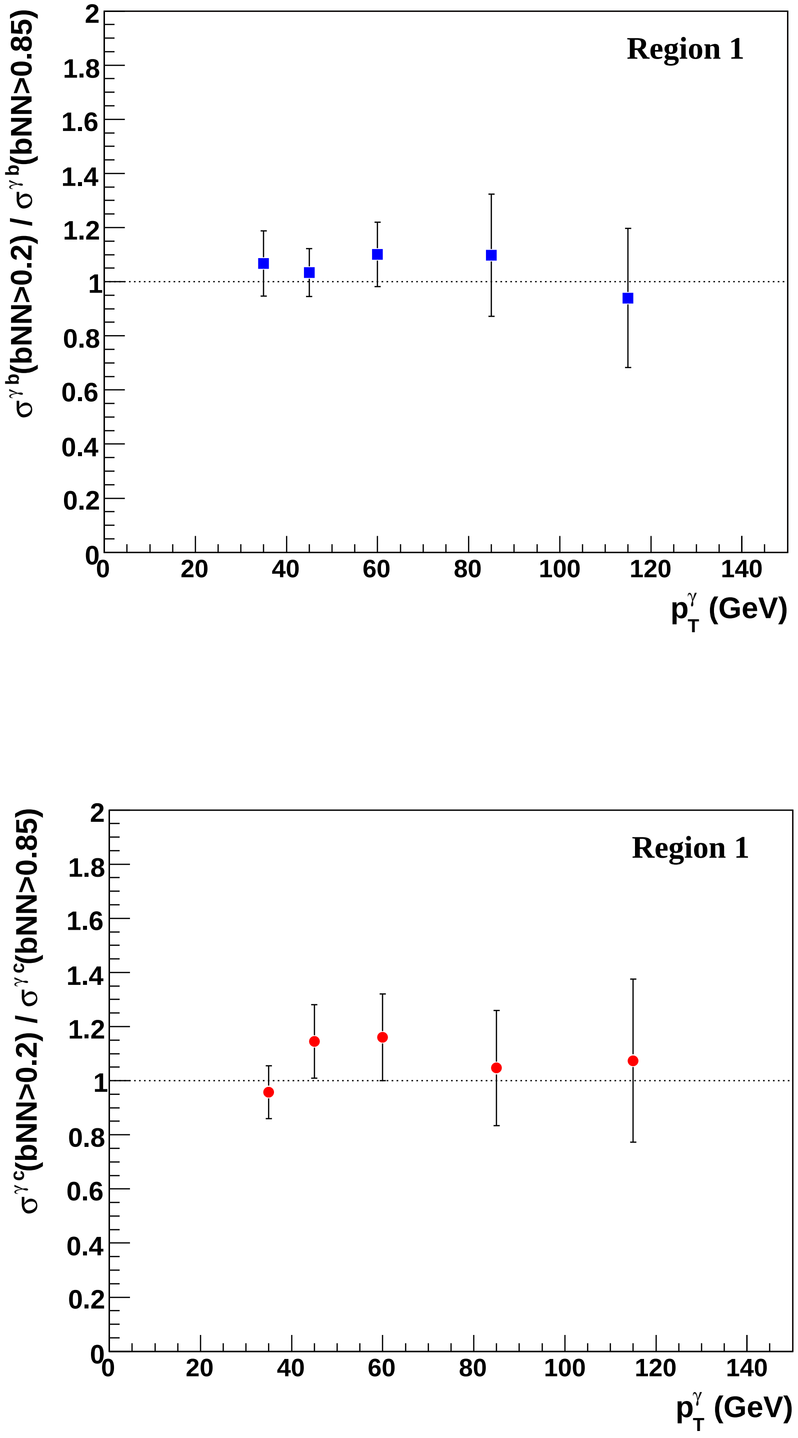
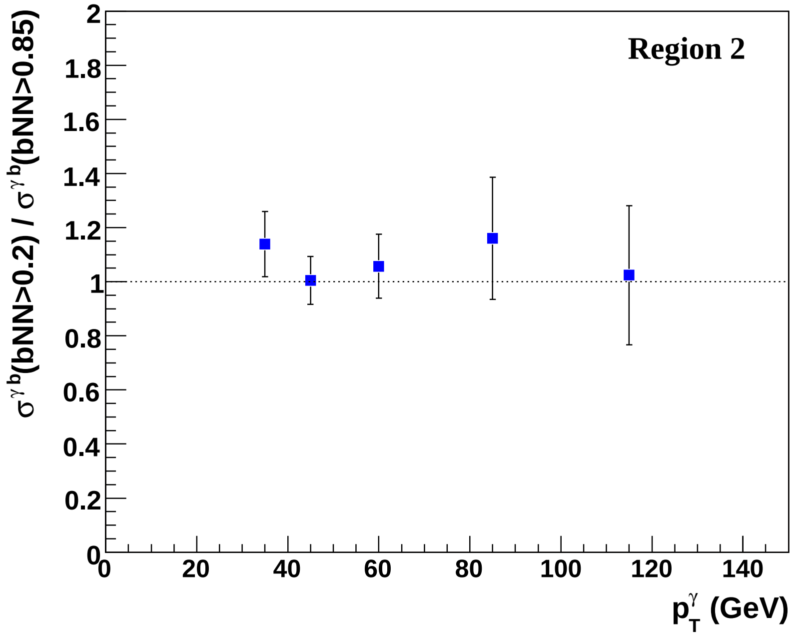


Figure A.1: Ratio of $\gamma + \mathrm{b}$ (top) and $\gamma + \mathrm{c}$ (bottom) cross sections calculated with operating points bNN>0.2 and bNN>0.85 in Region 1.



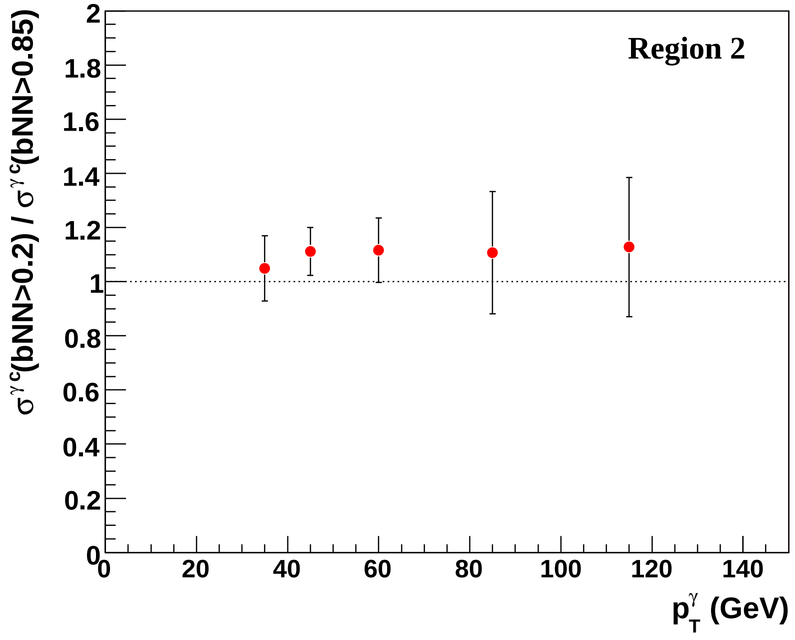
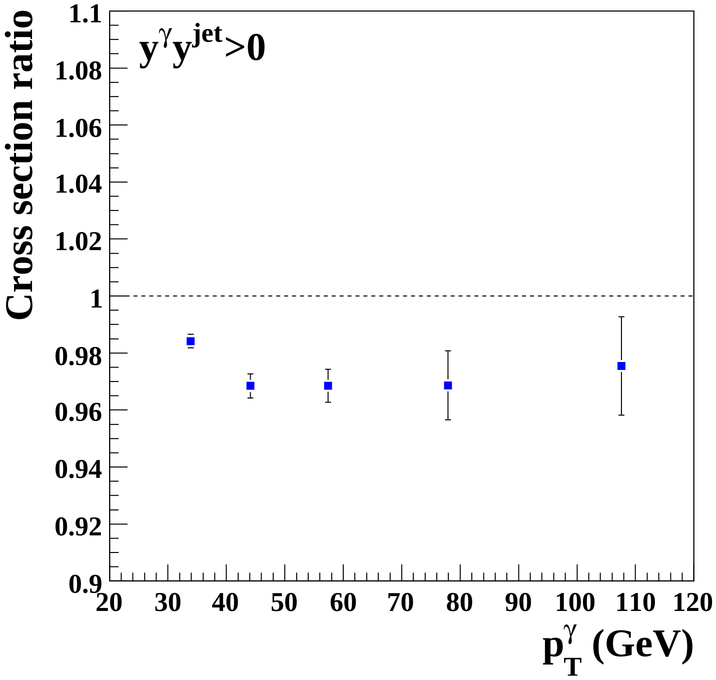


Figure A.2: Same as in Fig. A.1 for $\gamma + \mathrm{b}$ (top) and $\gamma + \mathrm{c}$ (bottom), but for Region 2.



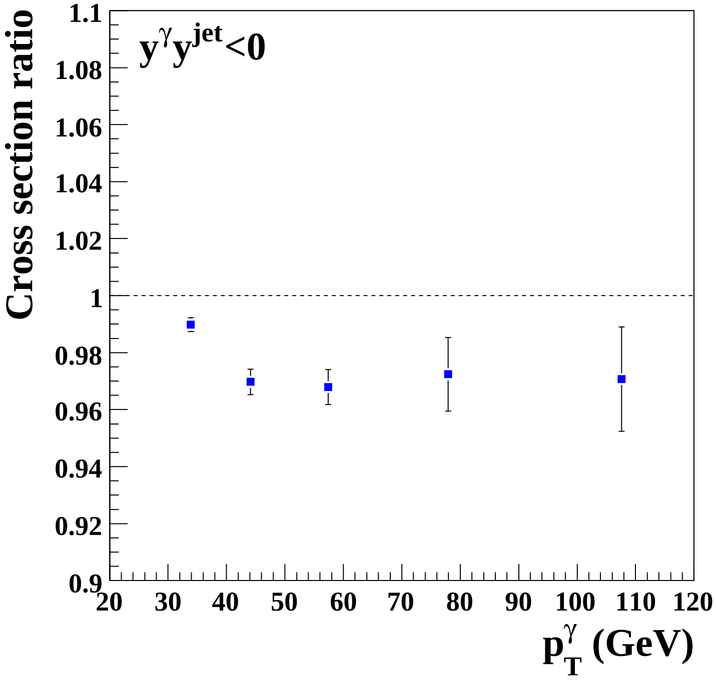


Figure A.3: Ratio of the inclusive cone$_{\mathcal{R}=0.7}$ cross section to the inclusive cone$_{\mathcal{R}=0.5}$ cross section for Region 1 (top) and Region 2 (bottom).

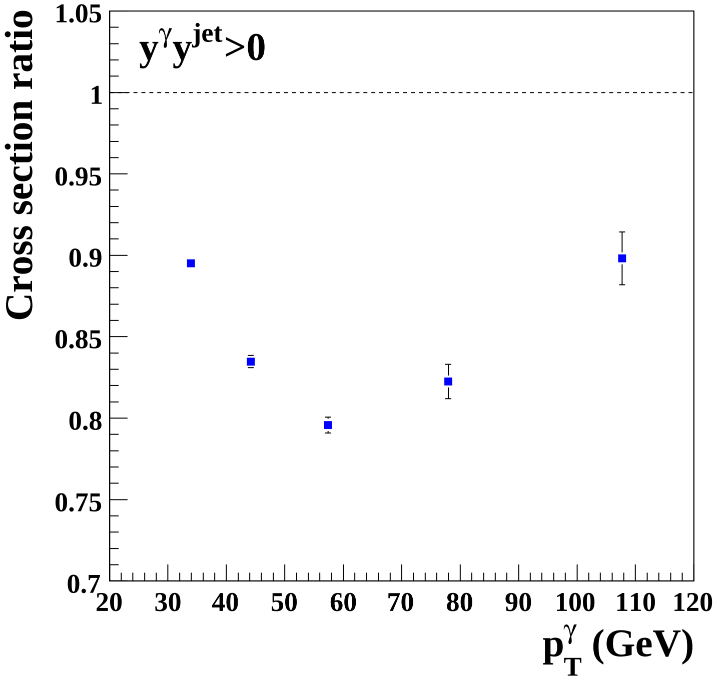
A.3 Inclusive cone$_{\mathcal{R}=0.5}$ cross section vs. taggable cone$_{\mathcal{R}=0.5}$ cross section

Cross checks were also performed to ensure that the introduction of the jet taggability requirement did not result in any bias in the cross section. To this end, the ratio of the taggable cone$_{\mathcal{R}=0.5}$ "γ+jet" cross section to the inclusive cone$_{\mathcal{R}=0.5}$ "γ+jet" cross section was made. Since events with a primary vertex z position outside of 50 cm remained after imposing the taggability requirement, a cut of $|PV_z| < 50$ cm was applied to the taggable jet sample. The efficiency for this criteria was $\sim 99.9\%$ and was flat in both p_T^γ and instantaneous luminosity. The resulting ratio showed an excess in the taggable cross section up to $\sim 20\%$ compared to the inclusive "γ+jet" cross section, as shown in Fig. A.4.

Because taggability is calculated as a function of jet p_T, η and the primary vertex z position, we plotted the average efficiency against these quantities. The resulting efficiency plots showed an irregular and low taggability efficiency for primary vertex values $|PV_z| >$ 35 cm, shown in Fig. A.5. We applied a cut of $|PV_z| < 35$ cm for the taggable cross section case and found a flat efficiency in both p_T^γ and instantaneous luminosity of 95.3%. The cross section was recalculated and the ratio taken with respect to the inclusive cone$_{\mathcal{R}=0.5}$ "γ+jet" cross section. Fig. A.6 shows that the agreement between cross sections is very much improved and within a few percent for all values of p_T^γ. Due to the erratic nature of these efficiencies and the potential effect on the b-tagged cross section, the same $|PV_z| < 35$ cm criteria was imposed to safeguard against a possible bias from the more unstable and lower taggability efficiencies, which is also shown in Fig. A.5.

A.4 Taggable Cross Section vs. Flavor Summed Cross Section

As a final cross check, the taggable cone$_{\mathcal{R}=0.5}$ "γ+jet" cross section was compared to the sum of the $\gamma + b$ jet, $\gamma + c$ jet, and $\gamma+$ light jet cross sections. Fig. A.7 shows the resulting ratio of this flavor summed cross section to the taggable cross section. The ratio shows an excess for the lower p_T^γ range, and agreement within uncertainties in both photon-jet rapidity regions for $p_T^\gamma > 70$ GeV. The most likely explanation for this is due to the light jet efficiency's strong p_T^γ dependence and the large scaling differences from our calculations and the official ones given from the b-ID group, as shown in Fig. A.8. The p_T^γ dependence of the light jet $\epsilon_{\text{b-tagged}}$ is not seen in either the c or b jet case. Additionally, the cross section correction factor derived from the ratio of these efficiencies shows differences of $\sim 10 - 20\%$ more than that of the c or b jets. Because the inclusive "γ+jet" cross section is dominated by $\gamma+$ light jets ($> 92\%$), these effects can potentially play a large role in the poor estimation of the inclusive cross section using a combined jet flavor cross section.



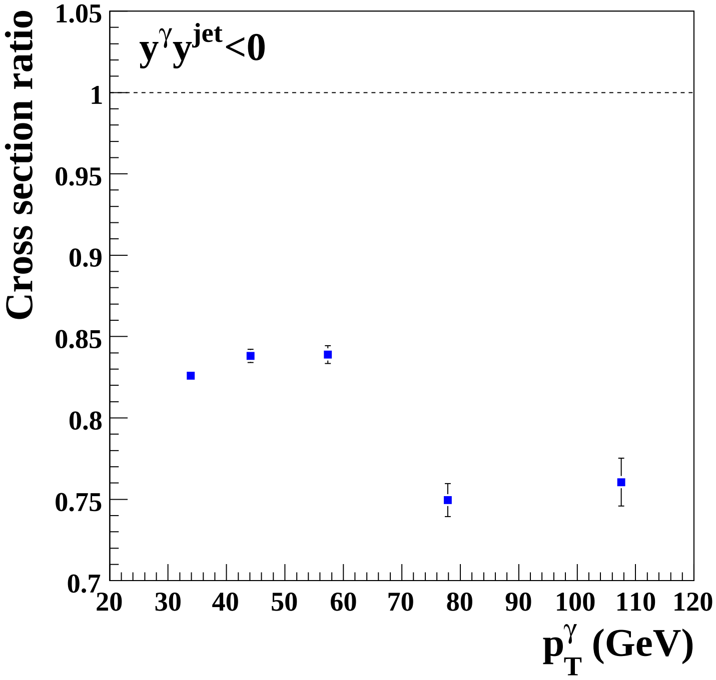
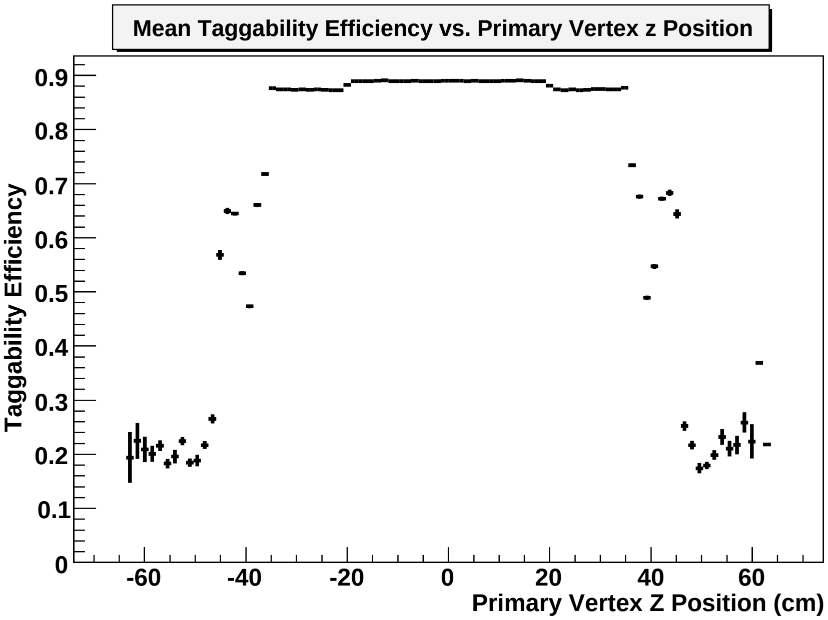


Figure A.4: Ratio of the taggable cone$_{\mathcal{R}=0.5}$ cross section with a primary vertex z cut at 50 cm to the inclusive cone$_{\mathcal{R}=0.5}$ cross section for Region 1 (top) and Region 2 (bottom).



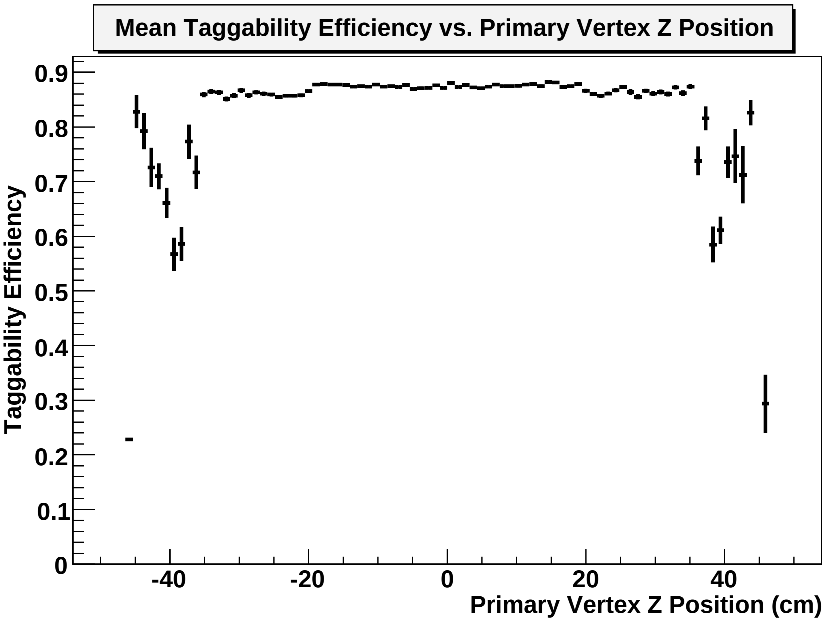
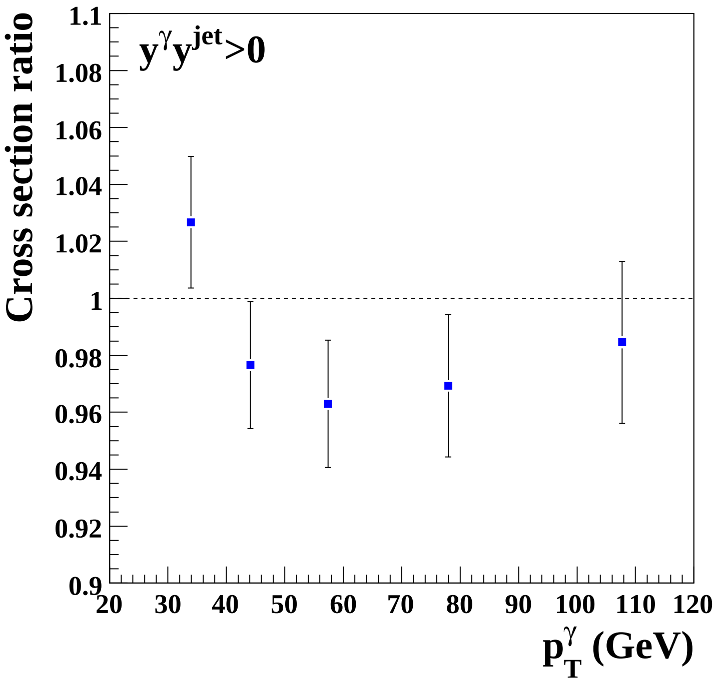


Figure A.5: Taggability efficiency as a function of the primary vertex z position in both the inclusive taggable case (top) and the b-tagged case (bottom).



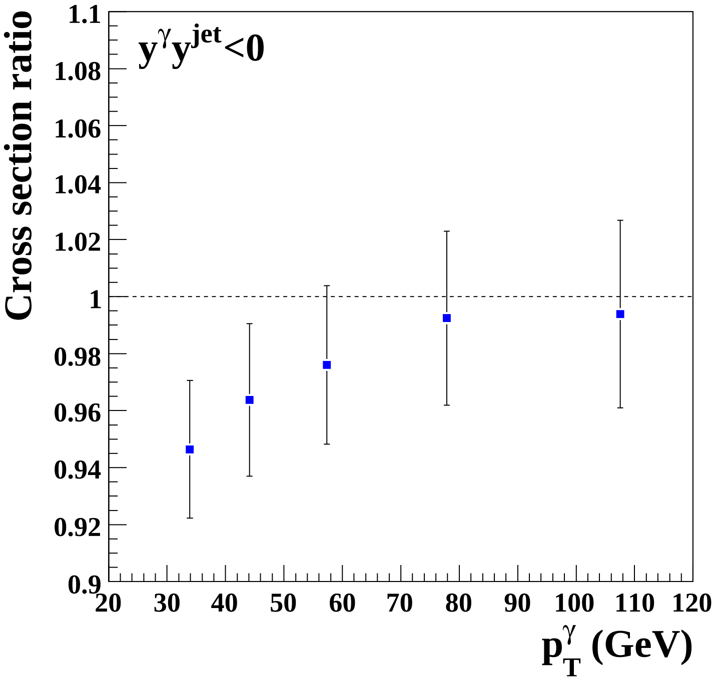


Figure A.6: Ratio of the taggable cone$_{\mathcal{R}=0.5}$ cross section with a primary vertex z cut at 35cm to the inclusive cone$_{\mathcal{R}=0.5}$ cross section for Region 1 (top) and Region 2 (bottom).

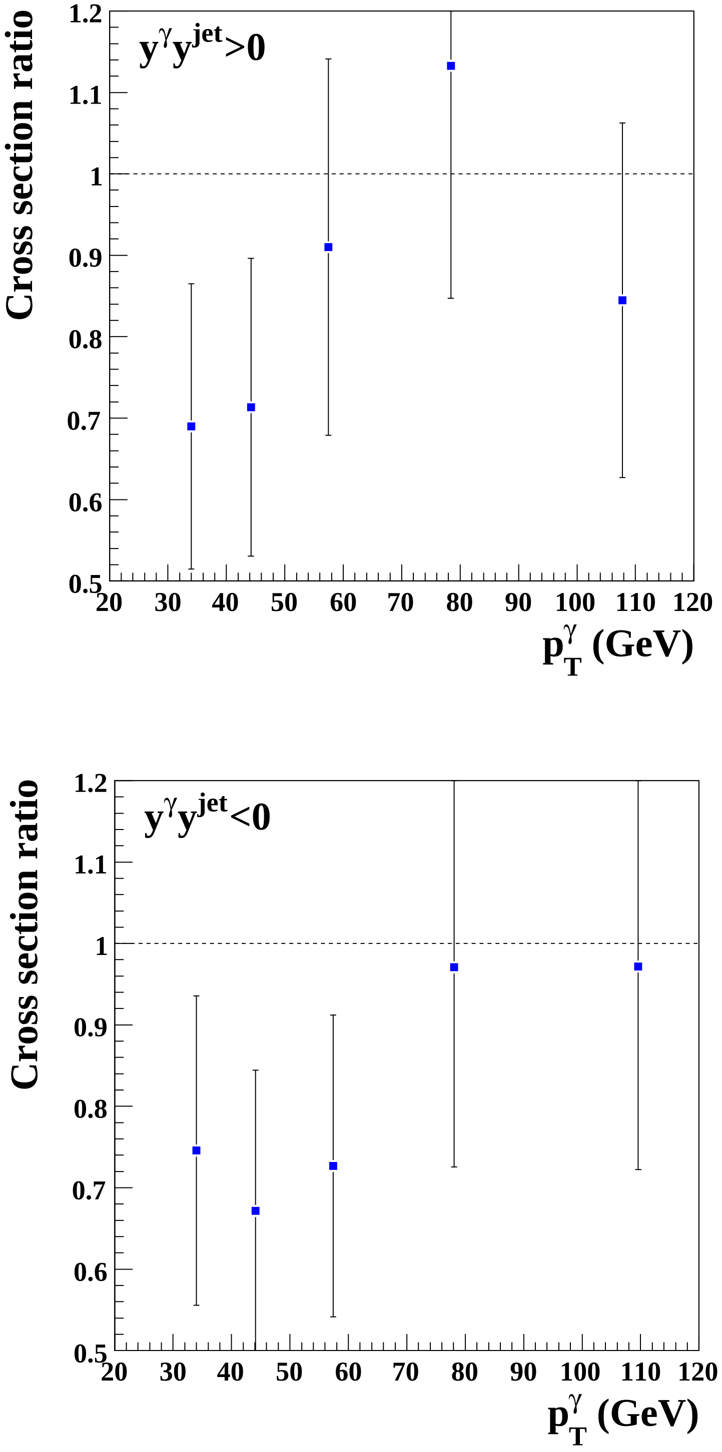


Figure A.7: The ratio of the taggable cone$_{\mathcal{R}=0.5}$ cross section over the cross section calculated from the individual jet flavor cross sections for Region 1 (top) and Region 2 (bottom).

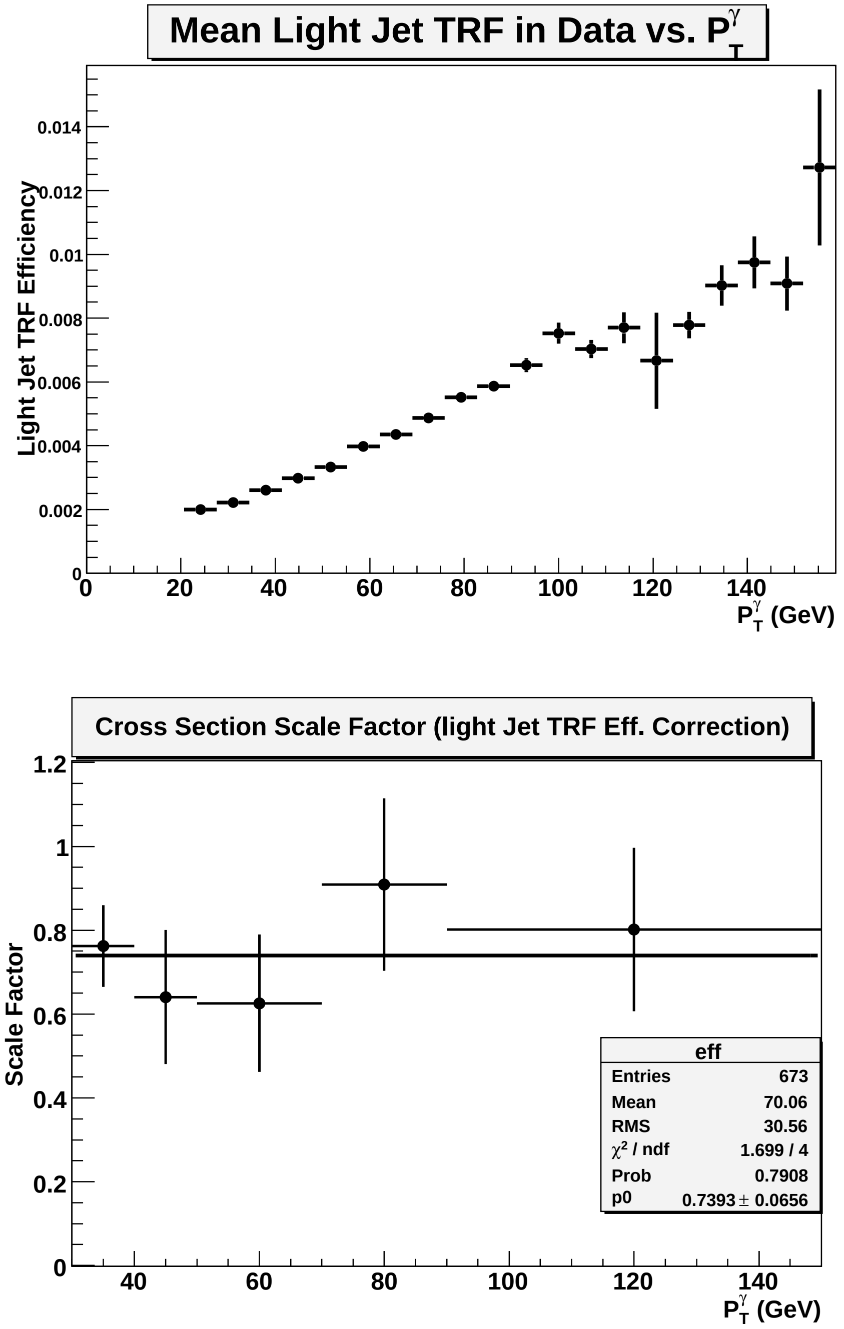


Figure A.8: The light jet b-tagging efficiency has a large dependence on p_T (top) and the correction factor determined between the signal process efficiencies and the official light jet efficiency (bottom) is sizable and has large statistical uncertainties.

APPENDIX B

HEAVY FLAVOR IDENTIFICATION CROSS CHECKS

In Section 5.4.9, the technique of the flavor fraction determination was detailed for the "$\gamma+b$ jet" and "$\gamma+c$ jet" cross sections. To test the robustness of these results, as well as the rJLIP variable itself, further investigations were performed. The following cross-checks [49] confirm the initial findings.

(1) To test the accuracy of the fitting procedure, we performed a closure test using simulated events containing light, c, and b flavored jets. These events were preselected using the criterion of bNN > 0.20. Afterward they were mixed into a single sample, where events containing each type of jet flavor had a 33.3% contribution to the overall number of events. The events in this sample were next required to pass the bNN > 0.85 criterion. The resulting fractions of each flavor were known because the jets were identified by their flavor. Separately, the sample of events passing the bNN > 0.85 requirement was fit using the rJLIP fitting technique, and the resulting flavor fractions compared to the "true" results from simulation are shown in Fig. B.1. We see that the fitted fractions are in very good agreement with expectations.

(2) To test the ability of the rJLIP variable to describe the data, the c and b jet fractions were calculated in two other regimes: replacing the bNN > 0.85 requirement with one of bNN > 0.20, and in this regime requiring an identified muon [60] to be found within the reconstructed jet[1]. The first requirement lowers the flavor fractions of both c and b jets due to the larger fraction of light jets allowed from loosening this operating point. The muon requirement, however, increases the heavy flavor fractions as the mesons within these jets

---

[1] Due to the limited statistics of this cross-check, only the first three p_T^{γ} bins were used.

are much more likely to decay semileptonically. The obtained flavor fractions for the two cases are shown in Fig. B.2. These fitted fractions were used to weight sum of light, c and b jet rJLIP distributions to the data, which exactly follows the prescription used in Figs. 5.34 and 5.35. The results for the first requirement are shown in Figs. B.3 with a linear y-axis scale and B.4 with a $\log_{10}$ y-axis scale. The muon enhanced sample's results are shown in Figs. B.5 and B.6. Note that the points contain only statistical errors and that they do not contain uncertainties from the fitting procedure. The average fraction of light jets in the interval $30 < p_T^\gamma < 50$ GeV for the muon enhanced data sample has been found to be $< 10^{-3}$ and 3.2% for $50 < p_T^\gamma < 70$ GeV.

Both regimes show excellent agreement in the shape comparisons of the data to the summed flavor templates, providing further evidence that the simulation accurately describes the data for this variable and this technique. An advantage of these two specific selections is that the shape of the data and its agreement to the flavor summed templates for values of rJLIP < 3 are largely dependent on the light jet template shape in the first case and the c jet shape in the second case. In both cases, the b jet template drives the summed flavor fraction shape for values of rJLIP > 6, as can be seen in Figs. 5.35 and B.6.

(3) The requirement of an identified muon in the jet reduced the number of events used in the cross check described above, leading to large statistical uncertainty. To reduce this statistical uncertainty, the photon identification requirements were heavily relaxed to increase the number of events in data by a factor of $\sim 2.5 - 3.5$. After this, the prescription of the cross-checks from (2) were repeated, except now at the operating point of bNN > 0.85. Figs. B.7 and B.8 show the results of the first of these comparisons to data with a linear y-axis and a $\log_{10}$ y-axis using the found $b/c/$light fractions but with no muon requirement. The results for the fitting procedure at this operating point when requiring a muon to be within the jet are shown in Figs. B.9 and B.10. As in the previous comparisons, only the statistical uncertainties are shown in these plots. In both regimes, the data is again well described by the summed flavor jet templates.

(4) The last cross-check involved the direct use of the bNN output. For this test, all events were preselected with a bNN> 0.20 requirement. To reduce the sensitivity to the bNN shape

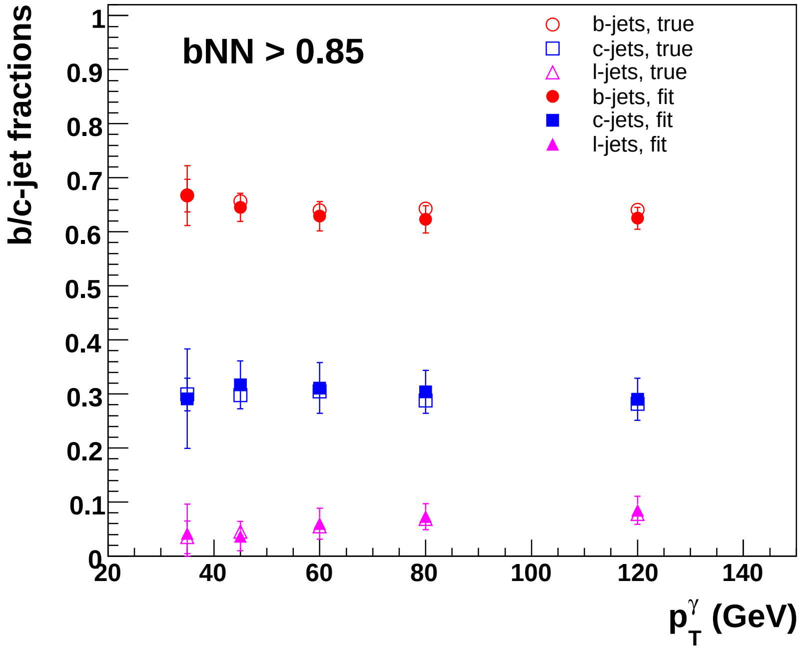


Figure B.1: Closure test of the flavor fractions for all five p_T^γ intervals. Only statistical uncertainties are shown for these points.

for c and b jets, the efficiency technique was used, as described in Section 5.4.9, except now the efficiencies are taken from events passing bNN requirements. The operating point cut pair used was $[0.45, 0.85]$, and the results are shown in Fig. B.11. The heavy flavor fractions for both variables are in good agreement.

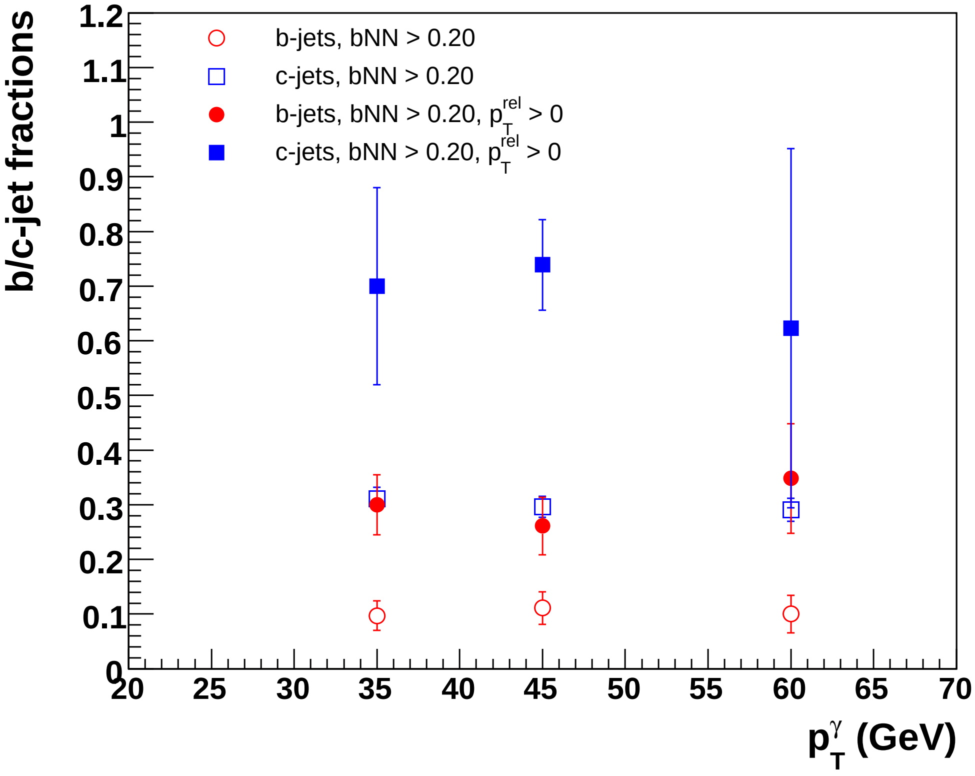


Figure B.2: The c and b jet fractions found for events selected for the case with bNN$>$ 0.20 (open symbols), and for the case with bNN$>$ 0.20 and a muon to be found within the jet (closed symbols).

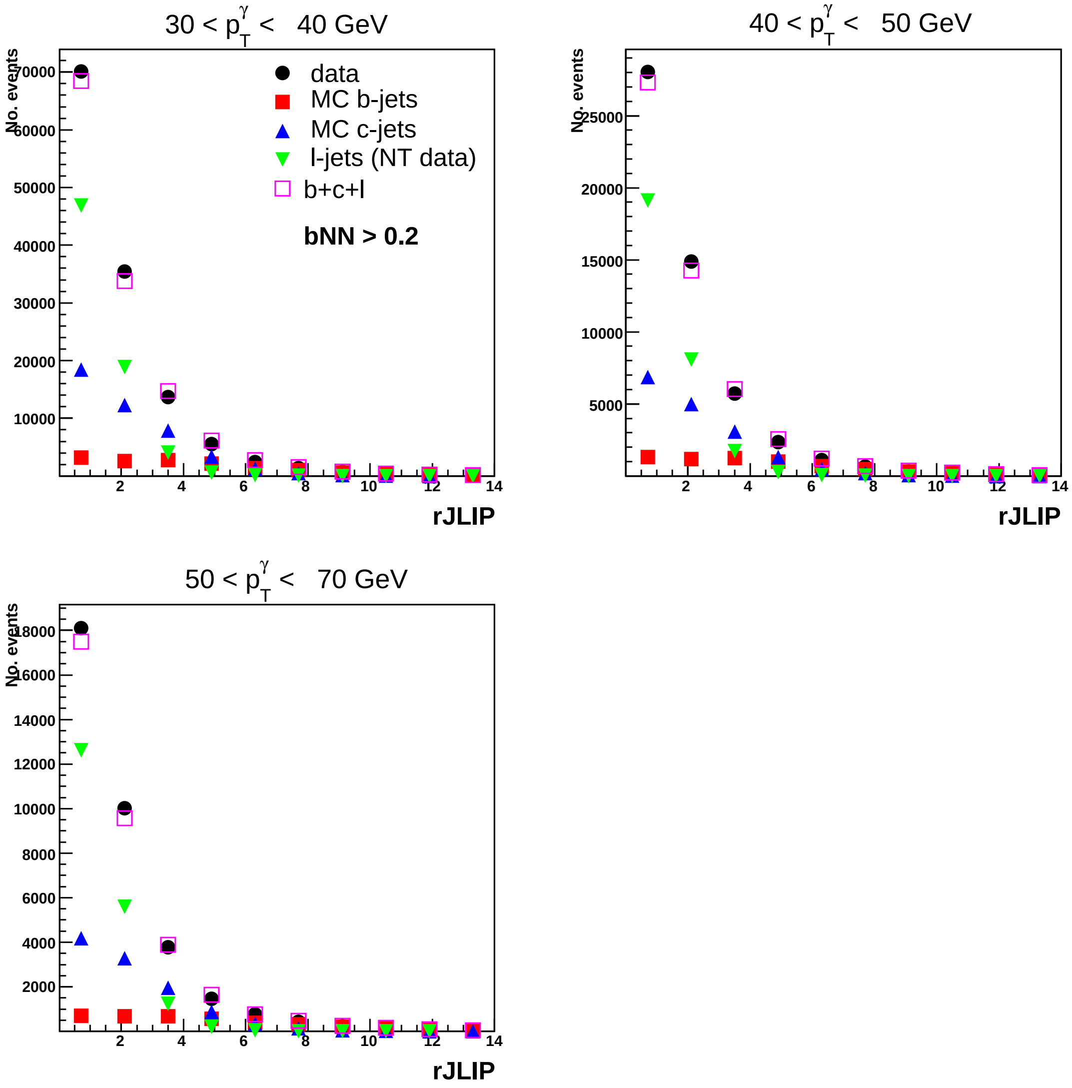


Figure B.3: Distribution of the number of events in data over rJLIP after all selection cuts and with bNN > 0.20. The distributions of simulated b and c jets, and light jets from data are also shown, weighted according to their found flavor fractions. The sum of their distributions are normalized to the data.

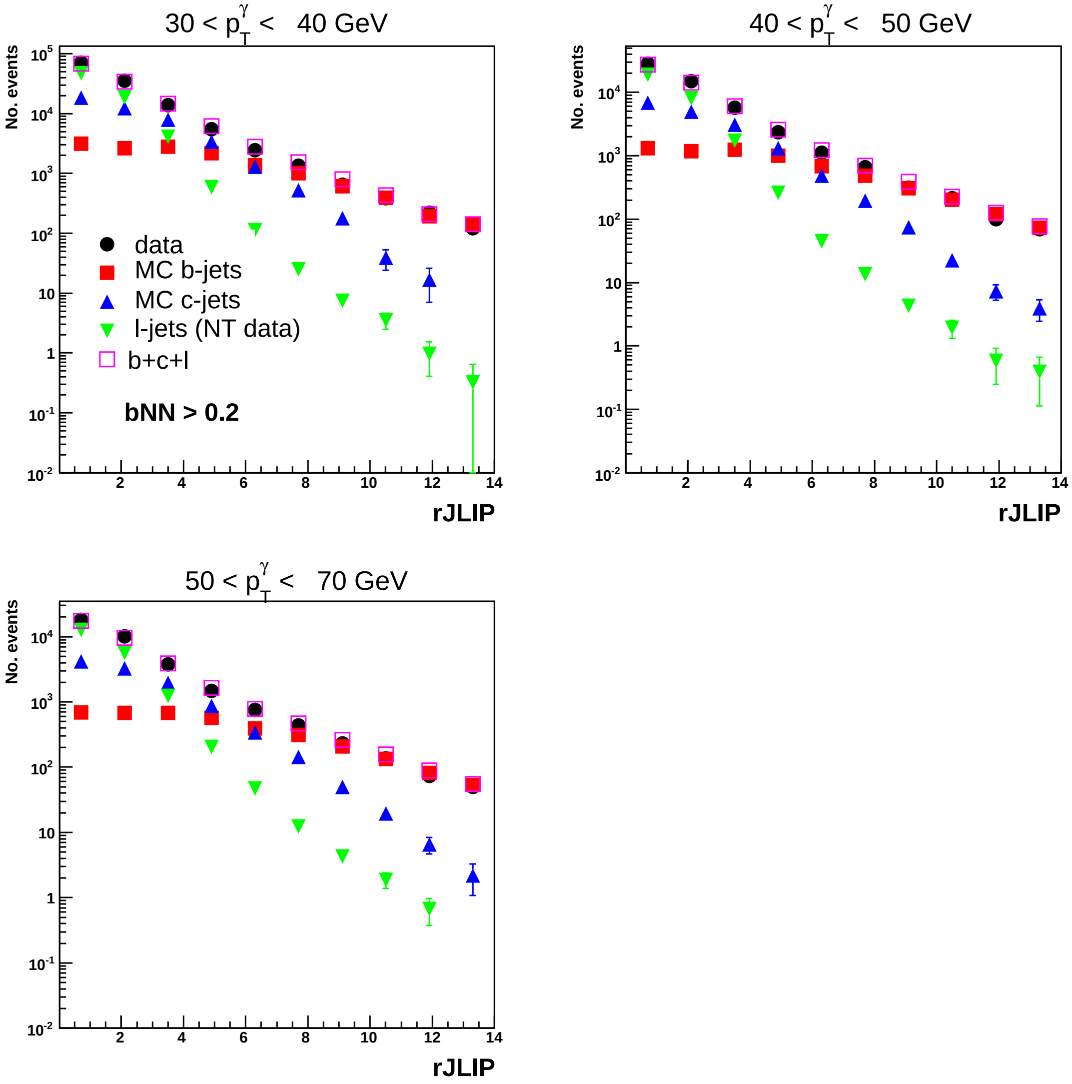


Figure B.4: Same as in Fig. B.3 but with a $\log_{10}$ y-axis scale.

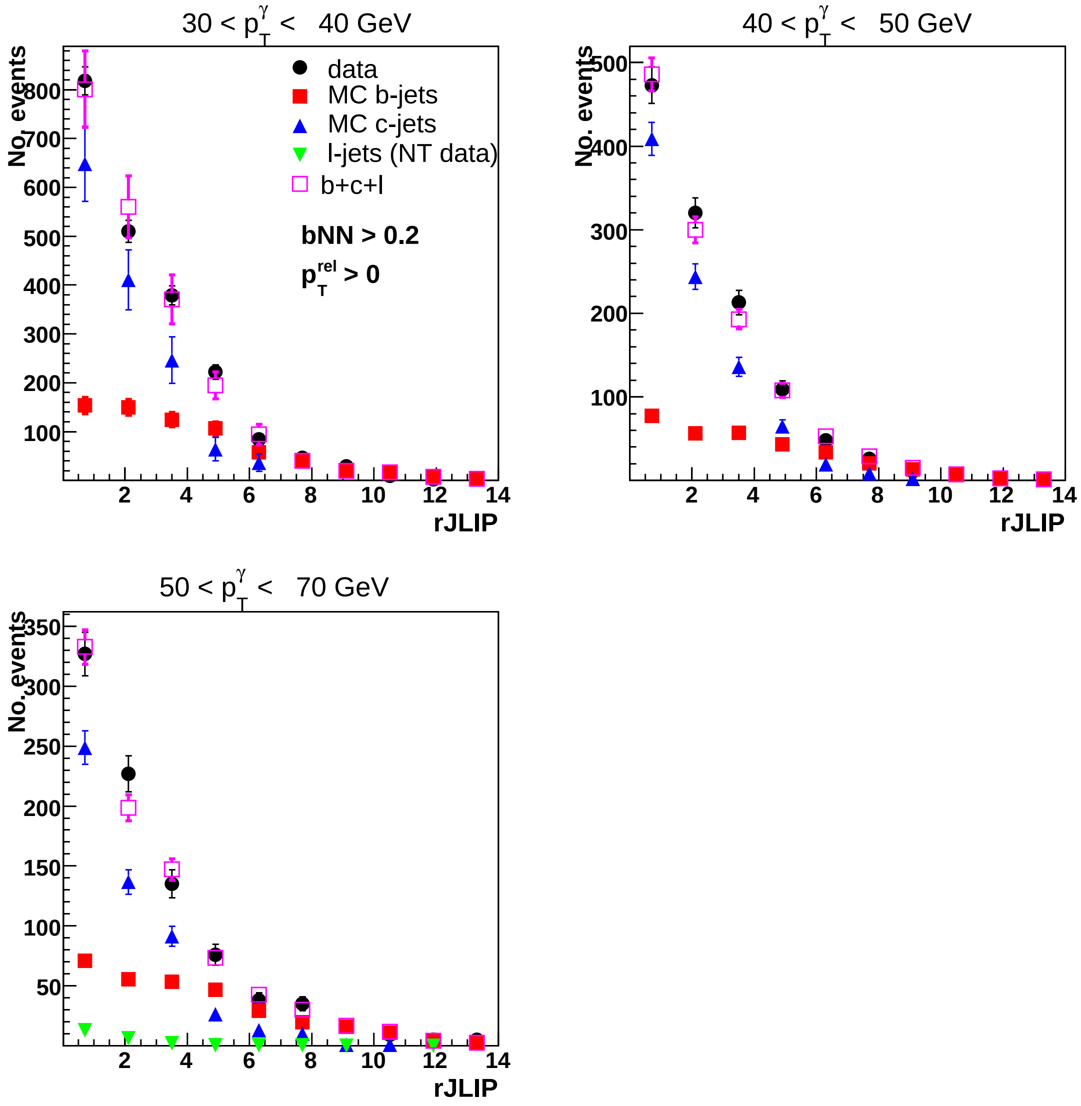


Figure B.5: Same as in Fig. B.3 but now with the requirement that a muon be found in the jet in all samples.

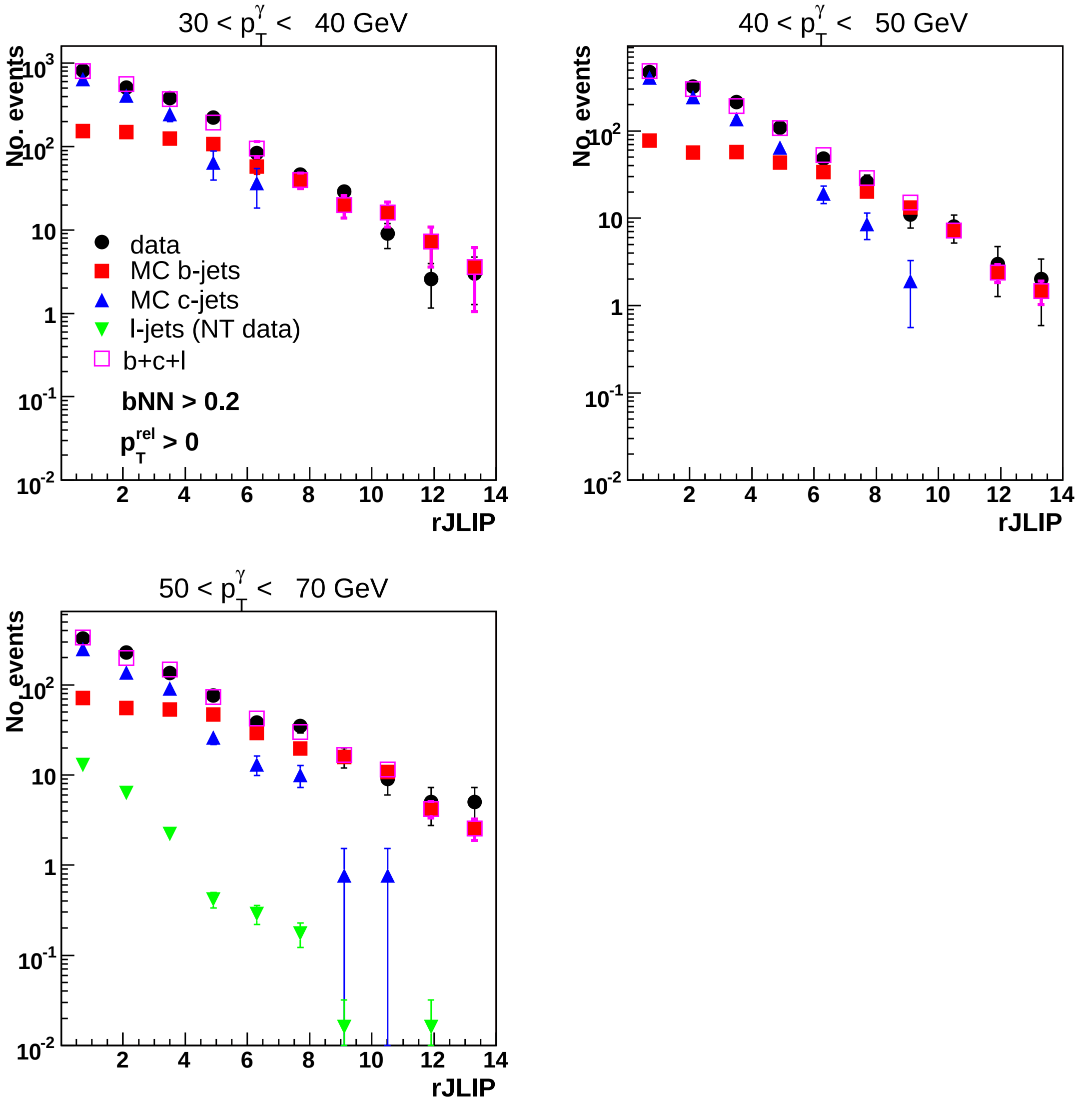


Figure B.6: Same as in Fig. B.5 but with a $\log_{10}$ y-axis scale.

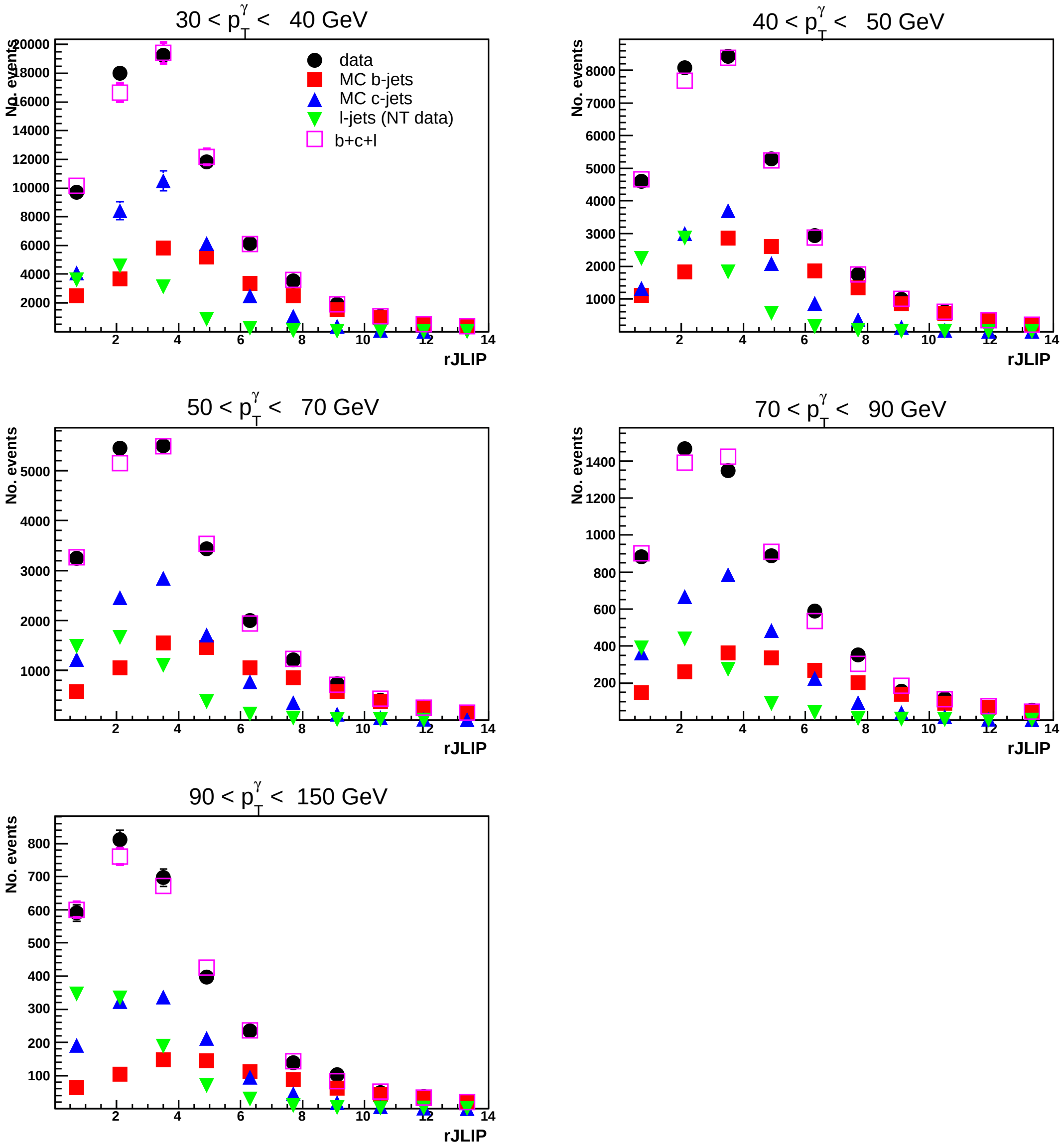


Figure B.7: rJLIP distribution for data all selection criteria have been imposed except photon identification requirements, including the bNN > 0.85 cut. The distributions of simulated b and c jets, and light jets from data are also shown, weighted according to their found flavor fractions. The sum of their distributions are normalized to the data.

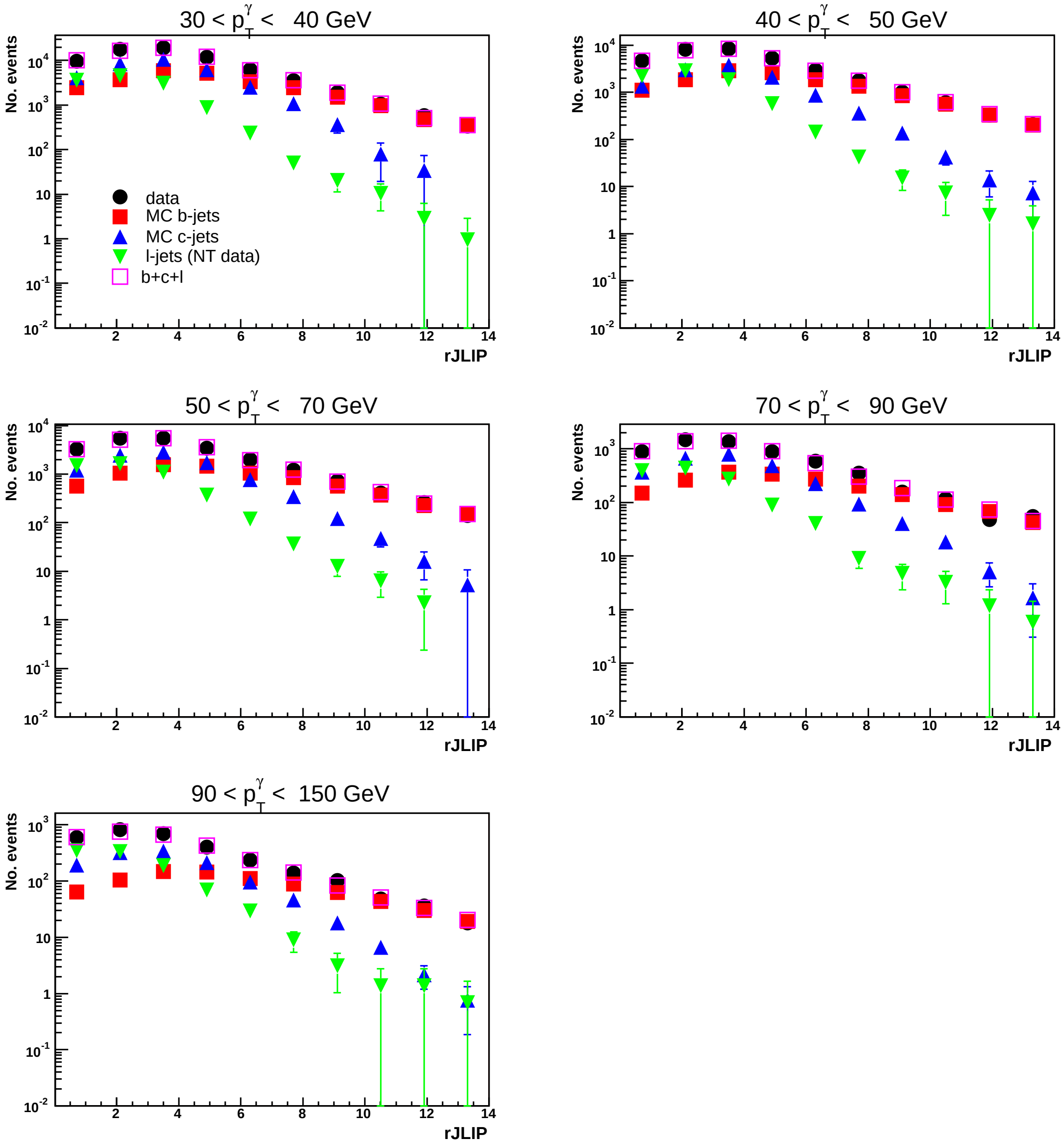


Figure B.8: Same as in Fig. B.7 but with a $\log_{10}$ y-axis scale.

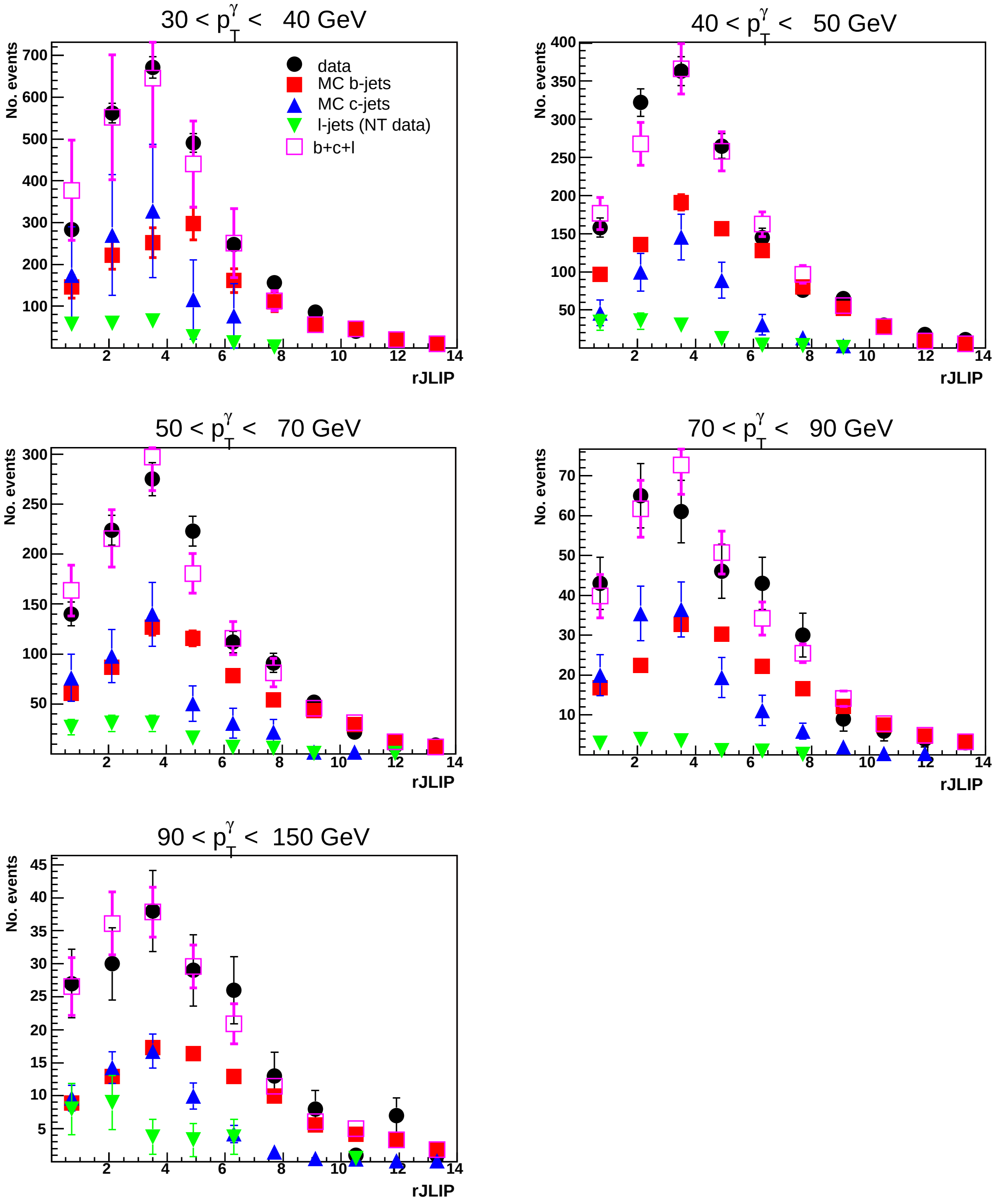


Figure B.9: Same as in Fig. B.7 but now with the requirement that a muon be found in the jet for all samples.

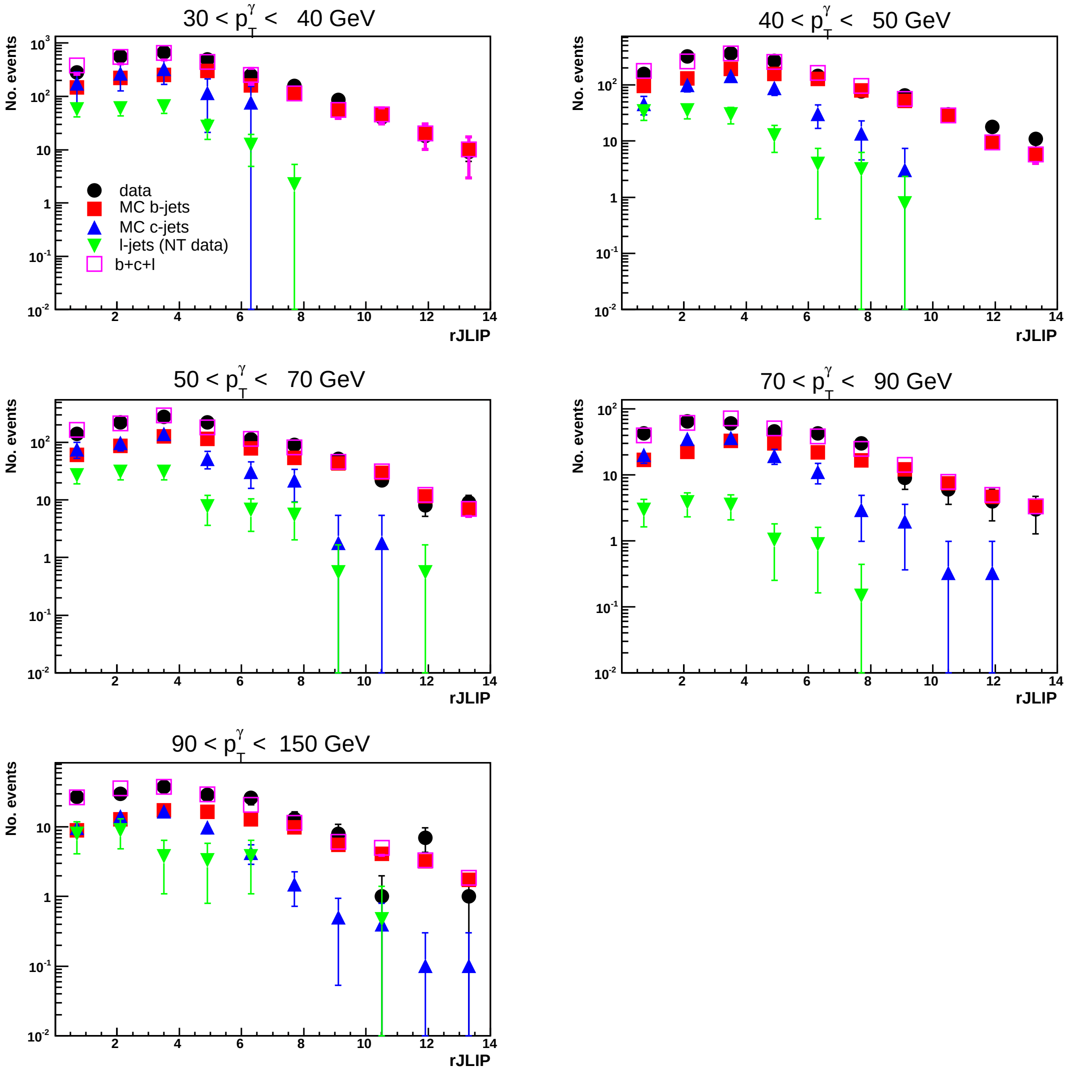


Figure B.10: Same as in Fig. B.9 but with a $\log_{10}$ y-axis scale.

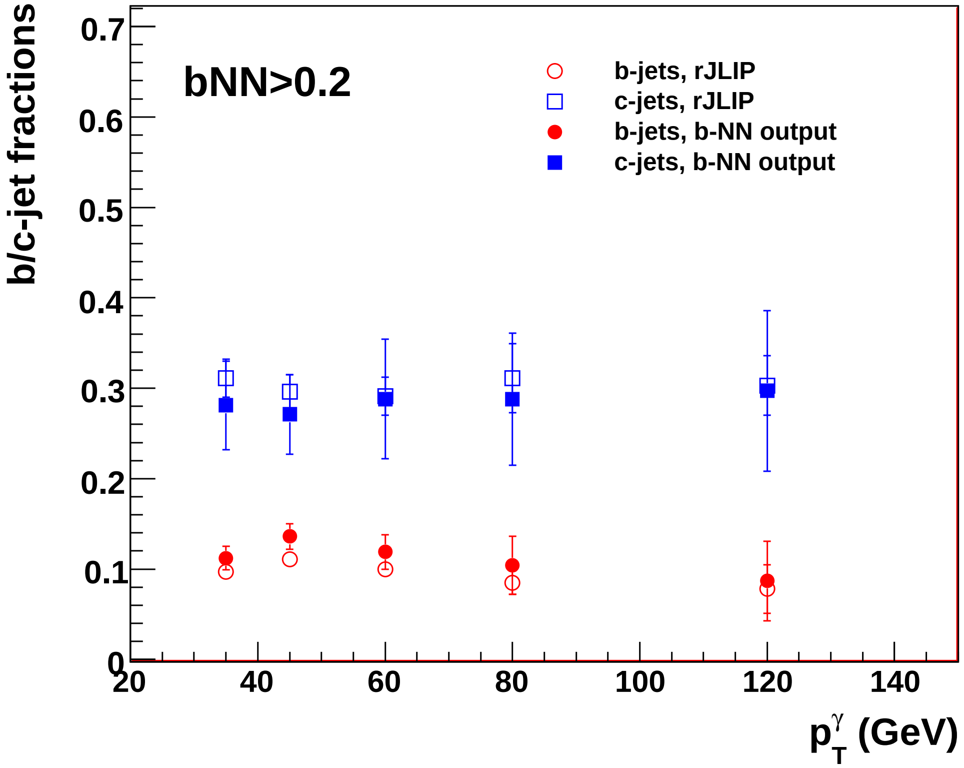


Figure B.11: Fraction of b and c jets in the data sample preselected with bNN $>$ 0.20. The open symbols correspond to the flavor fractions found using the rJLIP template fitting technique and the filled ones correspond to the efficiency method using the bNN.

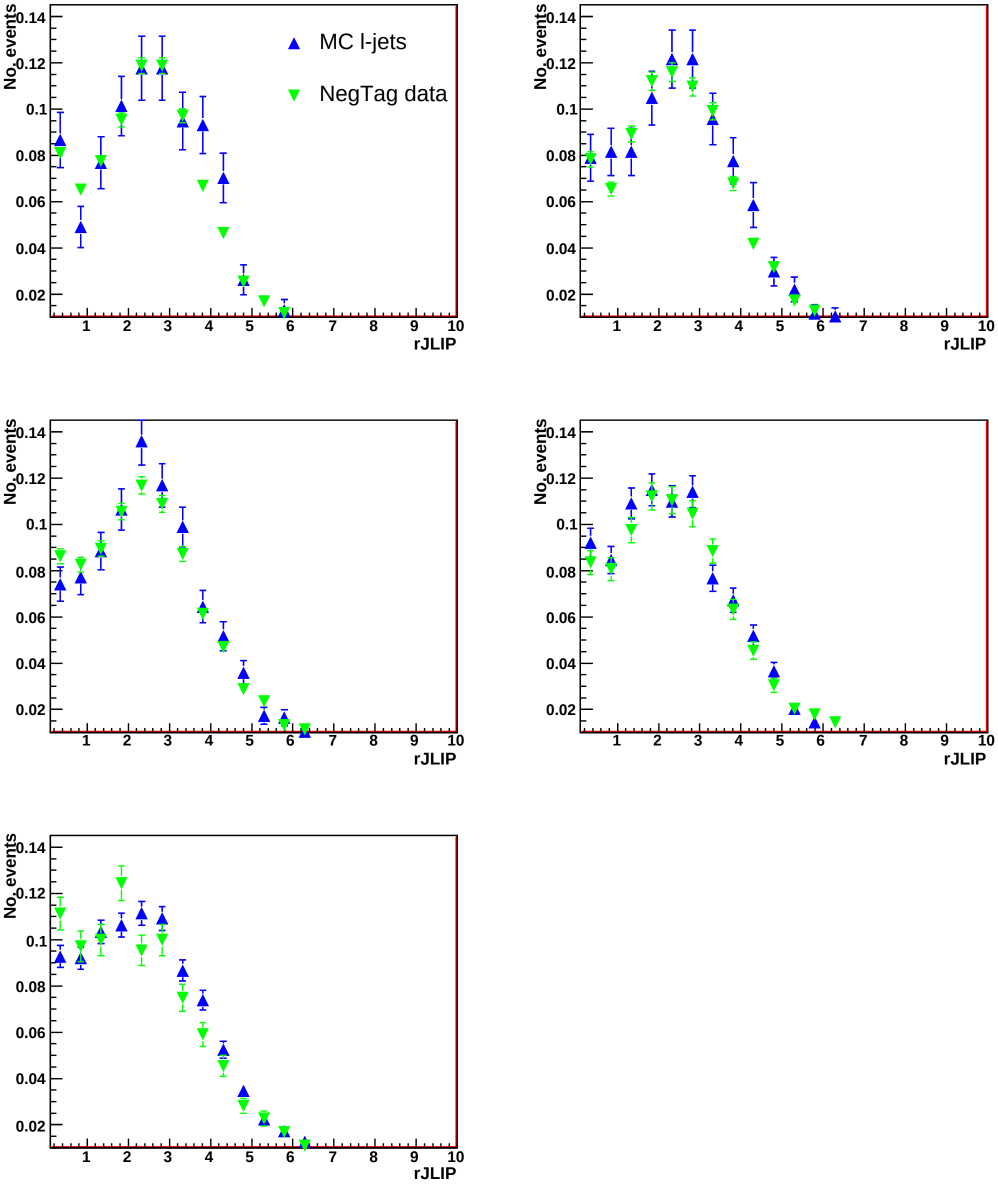


Figure B.12: Normalized distributions of the (negatively tagged) light jets from data and MC in the five p_T^γ bins satisfying the bNN>0.85 requirement.

APPENDIX C

JET P$_T$ THRESHOLD COMPARISONS

Because the "unsmearing" procedure is only applied to the photon's energy, the size of effects due to the jet's p_T resolution had to be estimated. The main effect is due to differences between the true particle's p_T versus its reconstructed p_T near the 15 GeV threshold. Simulated jets were used to study these effects, specifically comparisons were done using the simulated jets before being reconstructed ('particle' jets) and after being reconstructed ('reconstructed' jets). One study that was performed compared the particle jet efficiency to the reconstructed jet efficiency to pass the 15 GeV p_T threshold. The agreement is between the two efficiencies is found to be within $\sim 1\%$ for all p_T^γ bins, as shown in Fig. C.1.

Additionally, potential effects due to jet migrations near the jet p_T threshold were investigated as shown in Fig. C.2. First, the fraction of events where the particle jet p_T is above the 15 GeV threshold and the reconstructed jet p_T is below 15 GeV were investigated. This effect is less than 0.5% for all p_T^γ. From the jet p_T distribution in Fig. C.1, this is understandable because of the larger tail at low jet p_T. The second plot shows the fraction of events where the reconstructed jet p_T is above 15 GeV and the particle jet p_T is below 15 GeV. As to be expected, this is a larger effect than the first, again due to the larger tail at low jet p_T for the particle jet.

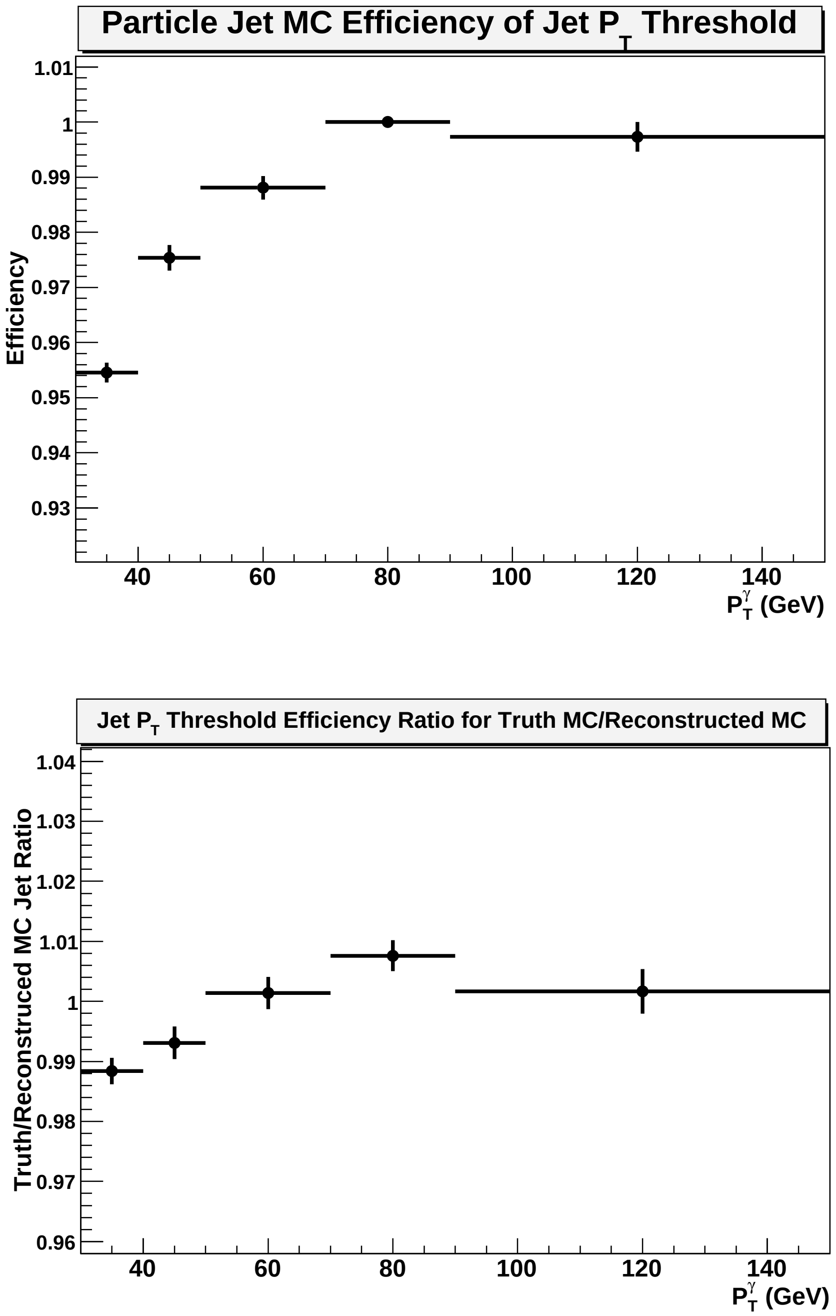


Figure C.1: The efficiency of a particle jet to pass the jet p_T cut of 15 GeV is shown in the top plot. In the bottom plot, the ratio of this efficiency to efficiency of the reconstructed Monte Carlo jet to pass the same cut is shown. The largest difference between the two efficiencies is found to $\sim 1.2\%$. This discrepancy is caused by a small difference in the shape of the particle and reconstructed jet p_T distributions as displayed in Fig. C.2.

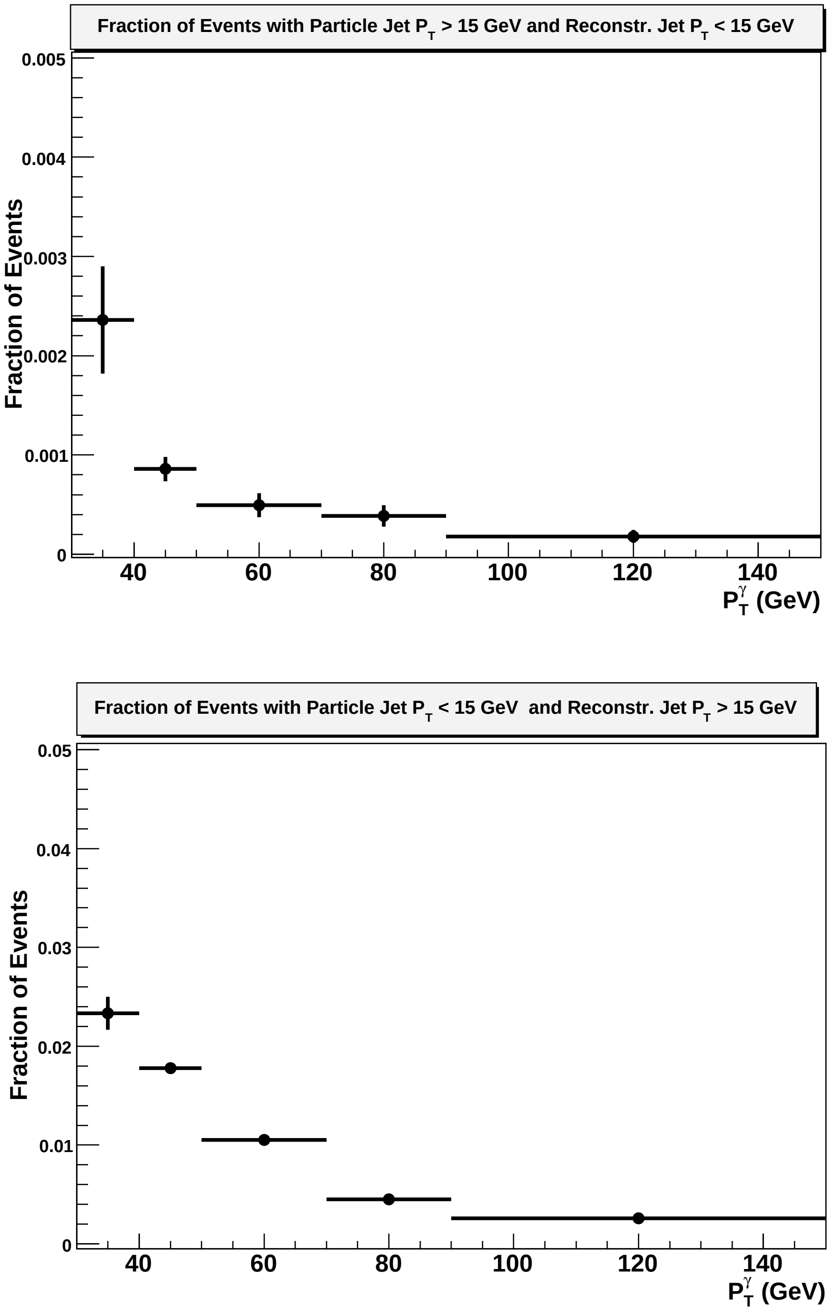


Figure C.2: The fraction of events in Monte Carlo is shown for when the particle jet p_T is above the jet p_T threshold and the reconstructed jet p_T is below it (top) and when the particle jet p_T is below the threshold and the reconstructed jet p_T is above (bottom).

APPENDIX D

NEURAL NETWORKS

Neural networks provide the ability to perform multidimensional cuts, reducing background events further by incorporating correlations among multiple discriminants. In essence, the neural network's main advantage is that it can be trained to recognize patterns among identification variables. The process [61] involves a reduction of dimensionality, such that the number of inputs is greater than the number of outputs. Two classes are defined, signal and background. The neural network is trained such that its output for signal events peaks at one, while that for background yields an output of zero. The neural network learns by building and subsequently testing relationships and connections between inputs. This specific model is a "Feed Forward" one, which means that there is a one directional flow of information from the inputs to the output.

The mathematical model of the neural network is that of a neuron firing in a biological system. All information coming from the inputs is summed at each point where they connect, called a node. If this sum is greater than the threshold set for that node, the neuron fires. After firing, the neuron resets to its initial state and sends a signal indicating its current state to its neighboring nodes. This model is called an elementary perceptron, and can be written in the form as shown in Equation D.1.

$$O(x_1, ..., x_n) = g(\beta \sum_i \omega_i x_i + \theta) \quad \text{where } g(t) = \frac{1}{1 + e^{-2t}} \tag{D.1}$$

In this equation, $(x_1, ..., x_n)$ represent the input nodes, O is the output, θ is the firing threshold, g acts as a non-linear signed transfer function, β defines the slope of g, and ω_i represents a weight. The weights are assigned to every node-node relationship and the values of the weights are determined from signal and background training. These weights are the physical embodiment of the correlations between inputs that boosts the discrimination

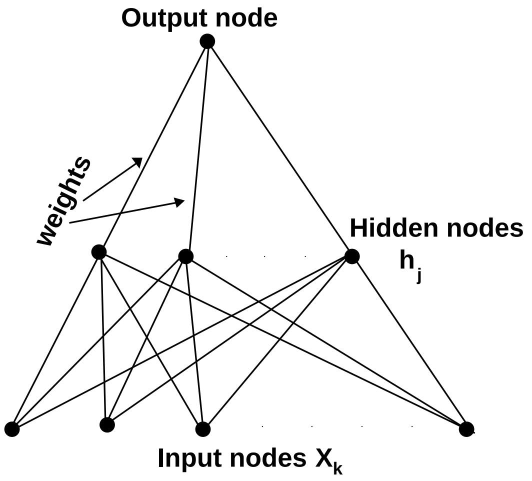


Figure D.1: Schematic of the "Feed Forward" style of neural network, which contains a layer of input nodes, hidden nodes, and output nodes. These layers are all connected by individual weights.

between signal and background. For the specific case of the "Feed Forward" neural network. weights connect the input nodes to a layer of hidden nodes (h_j), and then connects the hidden nodes to the output node as shown in Fig. D.1.

The process to determine the optimal set of weights to separate signal from background events is called "training". Training depends on the number of independent weights and thresholds used in the neural network as seen in Equation D.2.

$$N_{ind} = (N_{in} + N_{on}) \cdot N_{hn} + N_{ht} + N_{ot} \tag{D.2}$$

In this equation, the total number of independent weights and thresholds (N_{ind}) is a function of the number of input nodes (N_{in}), output nodes (N_{on}), hidden nodes (N_{hn}), hidden thresholds (N_{ht}), and output thresholds (N_{ot}). The training procedure is performed in two parts. The first part is to provide "training" data, the input patterns to the perceptron. This involves setting the total number of input nodes, hidden nodes, output nodes and thresholds to the desired amount, and providing enough data such that the number of events provided ($N_{events} \gg N_{ind}$). The second part is to set the desired target output of each pattern, which for the most simplest case of the "Feed Forward" style, corresponds to a signal output of one and a background output of zero. Now that the parameters have been set for the training procedure, the actual process can begin. Signal events and background events are listed separately, and for each event, its output is compared to the training target output. From this comparison, an error is estimated using the explicit calculation shown in Equation D.3, where $t^{(p)}$ is the target output for either a given signal or background event (0 or 1 in this case), and $O^{(p)}$ is the output calculated from the neural network using a given set of weights. By minimizing the error, the best set of weights is determined.

$$Error = \frac{1}{2N_{events}} \sum_{p=1}^{N_{events}} (O^{(p)} - t^{(p)})^2 \tag{D.3}$$

Once these weights are found, the neural network is complete. It can then be used as a basic function whose variables are the inputs of the neural network, and whose output indicates how "signal-like" or "background-like" any given event is.

REFERENCES

[1] C. Amsler *et al.,* Phys. Lett. **B592**, 1 (2008).

[2] P. Aurenche, J. Lindfors, Nucl. Phys. **B168**, 296 (1980).

[3] E.N. Argyres, A.P. Contogouris, N. Mebarki and S. D.P. Vlassopulos, Phys. Rev. D **35**, 1584 (1987).

[4] J.F. Owens, Rev. Mod. Phys. **59**, 465 (1987).

[5] P. Aurenche, R. Baier, M. Fontannaz, J.F. Owens and M. Werlen, Phys. Rev. D **39**, 3275 (1989).

[6] E.L. Berger and J. Qiu, Phys. Rev. D **44**, 2002 (1991).

[7] W. Vogelsang and A. Vogt, Nucl. Phys. **B453**, 334 (1995).

[8] S. Frixione and W. Vogelsang, CERN-TH/99-247, hep-ph/9908387.

[9] S. Catani, M. Fontannaz, J.Ph. Guillet and E. Pilon, JHEP **205**, 28 (2002).

[10] B. Bailey, E.L. Berger, L.E. Gordon, Phys. Rev. D **54**, 1896-1907 (1996). E.L. Berger, L.E. Gordon, Phys. Rev. D **54**, 2279-2294 (1996).

[11] M. Gluck, *et al.*, Phys. Lett. **B664** 133-138 (2008).

[12] V. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **85**, 5068 (2000). See also references inside.

[13] CDF Collaboration, *Preliminary*, `http://www-cdf.fnal.gov/physics/new/qcd/QCD.html`.

[14] V. Abazov *et al.* (DØ Collaboration), Phys.Rev.Lett. **94**, 161801 (2005).

[15] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **74**, 032008 (2006).

[16] A. Alton *et al.*, Phys. Rev. D **64**, 012002 (2001).

[17] J.J. Aubert *et al.*, Z. Phys. C **72**, 593 (1996).

[18] W.K. Tung, H.L. Lai, A. Belyaev, J. Pumplin, D. Stump and C.P. Yuan, JHEP **0702**, 053 (2007).

[19] C. Amsler, Phys. Lett. B **667**, 1 (2008).

[20] V. Abazov *et al.* (DØ Collaboration), Phys. Lett. B **666**, 435 (2008).

[21] V. Abazov *et al.* (DØ Collaboration), Phys. Lett. **B639**, 151-158 (2006).

[22] J.C. Collins, D.E. Soper, G. Sterman, Phys. Rev. D **54**, 2279 (1996).

[23] V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
G. Altarelli, G. Parisi, Nucl. Phys. **B126**, 298 (1977).
Yu. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).

[24] T. Stavreva, J.F. Owens, arXiv:0901.3791v1 (2009). Accepted by Phys. Rev. D.

[25] B.W. Harris, J.F. Owens, Phys. Rev. D **65**, 094032 (2002).

[26] P.M. Nadolsky *et al.*, Phys. Rev. D **78**, 013004 (2008).

[27] J. Pumplin *et al.*, Phys. Rev. D **75** 054029 (2007).

[28] V. Abazov *et al* (DØ Collaboration), Nucl. Instrum. Methods A **565**, 463 (2006).

[29] M. Abolins *et al.*, Nucl. Instrum. Methods A **584**, 75-97 (2008).

[30] M. Johnson *et al.*, Nucl. Instrum. Methods A **579**, 718 (2007).

[31] S. Abachi *et al.* (DØ Collaboration), Nucl. Instrum. Methods A **338**, 185 (1994).

[32] T. Andeen *et al.*, FERMILAB-TM-2365 (2007).

[33] G. Borissov, "Ordering a Chaos or... Technical Details of AA Tracking", `http://www-d0.fnal.gov/global_tracking/talks/20030228/talk-adm-030228.ps`.

[34] A. Garcia-Bellido, S. Lager, F. Rizatdinova, A. Schwartzman, G. Watts, "Primary Vertex certification in p14", DØ Note 4320.

[35] F. Fleuret, "The DØ Electron/Photon Analysis Package EMAnalyze", DØ Note 3888. `http://www-d0.fnal.gov/phys_id/emid/d0_private/emid.html`.

[36] L. Duflot, M. Ridel, "The CellNN algorithm: cell level clustering in the DØ calorimeter", DØ Note 3923.

[37] D. Bandurin, O. Atramentov, Y. Maravin, "Photon Energy Scale For Jet Energy Scaling", DØ Note 4974.

[38] G.C. Blazey *et al.*, arXiv:hep-ex/0005012, (2000).

[39] "Jet Energy Scale Determination at DØ Run II", DØ Note 5382.

[40] A. Schwartzman, M. Narain, "Secondary Vertex Reconstruction using the Kalman Filter", DØ Note 3908.

[41] A. Schwartzman, M. Narain, “b-quark Jet Identification via Secondary Vertex Reconstruction”, DØ Note 4080.

[42] R. Demina, A. Khanov, F. Rizatdinova, “b-tagging with Counting Signed Impact Parameter Method”, DØ Note 4049.

[43] T. Scanlon, FERMILAB-THESIS-2006-43.

[44] D. Bloch, *et al*, “Update of the JLIP b-tagger Performance in p14/pass2 with JES 5.3” DØ Note 4824,
D. Bloch, *et al*, “Performance of the JLIP b-tagger in p14” DØ Note 4348.

[45] T. Scanlon, “Improvements to the b-tagging Performance Measurement in p17 Data”, DØ Note 5402.

[46] T. Sjöstrand *et al.*, Comput. Phys. Commun. **135** 238 (2001).

[47] R. Brun, F. Carminati, CERN Program Library Long Writeup **W5013**, (1993), unpublished.

[48] D. Bandurin, G. Golovanov, D. Korablev, N. Skachkov, DØ Note 5368.

[49] D. Bandurin, M. Begel, J. Cammin, D. Duggan, DØ Note 5702.

[50] J. Hays, V. Kaushik, J. Mitrevski, O. Mundal, C. Schwanenberger, “Electron Trigger Efficiencies using Calorimeter Information in p17 Data”, DØ Note 5138.

[51] E. Carrera, “Search for Large Extra Dimensions in the Monophoton Final State”, DØ Note 5617.

[52] O. Atramentov, A. Askew, D. Bandurin, D. Duggan, A. Ferapontov, Y. Gershtein, Y. Maravin and G. Pawloski “Photon Identification in p17 data”, DØ Note 4976.

[53] D. Bandurin, “Artificial neural network for photon selection.”, DØ Note 5245.

[54] A. Harel, “Jet ID Optimization”, DØ Note 4919.

[55] C. Gerber, E. Shabalina, G. Otero y Garzon, “Taggability in Pass2 p14 Data”, DØ Note 4995;
I. Anghel, C. Gerber, L. Shabalina, T. Ten, “Studies of Taggability vs. nPV for p17”, DØ Note 5240.

[56] R. Barlow, C. Beeston, Comp.Phys.Comm. **77** (1993)219-228. See also p.120-131 of HBOOK manual.

[57] S. Jain, “Scale and oversmearing for high-p_T electrons”, DØ Note 4402.

[58] F. Abachi *et al.* (DØ Collaboration), Phys. Rev. Lett. **77**, 5011 (1996);
B. Abbott *et al.* (DØ Collaboration), Phys. Rev. Lett. **84**, 2786 (2000);
V. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **87**, 251805 (2001).

[59] N. Makovec, J.F. Grivaz, "Shifting, Smearing and Removing Simulated Jets", DØ Note 4914.

[60] Hwidong Yoo, Richard Partridge, DØ Note 5557.

[61] C. Peterson, T. Rognvaldsson and L. Lonnblad, "JETNET 3.0. A versatile Artificial Neural Network Package", Lund University Preprint LU-TP 93-29. Version 3.5 is used.

BIOGRAPHICAL SKETCH

Daniel J. Duggan

Daniel Duggan was born in Omaha, Nebraska on April 12th, 1980. As the youngest of five children, he quickly learned the ways of Nebraskans, decency, humor, and science. He graduated high school as valedictorian from Mount Michael Benedictine in 1998, and in 2002 earned degrees from Indiana University in Physics, Mathematics, and French. In Fall of 2003, he began his graduate studies at Florida State University under the supervision of Professor Horst Wahl, which led to research in Experimental High Energy Particle Physics at the Fermilab Tevatron. After receiving his Masters of Physics degree from Florida State, he moved to Chicago, Illinois to continue his research in earnest while helping to maintain the DØ experiment's calorimeter. In Spring of 2009, he successfully defended his dissertation and plans to continue his research in particle physics for the CMS collaboration at the Large Hadron Collider.