Study of Quark Substructure Using Proton-Proton Collisions

by

Daniel Miner

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Regina Demina

Department of Physics and Astronomy

Arts, Sciences and Engineering

School of Arts and Sciences

University of Rochester

Rochester, New York

2011

Curriculum Vitae

The author was born on September 17, 1982 in Philadelphia, Pennsylvania. He attended Brandeis University from 2001 to 2005, where he received a Bachelor of Science degree in physics with a minor in mathematics. He subsequently attended the University of Rochester for his graduate studies beginning in 2005. Having completed the Master of Arts portion of his studies, he moved to Geneva, Switzerland in 2007 to pursue his doctoral research on the Large Hadron Collider's CMS experiment at CERN. In 2010 he moved back to Rochester to write his dissertation on the research he'd done at CERN.

List of Publications:

The CMS Collaboration. Search for Quark Compositeness with the Dijet Centrality Ratio in pp Collisions at $\sqrt{s} = 7$ Tev. Phys. Rev. Lett., 105(26), December 2010.

Acknowledgements

Firstly, I'd like to express my gratitude to my advisor, Regina Demina. Her support, advice and guidance have been invaluable. I'd especially like to thank her for her continued support through my recovery from an injury that temporarily prevented me from working and threatened to throw me off track. I would like to thank the Rochester CMS group, particularly Pawel De Barbaro, Amnon Harel and Marek Zielinski. I would like to thank the Fermi National Accelerator Laboratory LPC JetMet group, particularly Robert Harris. I would like to thank Jim Hirschauer, whose leadership of the dijet centrality ratio group during my recovery from the aforementioned injury and whose subsequent contributions were vital to our early success and the rapid publication of our results. I am grateful to the Department of Energy for providing the grants which enabled my continued financial support and travel to and from CERN. I would like to thank my close friend Kristina Wilson, whose support during my recovery and subsequent close friendship helped to provide me with the necessary motivation to continue forward in an ambitious fashion. Finally, I'd like to thank my brother, Gabriel, my grandparents, Joe and Shirley Shrager, and my parents, David and Debbie.

Abstract

Two or more jets are frequently produced as a result of proton-proton collisions. The angular distribution of these jets can provide a wealth of information about the underlying interactions taking place as the protons collide. In this study, we use a simple observable called the dijet centrality ratio to examine this angular distribution, and search a 35.7 pb^{-1} sample of data at $\sqrt{s} = 7$ TeV from CERN's Large Hadron Collider for evidence of quark substructure. We expect to set a 95% confidence limit on the energy scale of such interactions of $\Lambda < 4.9$ TeV and are in actuality able to set a limit of $\Lambda < 4.6$ TeV.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

- SM standard model

- QCD quantum chromodynamics

- QED quantum electrodynamics

- LO leading order

- NLO next to leading order

- GeV giga electron volts

- TeV tera electron volts

- LHC large hadron collider

- CMS compact muon solenoid

- ECAL electromagnetic calorimeter

- HCAL hadronic calorimeter

- PS parton showering

- ISR initial state radiation

- FSR final state radiation

- MPI multiple parton interaction

- HAD hadronization

- LLR log likelihood ratio

- CL_S signal confidence limit

- θ^* scattering angle

- η pseudorapidity

- ϕ azimuthal angle

- p_T transverse momentum

- E_T transverse energy

- $\not{E}_T$ missing transverse energy

- Λ contact interaction energy scale

- R_η dijet centrality ratio

Foreword

The author served as one of the leaders of the Dijet Centrality Ratio group on the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The CMS collaboration consists of over 3,000 scientists responsible for the construction, maintenance, and operation of the CMS detector at the LHC. The Dijet Centrality Ratio group (and, by extension, the CMS collaboration as a whole) published an early version of these results in Physical Review Letters [7].

Figures 10-17, 21-25 from Reference [4].

Figures 18-20 from Reference [5].

Figures 26-27 from Reference [2].

Figures 28-37 produced internally for the CMS collaboration by the Dijet Resonance Search and Dijet Centrality Ratio groups.

Figures 40-43, 45-54, 61-62 produced internally for the CMS collaboration by the Dijet Centrality Ratio group.

Specifically: From the Summer of 2007 until the first turn-on of the LHC in the Fall of 2008, the author participated in the hardware commissioning of the CMS detector HCAL. This included the installation of the LVTMM (low voltage temperature monitoring and measurement) system, cabling, and operation of the detector in diagnostic and cosmic ray triggered operation. HCAL operational activity and growth of expertise continued through the Fall of 2009. Simultaneously, beginning in the Fall of 2008, the author began research into the analysis on which this thesis is written. This was done with significant assistance from the JetMet (jets and missing transverse energy) group. In

the Summer of 2009, as it became apparent that statistical techniques which lay beyond the author's scope at the time were necessary, a senior postdoc, Amnon Harel, was brought on and became a major contributor to the study. In the Fall of 2009, immediately following the release of a preliminary internal analysis note overseen by the author, the author suffered significant traumatic injuries in a motor vehicle accident which left him unable to work effectively until the Spring of 2010. During this approximately six month period, Jim Hirschauer, a postdoc, joined Amnon and replaced the author as the spearheading figure on the study. Jason St John, a grad student, also contributed. Their progress and contributions were instrumental in setting up the study to be the second major result published by the CMS collaboration in the Winter of 2010, approximately six months after the author's working return to the collaboration. Between his return to work and the publication of the result, the author continued to work as an operational expert on the HCAL as well. In the Spring of 2011, the author presented an updated study result at the April meeting of the American Physical Society utilizing the entire 2010 data set, which contained approximately ten times the data used to generate the original published result. This thesis focuses on that updated result.

1 Introduction

The reductionist view of the universe which forms the philosophical basis of particle physics dates at least to ancient Greece, with the philosophy of atomism positing that all material in the universe is composed of a family of indivisible, fundamental particles. It is only in relatively recent history that this line of reasoning has been brought out of the realm of philosophy and explored experimentally. The late nineteenth and early twentieth century provided the underpinnings of this, with Thomson's discovery of the electron and Rutherford's discovery of the atomic nucleus. As simultaneous work on relativity and early quantum physics began to provide a framework for the theoretical models that would support these findings, further experimental findings (such as Chadwick's discovery of the neutron in the 1930s) pushed the empirical envelope as well. Throughout the twentieth century, the development of particle accelerators, first fixed-target and then colliding-beam (and their respective particle imaging and detection instruments) allowed the probing of first the structure of the atom, then its nucleus, and then the protons and neutrons which compose the atomic nucleus. As these developments moved forward, the Standard Model, the theoretical basis of the current status quo in particle physic, was developed in the 1970s [21].

In the Standard Model of particle physics, quarks are the fundamental heavy fermions and are assumed to be indivisible [21]. The Standard Model, however, has certain shortcomings, including but not limited to the problems of fine tuning [21] and the absence of gravity [21]. Certain extensions proposed to remedy said shortcomings suggest that quarks are in fact composite particles.

Given the unprecedented energies attainable at CERN's Large Hadron Collider (LHC) (see section 3.1) [20], we have been able to use the Compact Muon Solenoid (CMS) experiment, a general purpose particle detector (see section 3.2) [4], to probe the structure of quarks to an unrivaled degree [7].

The behavior of quarks is governed primarily by quantum chromodynamics (QCD). QCD processes lead to the production of jets in proton-proton collisions. One of the phenomenological consequences of QCD is color confinement, which is supported by extensive experimental evidence and postulates that free quarks and gluons cannot exist [21]. In the case of energetic free quarks, color confinement leads to hadronization, which is the process by which said energetic free quarks use some of their energy to "pull" additional quarks and gluons out of the vacuum and recombine to form hadrons (and, to a lesser degree, mesons). A jet is a tightly focused stream of energetic particles produced by the hadronization of a quark or gluon collision product.

A collision event with two or more energetic jets is known as a dijet event. At high collision energies such as those found in the LHC, QCD dijets tend to be produced in a highly forward fashion - that is to say with a shallow angle with respect to the beamline, or a high pseudorapidity $\eta = -\ln\tan\frac{\theta^*}{2}$ (where θ^* is the angle of the jet with respect to the beam). Many new physics processes, on the other hand (including the contact interactions predicted by composite quark models) produce more central dijet distributions [1,15,16,26]. The dijet centrality ratio is defined as the number of dijet events in which both jets are in the central region of the barrel divided by the number of dijet events in which both jets are in the forward region of the barrel. This is, in turn,

studied as a function of dijet invariant mass. Because many systematics cancel out to the first order in the ratio, this observable provides a robust and effective test of QCD and a high degree of sensitivity to new physics [22]. The author's group within the CMS collaboration published the second physics result paper to come out of the experiment [7] with 2.9 pb^{-1} of integrated luminosity. The study presented in this thesis increased that number by more than an order of magnitude to approximately 36 pb^{-1}. A similar ratio was used to set a 95% confidence limit on the contact interaction energy scale of $\Lambda < 2.8$ TeV at the DZero experiment at the Tevatron [9, 10].

2   Theory

2.1   The Standard Model
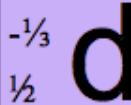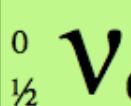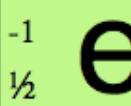
Three Generations of Matter (Fermions)

	I	II	III	
mass→	2.4 MeV	1.27 GeV	171.2 GeV	0
charge→	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0
spin→	$\frac{1}{2}$ **u**	$\frac{1}{2}$ **c**	$\frac{1}{2}$ **t**	1 **γ**
name→	up	charm	top	photon
	4.8 MeV	104 MeV	4.2 GeV	0
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0
	$\frac{1}{2}$ **d**	$\frac{1}{2}$ **s**	$\frac{1}{2}$ **b**	1 **g**
	down	strange	bottom	gluon
	<2.2 eV	<0.17 MeV	<15.5 MeV	91.2 GeV
	0	0	0	0
	$\frac{1}{2}$ ν_e	$\frac{1}{2}$ ν_μ	$\frac{1}{2}$ ν_τ	1 Z^0
	electron neutrino	muon neutrino	tau neutrino	weak force
	0.511 MeV	105.7 MeV	1.777 GeV	80.4 GeV
	-1	-1	-1	±1
	$\frac{1}{2}$ **e**	$\frac{1}{2}$ μ	$\frac{1}{2}$ τ	1 $W^\pm$
	electron	muon	tau	weak force

Quarks · Leptons · Bosons (Forces)

Figure 1: The particles of the standard model (confirmed by discovery) [Wikimedia Commons]

The Standard Model of particle physics is currently our best, most well-tested description of the fundamental behavior of the universe (gravity excepted). In the Standard Model, all matter is believed to be made up of fermionic quarks and leptons, while interactions are mediated by bosons. The fermionic particles are composed of three analogously-structured generations, as seen in figure 1. In addition, each particle has an antiparticle with opposite quantum numbers - electrical charge, for example. The quarks each have one of three different "color charges" (hence the term quantum "chromo"dynamics to describe strong nuclear interactions).

The Standard Model describes the electromagnetic, weak and strong nuclear interactions. The described forces are modeled as a non-Abelian gauge theory based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group, where C refers to the quark color charge, L refers to the left-handedness of the particles in the group, and Y refers to the hypercharge. The eight massless bosons (gluons) of the $SU(3)_C$ group mediate the strong nuclear interaction, while the electromagnetic and weak nuclear interactions are unified in the other two groups and mediated by the massive W^+, W^-, and Z bosons and the massless photon. Only particles with color charge (i.e. quarks) couple to the strong mediators (gluons). All electrically charged particles couple to the electromagnetic mediators (photons). Only the left-handed isospin doublets couple to the weak mediators (Ws and Zs).

Local gauge invariance would seem to dictate that the gauge bosons (such as the Ws and Zs) should be massless, but this is not the case. The mass generation mechanism is unverified, but the most popular theory is called the

Higgs mechanism, which introduces a new scalar doublet to the theory. This scalar doublet, called the Higgs field, spontaneously breaks the symmetry of the $SU(2)_L \times U(1)_Y$ group to the $U(1)_Q$ group, where Q is electrical charge. The Higgs field has four degrees of freedom, three of which give mass to the W and Z bosons, and the fourth of which generates a massive spin zero particle called the Higgs boson. Fermion masses are then generated by an ad-hoc Yukawa coupling to the Higgs boson, with the coupling strength proportional to the fermion mass. The discovery of a Higgs boson would confirm this theory.

The specific phenomenology involved in the described study arises primarily from quantum chromodynamics, which comes from the $SU(3)$ group with color charge taken as a local symmetry. In this group, three fundamental interactions are possible: a quark can radiate (or absorb) a gluon, a gluon can radiate (or absorb) a gluon, and a gluon can radiate (or absorb) a quark-antiquark pair. In order to calculate the likelihood of a specific interaction, the likelihoods of all possible routes to the same outgoing particles (stemming from the same incoming particles) are summed. In practice, this is often done as the (partial) sum of a power series expansion utilizing Feynman diagrams [21]. This is referred to as a perturbation series expansion, as the sum is an expansion or increasingly subtle perturbations "around" the set of simplest processes. The simplest (and often most likely) processes involved in dijet production possess two three-way interaction vertices and are referred to as leading order (LO) processes. The scattering diagrams shown in figure 2 are examples of leading order processes. Next to leading order (NLO) processes possess an additional vertex, typically in the form of a particle radiating an

extra particle immediately before or immediately after the primary interaction. These specific processes are referred to as initial state radiation (ISR) and final state radiation (FSR) respectively. Next to next to leading order (NNLO) processes possess two additional vertices (often forming closed loops) within the interaction diagram, and so on.
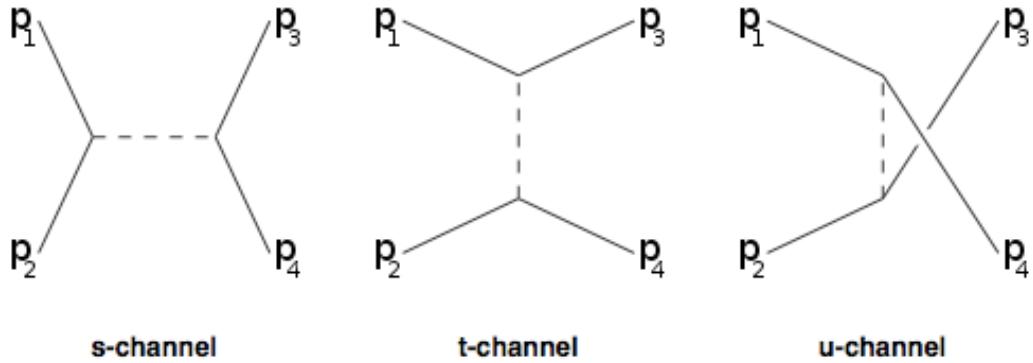
Figure 2: $\hat{t}$, $\hat{s}$ and $\hat{u}$ production channels [Wikimedia Commons]

2.2 Quark Contact Interactions

The particular form of new physics beyond the standard model to which the production of two or more jets (dijets) are most sensitive is a quark contact interaction (though it does maintain lesser but not-insignificant sensitivity to other new physics signals, such as resonant particle production [1, 15, 23, 26]). A contact interaction would be indicative of quark substructure, which the standard model does not include. There is great historical precedent for such reductionist thought - from the atom to the lepton and hadron to the parton. This reductionist chain of discoveries naturally leads one to question

and explore the hypothesized fundamental nature of the quark. Numerous and varied models exist, but the most general $SU(3)_C \times SU(2)_L \times U(1)_Y$-invariant light quark contact interaction is described by the following Lagrangian [15].

$$
\begin{aligned}
L_{qq} = \frac{g^2}{2\Lambda^2}(\quad & A_0 \overline{q}_L \gamma^\mu q_L \overline{q}_L \gamma_\mu q_L \\
& + A_1 \overline{q}_L \gamma^\mu \tfrac{\tau_a}{2} q_L \overline{q}_L \gamma_\mu \tfrac{\tau_a}{2} q_L \\
& + A_{0u} \overline{q}_L \gamma^\mu q_L \overline{u}_R \gamma_\mu u_R \\
& + A_{0d} \overline{q}_L \gamma^\mu q_L \overline{d}_R \gamma_\mu d_R \\
& + A_{8u} \overline{q}_L \gamma^\mu \tfrac{\lambda_b}{2} q_L \overline{u}_R \gamma_\mu \tfrac{\lambda_b}{2} u_R \\
& + A_{8d} \overline{q}_L \gamma^\mu \tfrac{\lambda_b}{2} q_L \overline{d}_R \gamma_\mu \tfrac{\lambda_b}{2} d_R \\
& + A_{uu} \overline{u}_R \gamma^\mu u_R \overline{u}_R \gamma_\mu u_R \\
& + A_{dd} \overline{d}_R \gamma^\mu d_R \overline{d}_R \gamma_\mu d_R \\
& + A_{ud} \overline{u}_R \gamma^\mu u_R \overline{d}_R \gamma_\mu d_R \\
& + A_{du} \overline{u}_R \gamma^\mu d_R \overline{d}_R \gamma_\mu u_R)
\end{aligned}
\tag{1}
$$

Here g defines the quark coupling strength, $q_L = (u_L, d_L)$ (the wavefunctions describing the left-handed up and down quark fields), λ_b ($b = (1, 2, ..., 8)$) are the $SU(3)_C$ color matrices, γ_μ are the Dirac Gamma matrices, τ_a ($a = (1, 2, 3)$) are the weak isospin Pauli matrices and A_{ij} define the sign and strength of the Lagrangian terms. We define Λ as the contact interaction energy scale (which goes to the Standard Model as it approaches infinity) and by convention set the square of the coupling g^2 to 4π and the largest possible

$|A_{ij}| = 1$. This Lagrangian can be extended in similar terms to cover the entire six-quark field. For the high energies achieved at the LHC we concern ourself with the dominant term in this extension, defining q_L to now be the entire quark field instead of just the light quark field [17]:

$$L_{qq} = \frac{A_0 g^2}{2\Lambda^2} \bar{q}_L \gamma^\mu q_L \bar{q}_L \gamma_\mu q_L \tag{2}$$

This term arises from a left-left isoscalar color singlet interaction [16]. This is added to the existing QCD Lagrangian for jet production. We take $A_0 = +1$ here, as this will later provide the most conservative coverage. From this the lowest order quark scattering angular cross-sections (area measures used to quantify interaction likelihoods in differential form) can be calculated to be:
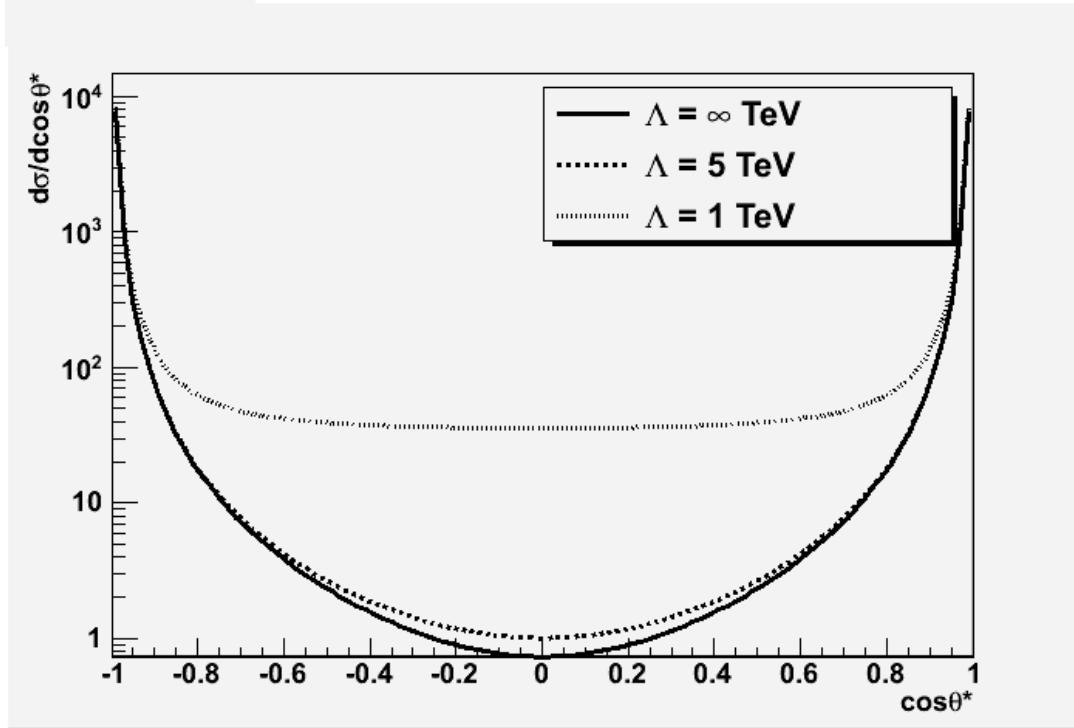
Figure 3: Angular cross-sections for $q_i q_i \rightarrow q_i q_i$

$$\frac{d\hat{\sigma}(q_i q_i \rightarrow q_i q_i)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}\left(\frac{4}{9}\alpha_S^2\left(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{s}^2}{\hat{u}^2} - \frac{2}{3}\frac{\hat{s}^2}{\hat{t}\hat{u}}\right) + \frac{8}{9}\frac{A_0\alpha_S}{\Lambda^2}\left(\frac{\hat{s}^2}{\hat{t}} + \frac{\hat{s}^2}{\hat{u}}\right) + \frac{8}{3}\frac{\hat{s}^2}{\Lambda^4}\right)$$

(3)
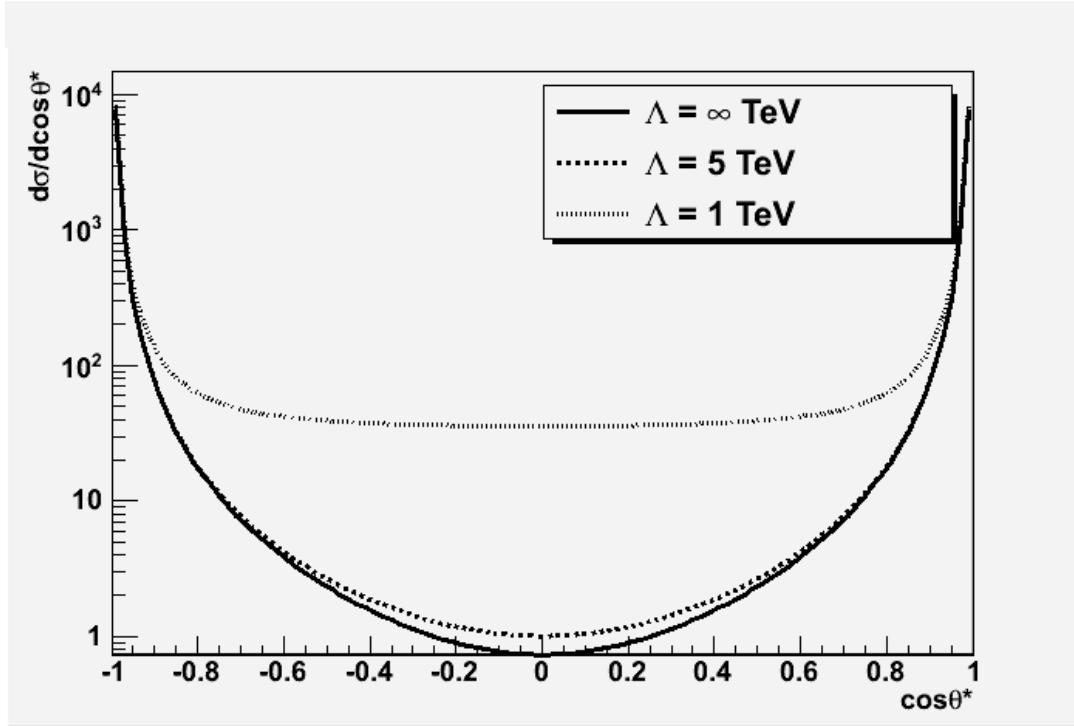
Figure 4: Angular cross-sections for $\overline{q}_i\overline{q}_i \rightarrow \overline{q}_i\overline{q}_i$

$$\frac{d\hat{\sigma}(\overline{q}_i\overline{q}_i \rightarrow \overline{q}_i\overline{q}_i)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}\left(\frac{4}{9}\alpha_S^2\left(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{s}^2}{\hat{u}^2} - \frac{2}{3}\frac{\hat{s}^2}{\hat{t}\hat{u}}\right) + \frac{8}{9}\frac{A_0\alpha_S}{\Lambda^2}\left(\frac{\hat{s}^2}{\hat{t}} + \frac{\hat{s}^2}{\hat{u}}\right) + \frac{8}{3}\frac{\hat{s}^2}{\Lambda^4}\right)$$

$$(4)$$
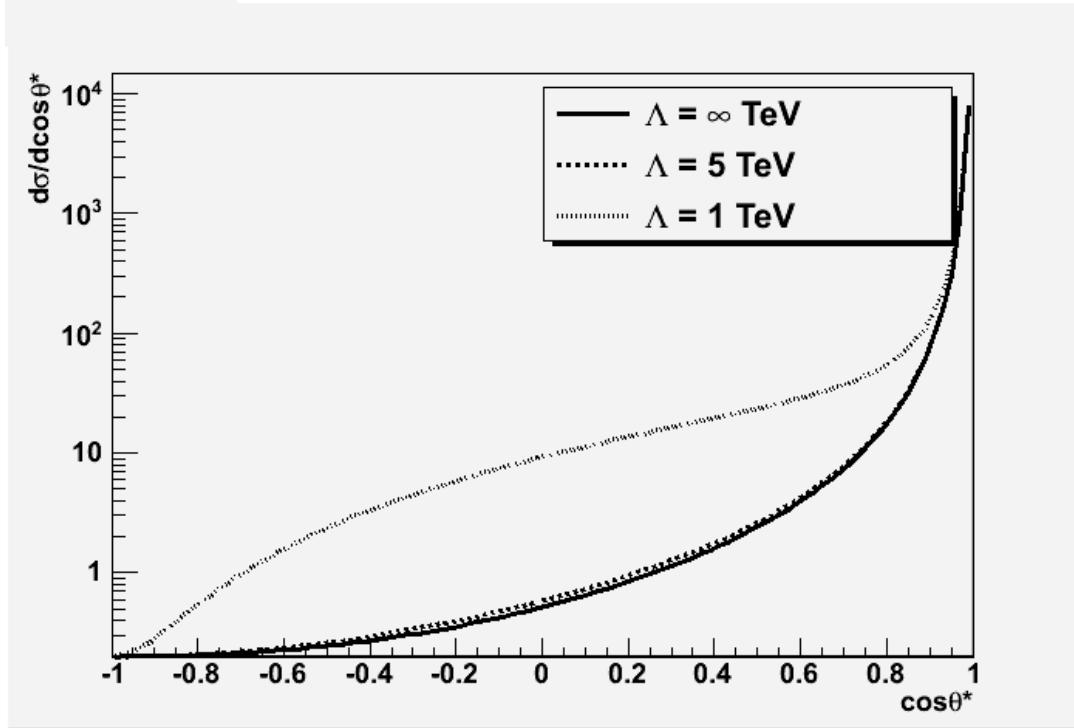
Figure 5: Angular cross-sections for $q_i \bar{q}_i \to q_i \bar{q}_i$

$$\frac{d\hat{\sigma}(q_i \bar{q}_i \to q_i \bar{q}_i)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}\left(\frac{4}{9}\alpha_S^2\left(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} + \frac{\hat{u}^2 + \hat{t}^2}{\hat{s}^2} - \frac{2}{3}\frac{\hat{u}^2}{\hat{s}\hat{t}}\right) + \frac{8}{9}\frac{A_0\alpha_S}{\Lambda^2}\left(\frac{\hat{u}^2}{\hat{t}} + \frac{\hat{u}^2}{\hat{s}}\right) + \frac{8}{3}\frac{\hat{u}^2}{\Lambda^4}\right)$$

$$(5)$$
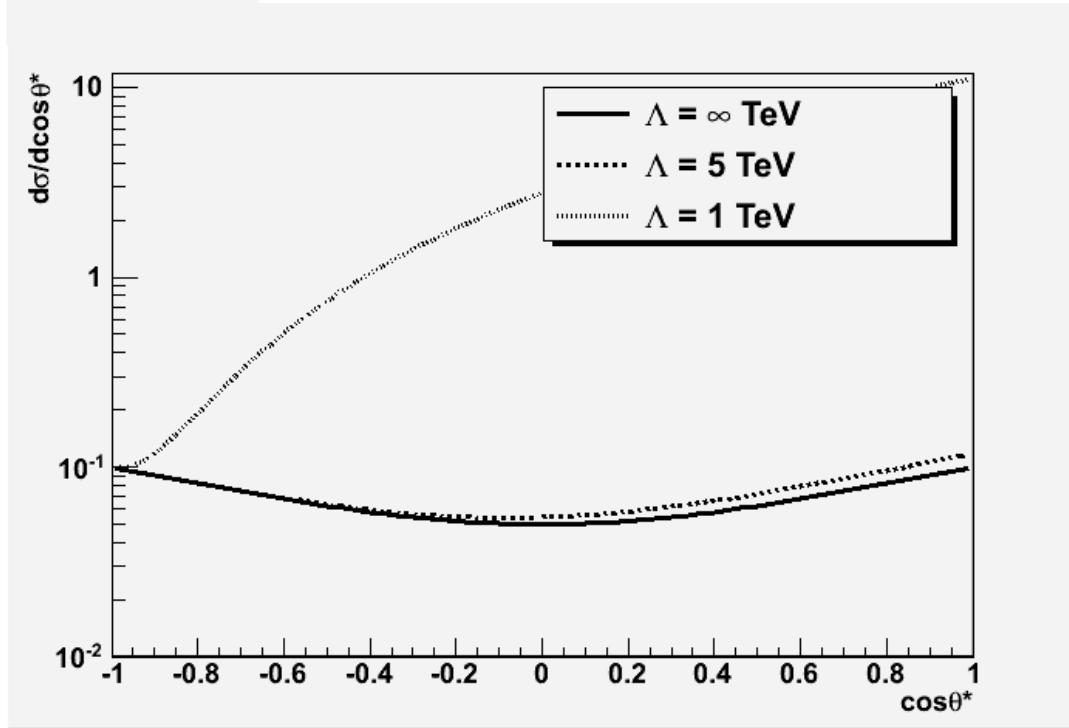
Figure 6: Angular cross-sections for $q_i\bar{q}_i \rightarrow q_j\bar{q}_j$

$$\frac{d\hat{\sigma}(q_i\bar{q}_i \rightarrow q_j\bar{q}_j)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}(\frac{4}{9}\alpha_S^2(\frac{\hat{u}^2 + \hat{t}^2}{\hat{s}^2}) + \frac{\hat{u}^2}{\Lambda^4}) \tag{6}$$
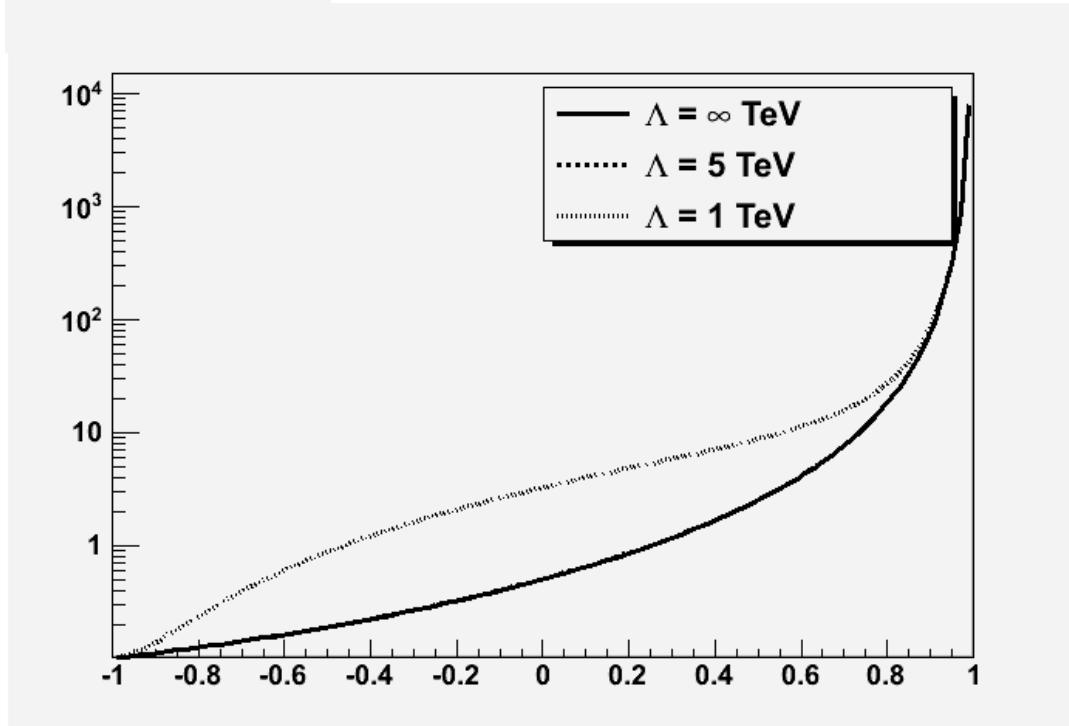
Figure 7: Angular cross-sections for $q_i\bar{q}_j \rightarrow q_i\bar{q}_j$

$$\frac{d\hat{\sigma}(q_i\bar{q}_j \rightarrow q_i\bar{q}_j)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}(\frac{4}{9}\alpha_S^2(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2}) + \frac{\hat{u}^2}{\Lambda^4}) \tag{7}$$
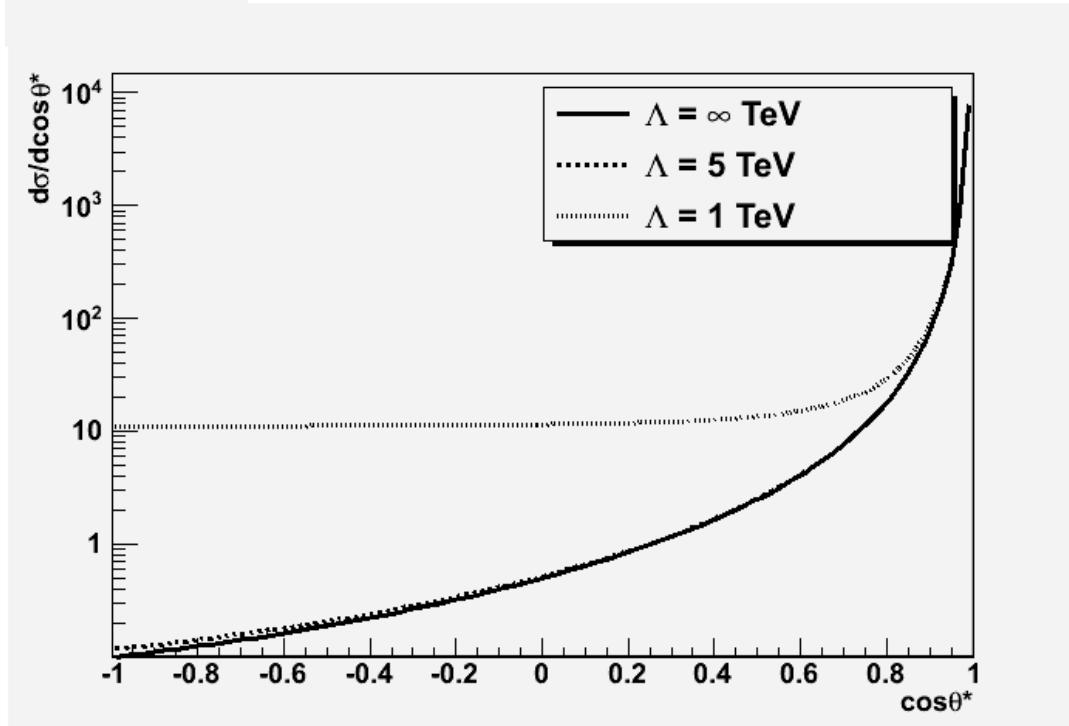
Figure 8: Angular cross-sections for $q_i q_j \rightarrow q_i q_j$

$$\frac{d\hat{\sigma}(q_i q_j \rightarrow q_i q_j)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}\left(\frac{4}{9}\alpha_S^2\left(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2}\right) + \frac{\hat{s}^2}{\Lambda^4}\right) \tag{8}$$
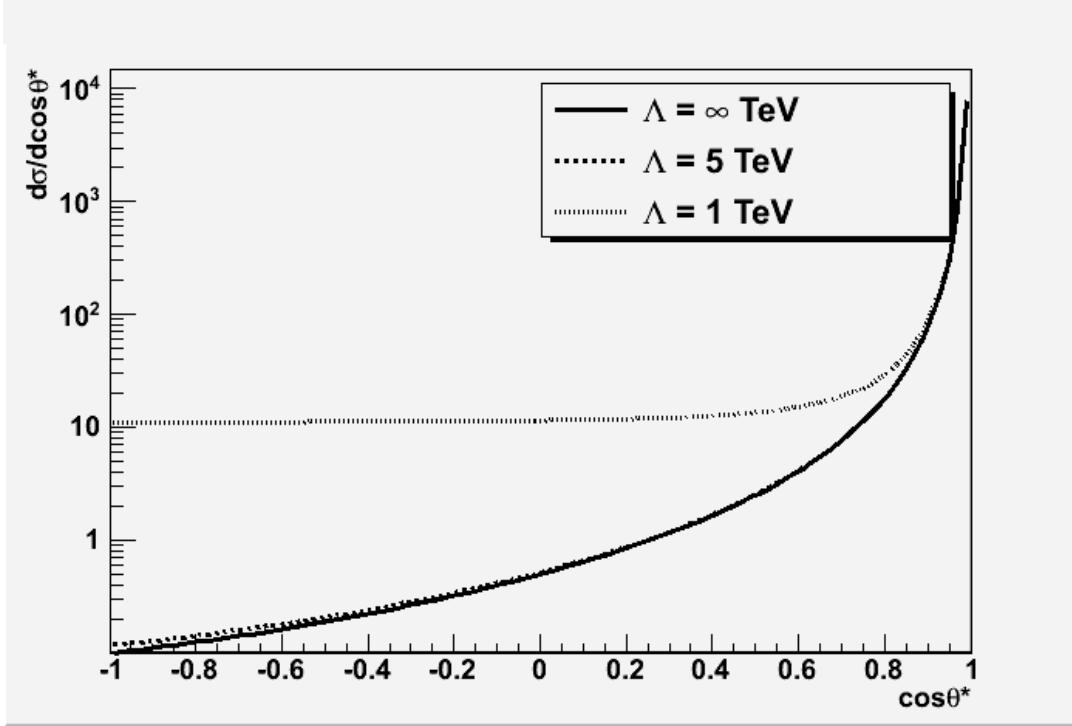
Figure 9: Angular cross-sections for $\overline{q}_i\overline{q}_j \rightarrow \overline{q}_i\overline{q}_j$

$$\frac{d\hat{\sigma}(\overline{q}_i\overline{q}_j \rightarrow \overline{q}_i\overline{q}_j)}{d\cos\theta^*} = \frac{\pi}{2\hat{s}}\left(\frac{4}{9}\alpha_S^2\left(\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2}\right) + \frac{\hat{s}^2}{\Lambda^4}\right) \tag{9}$$

Here, $\sqrt{\hat{s}}$ is the subprocess center of mass energy (effectively the dijet invariant mass), $\hat{t} = \hat{s}(1-\cos\theta^*)/2$ and $\hat{u} = \hat{s}(1+\cos\theta^*)/2$. The corresponding interactions can be seen in figure 2. α_S is the strong coupling constant and Λ remains the contact interaction energy scale. In figures 3-9, the $\Lambda = \infty$ corresponds to the Standard Model. These figures plot the angular cross-section for each scattering process at $\sqrt{\hat{s}} = 7$ TeV. Note, for example, that in equations 8 and 9 the QCD term occurs in the $\hat{t}$ channel ($\propto 1/\hat{t}^2$), which is dominated by forward scattering, while the contact interaction term occurs

in the $\hat{s}$ channel, which is isotropic. In general, in fact, a contact interaction-driven enhancement around $\cos\theta^* = 0$ is seen in all quark-quark scattering cross-sections. This enhanced isotropy can (and will be) used to search for contact interactions [17]. We do not concern ourselves with the distribution in ϕ, as the proton beams are not polarized and thus it must be isotropic.

Experiments at Fermilab's Tevatron accelerator have, prior to the LHC, excluded an contact interaction having $\Lambda < 2.8$ TeV with 95% confidence [9] [10].

3   Experimental Setup

3.1   The LHC
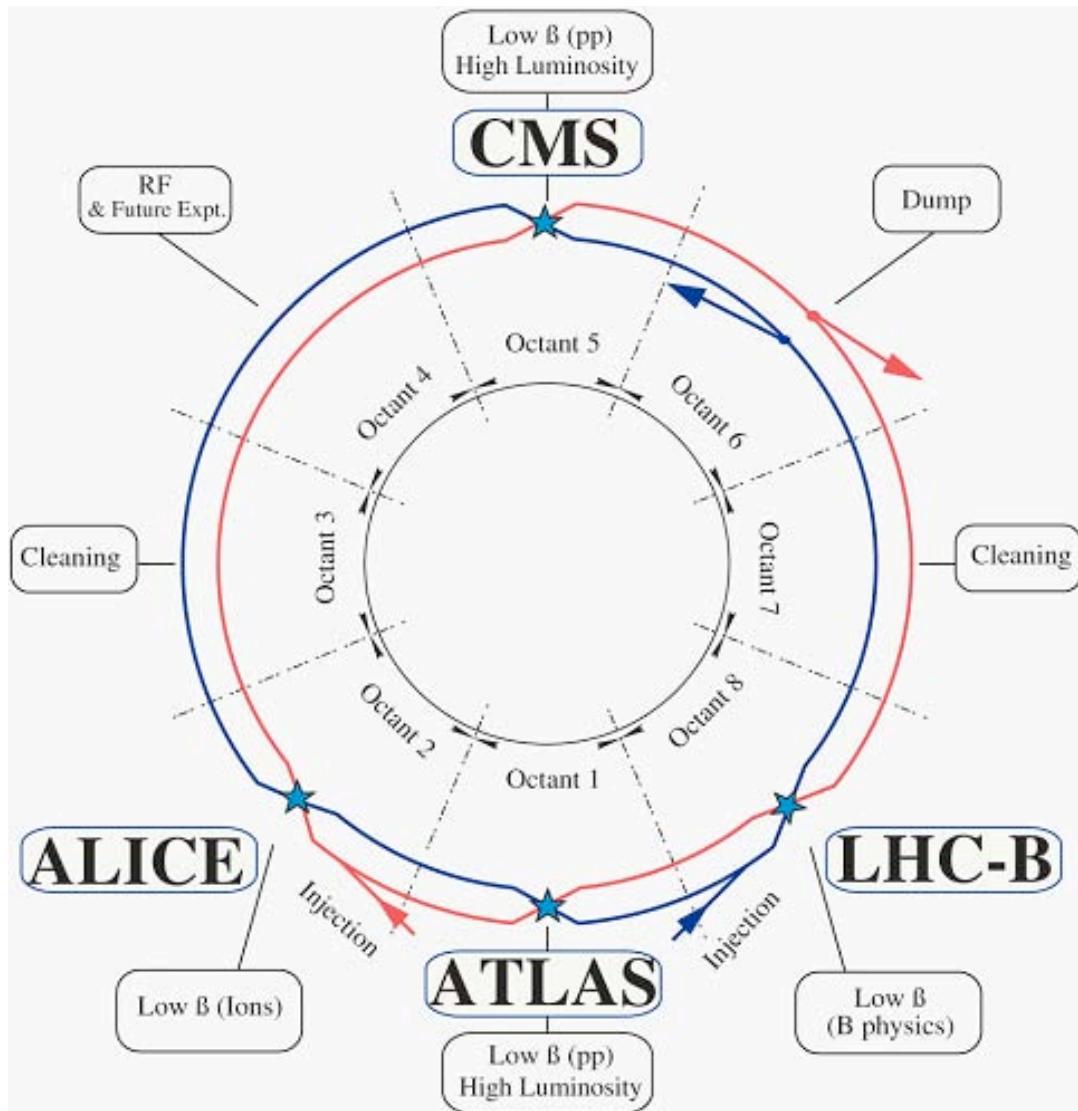
Figure 10: The LHC layout [20]

The data used in this dissertation was collected at the CMS detector, which receives its input from 7 TeV center of mass energy proton-proton collisions generated in the LHC (see figure 10). The LHC, in turn, receives its energetic protons from a chain of smaller particle accelerators. In order to reach the 7 TeV collision energies seen at the LHC, the proton acceleration must be completed in stages. Protons are first made available by the stripping of electrons from hydrogen atoms. These protons are accelerated in the LINAC2 linear accelerator to 50 MeV and injected into the Proton Synchrotron Booster, which accelerates them to an energy of 1.4 GeV. The Proton Synchrotron accelerates them to 28 GeV before they are injected into the Super Proton Synchrotron, which achieves an energy of 450 GeV. The 450 GeV protons are divided into two bunched beams and injected into the LHC, which after a ramp of approximately 20 minutes, reaches its 3.5 TeV per beam collision energy (see figure 11). This is scheduled to be upgraded to 7 TeV per beam in the future. At its full design luminosity (10^{34} cm^{-2} s^{-1}), the LHC will circulate and collide 2808 proton bunches containing 1.1×10^{11} protons at a rate of 40 MHz [20].
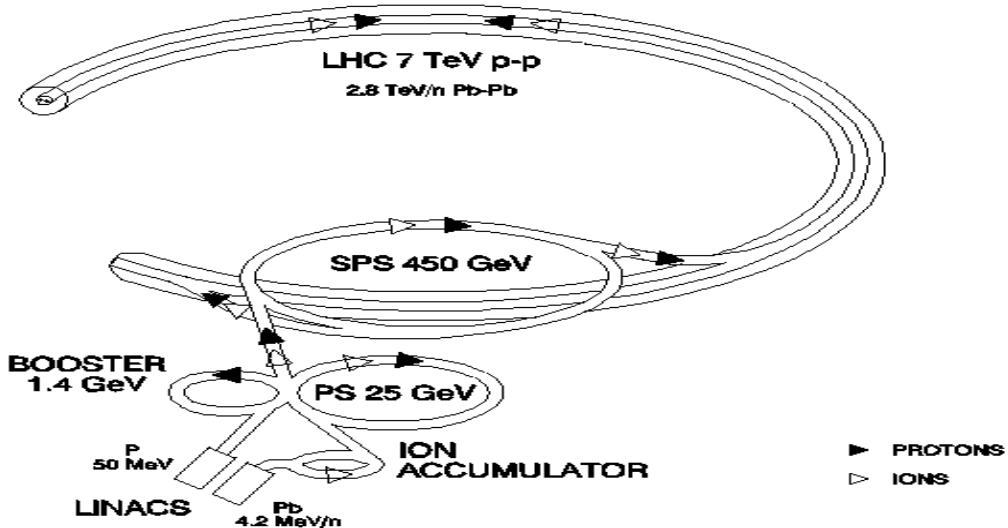
Figure 11: The LHC injection chain [20]

3.2   The CMS Detector

The CMS detector is a general-purpose particle detector, roughly cylindrical in shape, 14.6 m in diameter and 21.6 m long and weighing 12,500 t. Its centerpiece is a 6 m inner diameter, 12.5 m long superconducting solenoid with a design field strength of 4 T (3.8 T in practice). The coordinate system used has its origin at the ideal collision point, with the y-axis pointing upward, the x-axis pointing inward with respect to the LHC ring, and the z-axis pointing along the counter-clockwise beam direction.

Very-forward
Calorimeter

Superconducting Solenoid

Silicon Tracker

Pixel Detector

Preshower

Hadron
Calorimeter

Electromagnetic
Calorimeter

Muon
Detectors

Compact Muon Solenoid
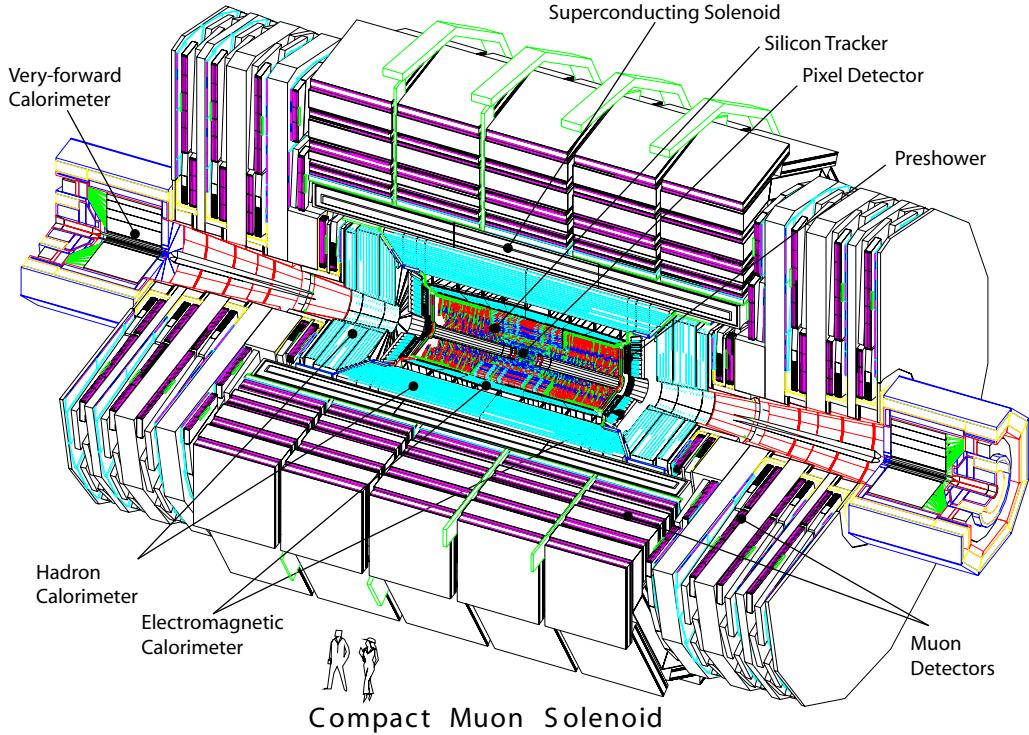
Figure 12: The CMS detector [4]

The detector elements inside the solenoid are a silicon pixel and strip based
tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass /
scintillator sampling hadron calorimeter within the solenoid volume. Out-
side the solenoid volume, integrated into the iron return yoke superstructure,
is a similar outer hadron calorimeter and multifaceted muon system consist-
ing of drift tubes, cathode strip chambers and resistive plate chambers. An
iron / quartz fiber forward calorimeter is located outside the iron return yoke
superstructure, as are several small complementary detectors, including the
CASTOR and ZDC calorimeters and the TOTEM tracker (see figure 12) [4].
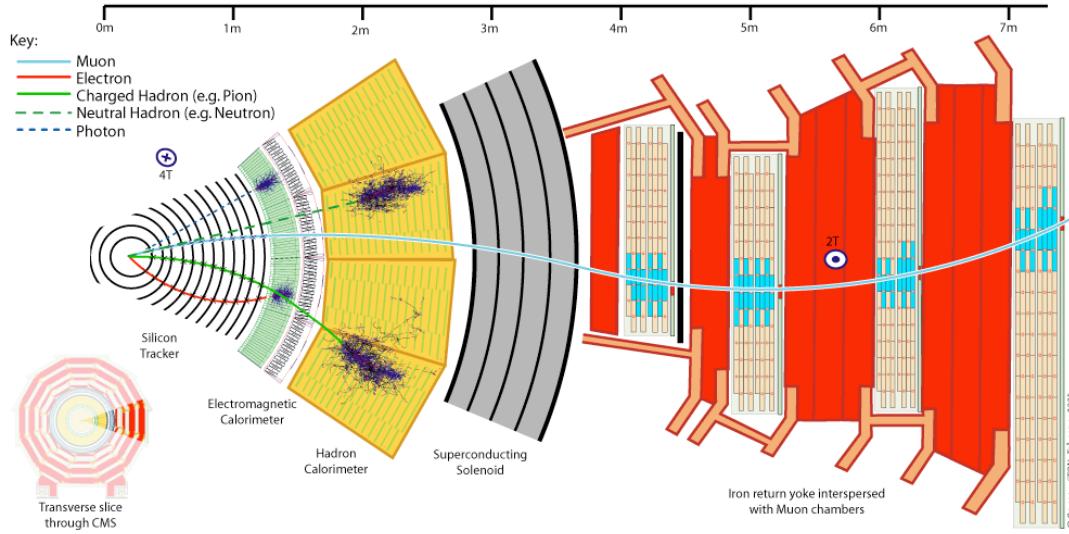
Figure 13: Particle interactions in the CMS detector [Wikimedia Commons]

Proton collisions can produce many different particles, including quarks, gluons, photons, electrons and muons. The quarks and gluons will hadronize and produce either charged or neutral hadrons (and jets, which are made up of both). Charged hadrons will leave "hits" in the tracker and energy deposits in the electromagnetic and hadronic calorimeters, while long-lived neutral hadrons will only leave energy deposits in the hadronic calorimeter. Photons will only leave energy deposits in the electromagnetic calorimeter. Electrons will leave "hits" in the tracker and energy deposits in the electromagnetic calorimeter. Muons will leave hits in the tracker and in the muon system (see figure 13).

3.2.1 The Solenoid

The eponymous solenoid of the CMS detector (see figure 14) is a 6 m inner bore, 12.5 m coil length, 4 layer superconducting NbTi coil co-extruded with pure aluminum, which is in turn reinforced with aluminum alloy, with a cold mass of 220 t. It is capable of producing a field strength of 4 T (though it is only energized to 3.8 T in practice) and holding a stored energy of 2.6 GJ. The magnetic flux is returned via a 10,000 t iron yoke system consisting of five wheels (including the 1,250 t central wheel upon which the solenoid is mounted) and six endcap discs. The stored energy density of the solenoid of 11.6 kJ per kilogram is unmatched among particle detector magnets. The magnet was fully commissioned and tested above ground in 2006 before being lowered into the detector cavern [4].
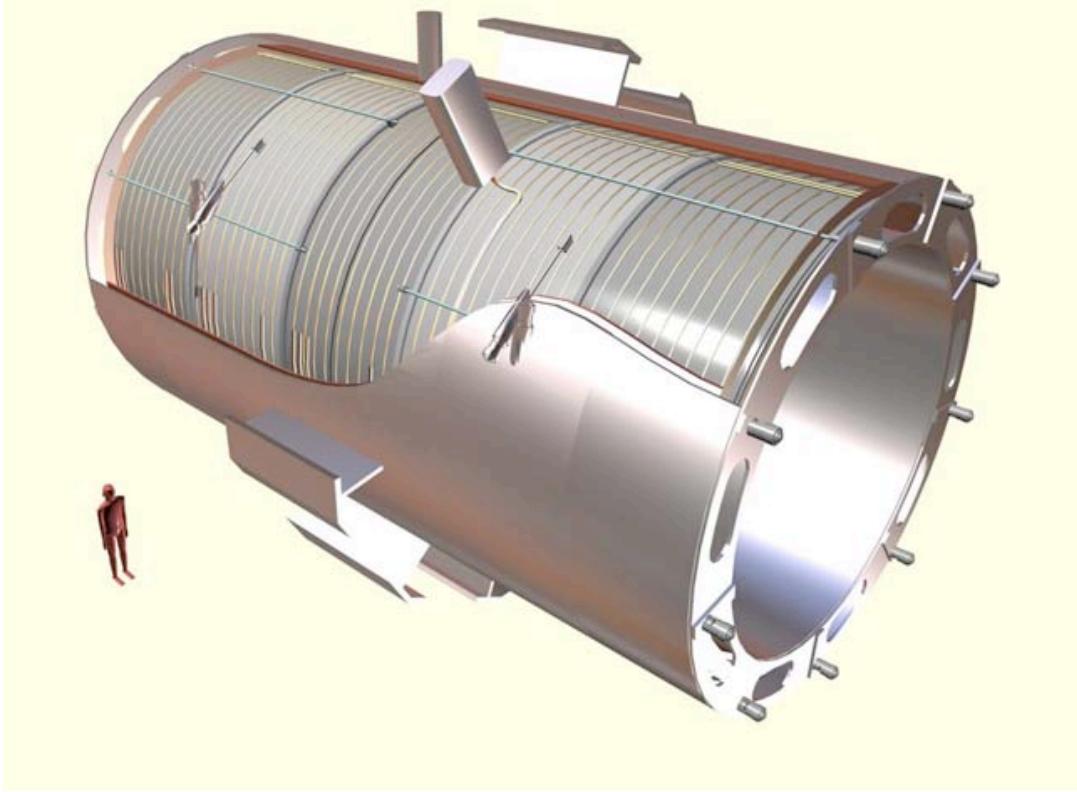
Figure 14: The CMS solenoid [4]

3.2.2 The Tracker

At the full LHC design luminosity, more than 1,000 particles from about 20 proton collisions per bunch crossing will populate the detector every 25 ns (that is, per bunch crossing). This presents unique technical challenges in terms of particle tracking, requiring an extremely high granularity and response rate from whatever system is utilized. Other technical challenges include ensuring sufficient radiation hardness for the system and facilitating the high power density and associated cooling required for such performance while minimizing the material budget in order to limit secondary particle interactions in

the tracking volume. The particle tracking duty and its associated challenges are addressed by an all-silicon system spanning a length of 5.8 m and a diameter of 2.5 m and containing about 200 m^2 of active silicon, making it the largest silicon tracker ever built. The silicon tracker operates by presenting a small doped region of silicon (either a narrow strip or a small pixel) so that it functions as a diode. Each diode region is reverse-biased at high voltage, and when a charged particle passes through it, the resultant ionization produces a current which is read out and measured. The innermost region, consisting of silicon pixels, has three barrel layers and two endcap discs out to 16 cm from the beam line. The region between radii of 20 cm and 116 cm contains the the silicon strip tracker, which is comprised of the tracker inner barrel (TIB), tracker inner discs (TID), tracker outer barrel (TOB), and tracker endcaps (TEC) (see figure 15).
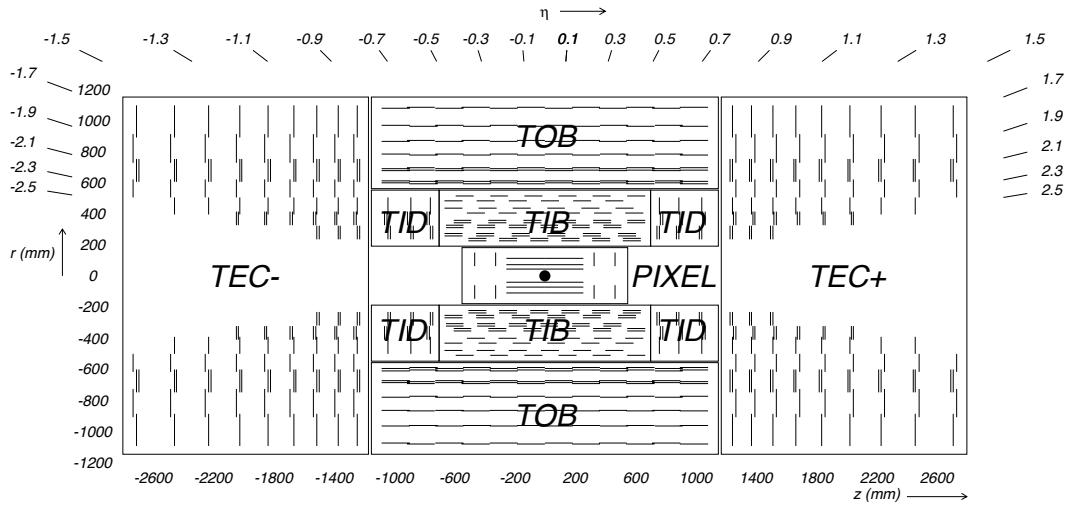
Figure 15: CMS Silicon Tracker Configuration [4]

The pixels, in addition to performing precision tracking of particles in high-occupancy region around the primary collision vertex, are also used to reconstruct, tag and measure the positions of secondary vertices resulting from the decay of long-lived particles such as bottom quarks. The pixel portion of the tracker has a cell size of 100 x 150 μm^2 and an angular coverage of $|\eta| < 2.5$. The three 53 cm barrel pixel modules are located at radii of 4.4 cm, 7.3 cm and 10.2 cm, and the four endcap discs extend from radii of 6 cm to 15 cm and are located from $|z| = 34.5$ cm to 46.5 cm and contain about 66 million pixels. Calculation of the centroid of the charge spread of each "hit" on the pixel detector allows a spatial resolution of 15 μm - 20 μm to be achieved.

As the particles from the collision move radially outward and the fine granularity of the pixel system becomes less necessary, tracking duties are taken over by the silicon strip tracker system. This system utilizes single-sided silicon microstrip sensors. The signals from ionizing particles are amplified, shaped and stored by a custom-designed integrated circuit which, upon receipt of an L1 trigger decision (see section 3.2.6), multiplexes and transmits the signal to the front end drivers where analog to digital conversion and subsequent signal information distribution takes place. Varying sensor geometries are utilized throughout the silicon strip tracker - two rectangular variants for the TIB, two rectangular variants for the TOB, and 11 wedge-shaped variants for the TEC, having either 512 or 768 strips each, reflecting the readout modularity of 256 channels. 24,244 sensor modules are utilized, leading to an active sensor area of approximately 198 m^2 and a channel count of about 9.3 million [4]. This allows the overall spatial resolution of tracks to be, as in the pixel detec-

tor, 15 μm - 20 μm for central, high transverse (with respect to the beam line) momentum tracks, with this degrading to 100 μm - 1000 μm for low transverse momentum, highly forward tracks. The transverse momentum resolution is similarly distributed between .5 % and 10 % of its average value [4].

3.2.3   The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) of the CMS detector is homogeneous and hermetic up to $|\eta| = 3.0$, and is constructed from 61,200 PbWO$_4$ crystals in the barrel region ($|\eta| < 1.479$) and 7,324 PbWO$_4$ crystals in the endcap region ($|\eta < 3.0$). In addition, preshower detectors are utilized in front of the endcaps. The crystals are read out by avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcap region. The high density crystals used allow for a fast, radiation-hard, high granularity subdetector, which is vital for effective particle detection in the LHC collision environment. The homogeneity also allows for excellent energy resolution (see figure 16).

The PbWO$_4$ crystals scintillate in the presence of electromagnetic particle showers, and the intensity and timing of this scintillation is measured for each crystal in order to determine the energy deposited. Electromagnetic particle showers occur when charged particles produce bremsstrahlung radiation while decelerating in the presence of nuclear matter. The bremsstrahlung photons then pair-produce additional charged particles, which in turn produce additional bremsstrahlung radiation. This process continues until the particles in the shower reach a low enough energy that they are absorbed by nuclear mat-

ter. Alternatively, the showering process can begin with a photon, which then undergoes the same basic cycle, starting from pair production. The characteristic radius of such a shower in a material is referred to as the material's Moliere radius.
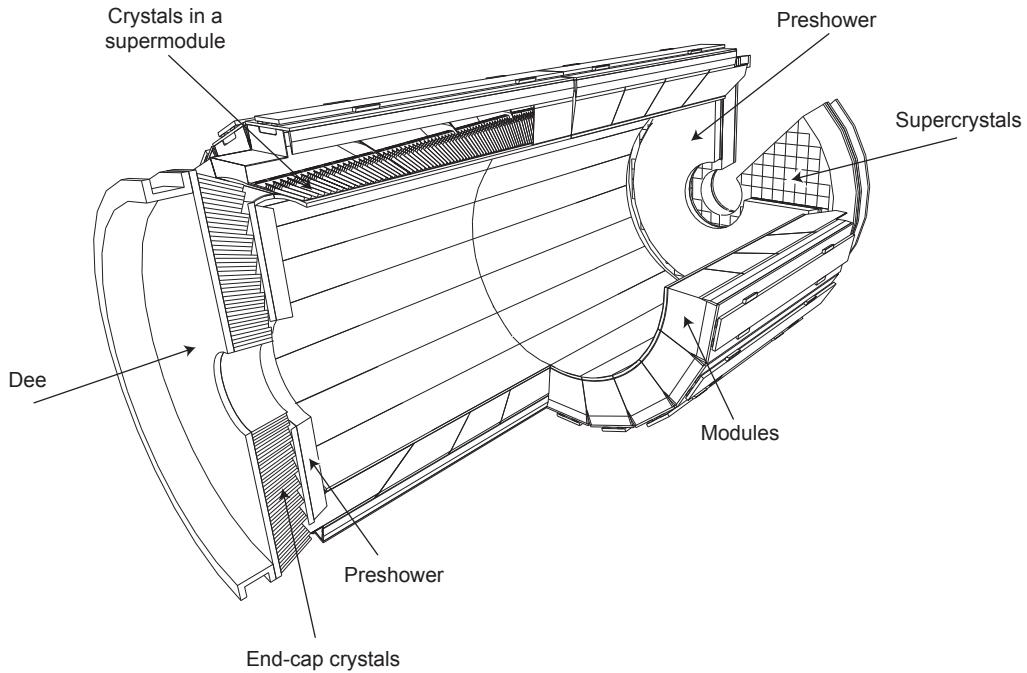
Figure 16: CMS ECAL configuration [4]

$PbWO_4$ was chosen as a scintillator because of its high density (8.28 g / cm^3), short radiation length (0.89 cm), small Moliere radius (2.2 cm), fast scintillation decay time (80 % of the light is emitted within the 25 ns LHC bunch crossing time) and general radiation-hardness. These properties allow for the construction and long-term operation of a compact, fine-granularity calorimeter in the LHC collision environment. The crystals are all polished to

exploit total internal reflection. The barrel has a pseudorapidity coverage of $|\eta| < 1.479$, with a 360-fold granularity in ϕ and a (2 x 85) fold granularity in η. The barrel crystals are, as mentioned previously, trapezoidally tapered, and arranged in a quasi-projective geometry to prevent the alignment of cracks with potential particle trajectories. The crystals have a depth of of approximately 25.8 radiation lengths. Likewise, the endcaps have a pseudorapidity coverage of $1.479 < |\eta| < 3.0$. The endcaps, which are split into two Dees on each side, are located 315.4 cm from the interaction point, and extend for 24.7 radiation lengths. The endcap crystals are arranged in a rectangular grid, with the crystals pointing at a focus 1,300 mm from the interaction point, giving off-pointing angles of 2 to 8 degrees. The preshower detector provides additional identification power for electrons and neutral pions in the fiducial region of the endcaps between $1.653 < |\eta| < 2.6$. Each unit is a 20 cm thick system made up of a lead radiator to initiate electromagnetic showers and silicon strip sensors to sample and profile the showers.

Each on-detector electronics module reads out a 5 x 5 trigger tower. Off the detector, each front end board transmits trigger primitive data to a timing trigger and control (TTC) card and readout data to a data concentration card (DCC), which communicate with the CCS (clock and control system) boards and selective readout processors (SRP). These manage timing and determine which crystals are fully read out in order to control event data size. Event data is then, upon receipt of an L1 trigger decision (see section 3.2.6), extracted from this system for calorimeter tower reconstruction.

The ECAL energy resolution is parametrized with respect to energy E as

having a stochastic term S, a noise term N and a constant term C; $(\sigma/E)^2 = (S/\sqrt{E})^2 + (N/E)^2 + C^2$. Test beam results indicate typical values of $S = 2.8\%$, $N = 0.12$ GeV and $C = 0.3$ %. In practice, this results in an overall energy resolution varying from a couple of percent at very low transverse (with respect to the beamline) momentum, dropping sharply and leveling out to about 0.4 % around a transverse momentum of 75 GeV [4].

3.2.4   The Hadronic Calorimeter

A hadronic calorimeter is necessary for measurement of both jet energy and missing transverse energy (that is, an imbalance in energy perpendicular to the direction of the beam), the latter of which may be the result of neutrinos or new and exotic particles. The second type of measurement in particular requires a high degree of angular coverage and hermeticity. Occupying the outermost region of the interior of the solenoid, from a radius of 1.77 m to 2.95 m, the barrel hadron calorimeter (HB), a brass-scintillator sampling calorimeter, provides angular coverage up to $|\eta| < 1.4$. This is complemented by a similar "tail-catcher" calorimeter layer outside the solenoid volume (HO). The endcap calorimeters (also of a brass-scintillator sampling design) provide angular coverage up to $|\eta| < 3.0$, while a Cerenkov-based iron and quartz fiber forward detector (HF) located 11.2 m from the interaction point increases coverage to $|\eta| < 5.2$ (see figure 17). Several small auxiliary detectors (CASTOR and ZDC) increase, with limitations, coverage up to $|\eta| < 8.3$, ensuring a tremendous degree of hermeticity.
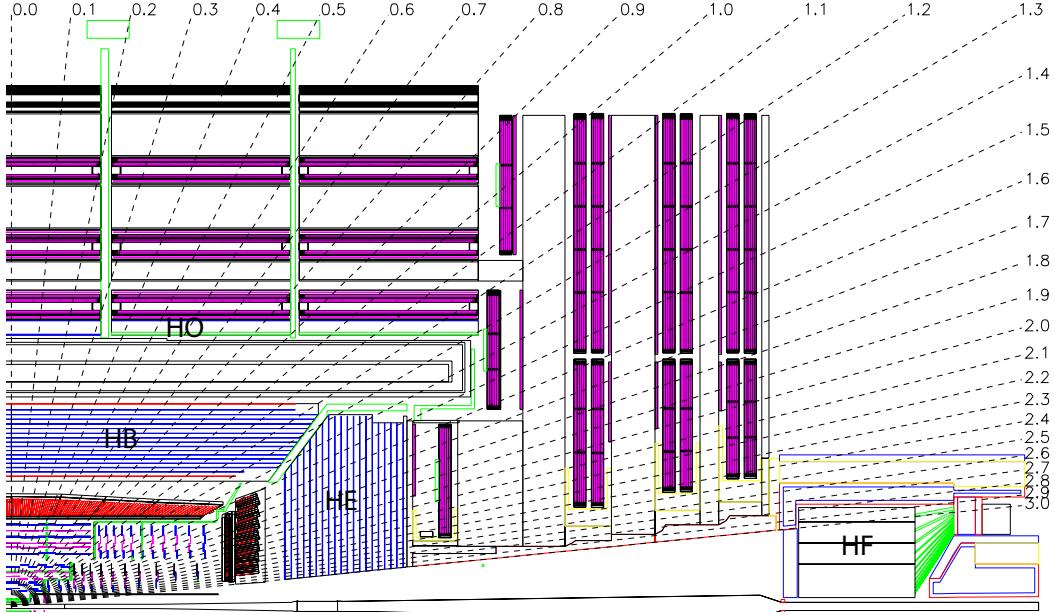
Figure 17: CMS HCAL configuration [4]

A hadronic calorimeter functions similarly to an electromagnetic calorimeter. In the CMS HCAL, hadron showers are produced via strong (QCD) nuclear interactions between the high-energy collision byproducts and an absorber material (brass, in this case). A scintillating material (a type of plastic, in this case) samples a predicted portion of the shower energy. In the case of charged hadrons, some of the energy is deposited electromagnetically as well, as in the ECAL.

The HB consists of 36 identical azimuthal wedges split into two halves (HB+ and HB-). Each wedge is built up of layers of plastic scintillator plates and flat brass absorber plates which are parallel to the beam axis and arranged in a staggered layering configuration to eliminate any projective dead

material. The inner and outermost absorber plates are made of steel to provide structural support. The scintillators are sectored into projective towers in η and ϕ. The total depth of the absorber plates perpendicular to the beam axis is 5.82 interaction lengths. Geometrically, this increases with increasing pseudorapidity, up to 10.6 interaction lengths at $|\eta| = 1.3$. The ECAL adds an additional 1.1 interaction lengths of material. The tiles of each wedge in ϕ are grouped in an individual tray, which allows for easy modular assembly, disassembly and replacement of the HB elements. The light from the scintillating tiles used in the barrel is brought out via green wavelength shifting fibers which are embedded in machined grooves in each tile. Additional tile features include a hollow 1 mm tube through which a radioactive source is moved for calibration and, in layer 9, additional quartz fibers which are used to inject ultraviolet laser light for calibration purposes. The wavelength shifting fibers are, after exiting the scintillator, spliced to clear double-clad fibers which in turn are coupled to optical connectors at the ends of the trays. These are then attached to optical decoding units (ODUs), which arrange the fibers into readout towers and couple them to hybrid photodiodes (HPD) for readout. An additional fiber is used to directly inject laser or LED light into each HPD for calibration purposes.

From pion test beam data, the energy resolution in the barrel was measured to be $\frac{\sigma}{E} = \frac{115.3}{\sqrt{E}}\% \oplus 5.5\%$ (with the symbol $\oplus$ indicating addition in quadrature) [8]. Upon reconstruction, jet p_T resolution is estimated (using Monte Carlo methods calibrated via test beam data) to vary as shown in figure 18 (here, JPTJets are "JetPlusTrack" jets - reconstructed jets that use tracker data in

addition to calorimeter data to refine the position of the jet, and PFJets are "ParticleFlow" jets - jets that use a reconstruction of each individual particle as input to the jet). Position resolutions in η-ϕ space are similarly determined (see figures 19 and 20) [5].
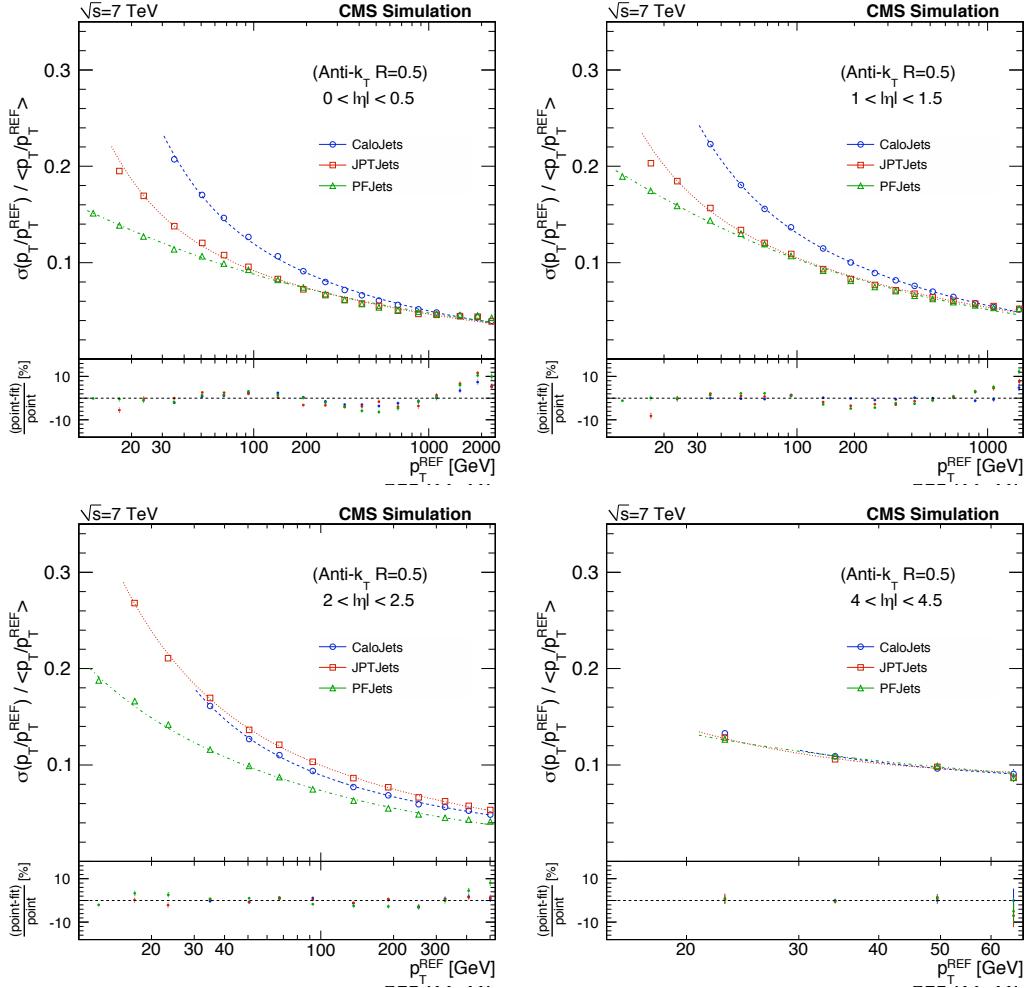
Figure 18: Jet p_T resolutions in four pseudorapidity regions as a function of p_T [5]
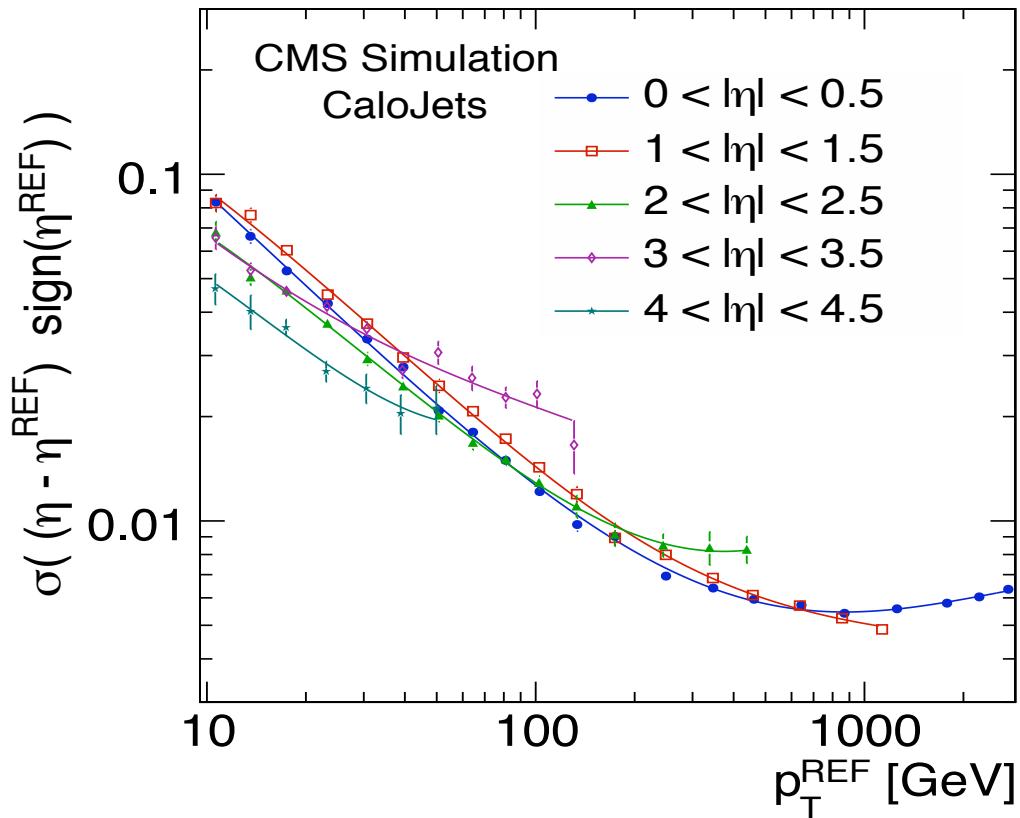
Figure 19: Jet η resolutions in five pseudorapidity regions as a function of p_T [5]
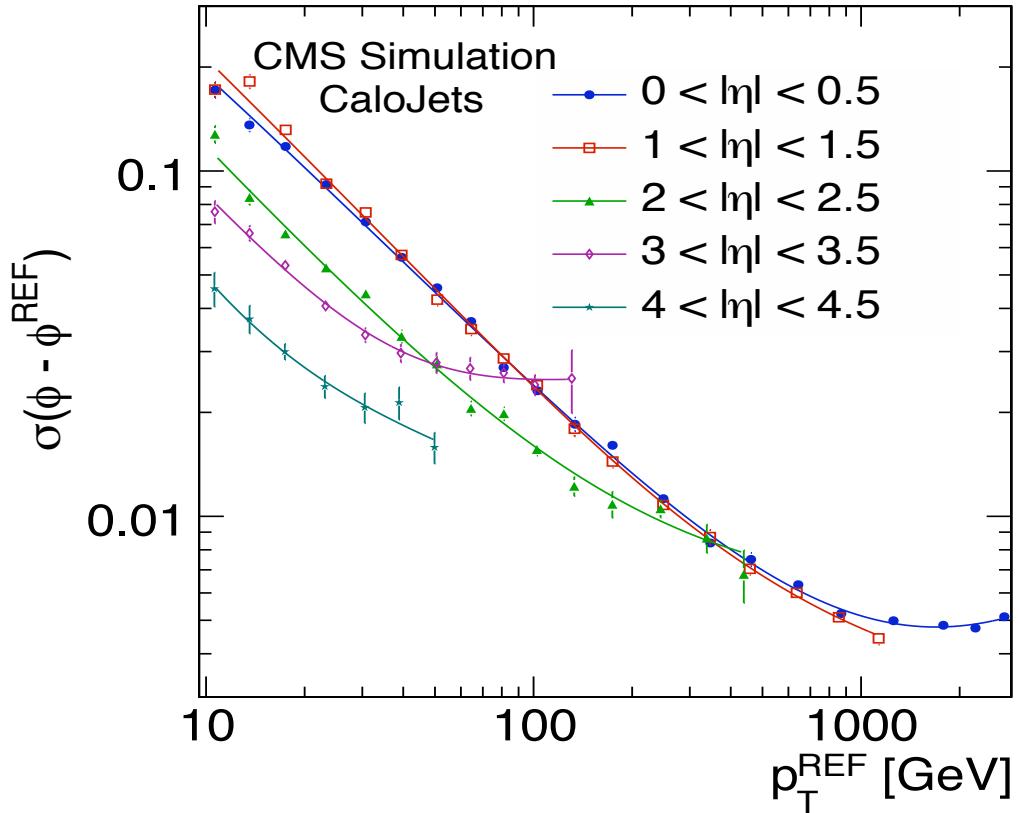
Figure 20: Jet ϕ resolutions in five pseudorapidity regions as a function of p_T [5]

The HE covers a pseudorapidity range of $1.3 < |\eta| < 3.0$, a portion of space containing a full 34 % of the final state particles produced in the LHC's proton-proton collisions. Due to the high radiation intensity in this region, the HE must be extremely radiation-hard. Additionally, due to its position in the superconducting solenoid, the absorber elements must be non-magnetic. As in the HB, brass is used as an absorber material. The absorber layers are assembled in a staggered configuration to eliminate projective dead material

and help ensure hermeticity. The outer regions have cutouts to allow for installation of photodetectors and front end electronics. Light is collected from the trapezoidally shaped scintillating tiles via wavelength shifting fibers. An ultraviolet nitrogen laser fiber-injects light into each of the tiles, allowing for diagnostic checks of the optical paths in the system. As in the HB, multipixel HPDs are used for light readout.
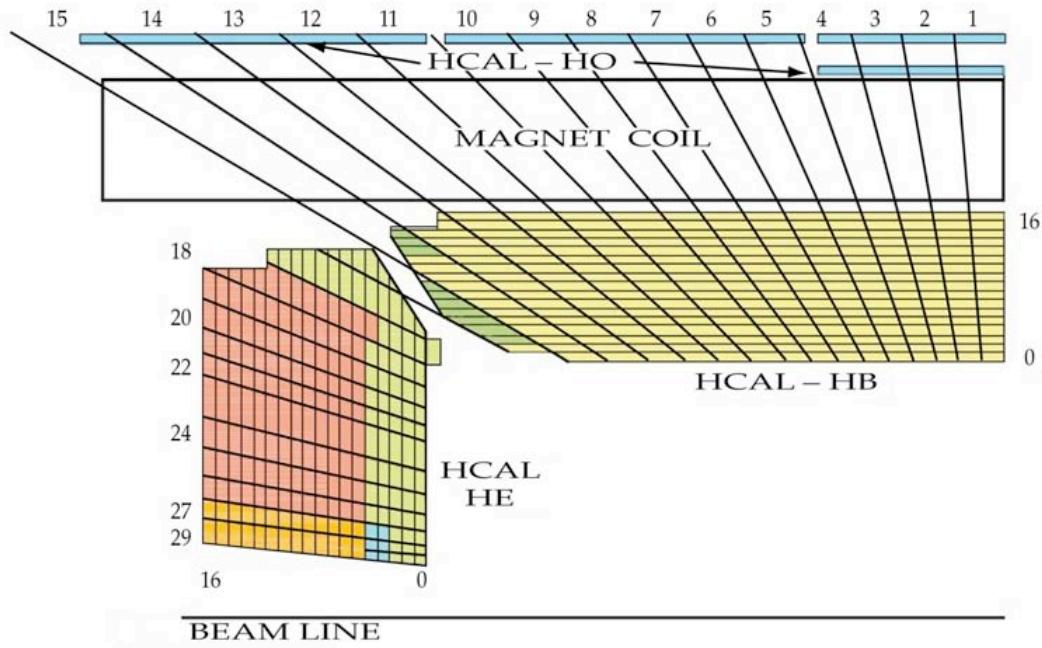
Figure 21: CMS HCAL segmentation [4]

The HO provides additional sampling depth for the $|\eta| < 1.3$ region, acting as a "tail-catcher" (the efficacy of which has been shown in simulations) and utilizing the solenoid as additional absorber material. It is installed as the first sensitive layer on each of the five central wheels of the iron return yoke superstructure. The centermost wheel has two active layers (layers zero and

one), while each of the other four wheels on which the HO is installed has one (layer one). The geometry of the HO is constrained by the configuration of the muon system (in addition to the return yoke superstructure and the feed lines to the superconducting solenoid), and thus it must have a relatively low-profile. Projectively, the HO tiles roughly map the tower structure of the HB. As in the HB and HE, wavelength shifting fibers are used to move light from the scintillating tiles to the photodetectors.
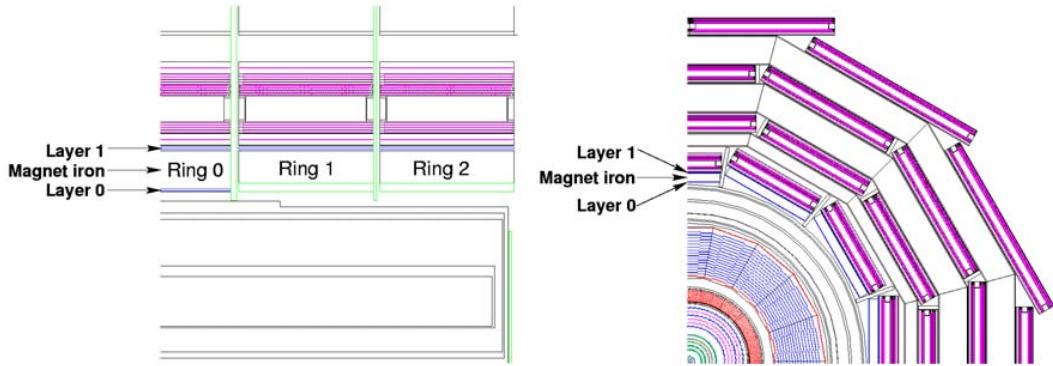
Figure 22: CMS HCAL HO configuration [4]

The forward region of the CMS detector will be subjected to an unprecedented level of particle flux in terms of both number and energy of particles. Sustained successful operation of the detector thus depends on the radiation hardness of the detection material. With this in mind, quartz fibers were chosen as the active detection material for the HF. Unlike the scintillators of the HB, HE and HO, the quartz fibers detect particles via the Cherenkov light generated by energetic particle showers above the Cherenkov threshold (190 keV for electrons). Due to this process, the HF ends up being extremely sensitive to the electromagnetic shower components (as opposed to the hadronically-

dominated scintillators). The quartz fibers are embedded in two steel absorber structures for each of the forward calorimeters. Half of the fibers run for the full length of the detector (1.65 m or approximately 10 interaction lengths) and the other half start 22 cm from the front face of the HF. They all run parallel to the beam line and are bundled to form towers. Like the HB, each forward calorimeter is subdivided into 36 modular wedges in ϕ. The whole assemblage is housed in radiation shielding, while a large rear "plug" structure acts as a shield as well. Air-core light guides carry the light from the fibers to photomultiplier tubes in the readout boxes. Due to the high radiation flux experienced by the HF, radiation damage (raddam) monitors are attached to several reference fibers in the detector and are used to monitor performance as a function of integrated luminosity. Additionally, the HF is used as an instantaneous luminosity monitor via zero-counting (in which the average fraction of empty towers is used to calculate the average number of interactions per bunch crossing). This is done via an HF Luminosity Transmitter (HLX) board attached to each of the HF HTRs which transmits the tower occupancy and luminosity data to the DAQ.

Throughout the HCAL, the electronic structure is of a similar architecture: some sort of optoelectrical transducer connected to a fast charge-integrating ADC, which is in turn connected via a gigabit optical connection to the off-detector electronics in the service cavern. These off-detector electronics take the signal and generate trigger primitives, which are sent to the calorimeter trigger. Upon receipt of an L1A trigger decision, the signal and trigger primitives are then pipelined to the DAQ (described in section 3.2.6). Several PCs in

the CMS control room control the high and low voltage supplies and provide a convenient interface for monitoring and recording temperature, humidity, and other environmental parameters. The system also sets pedestals and timing parameters as well as controlling the various calibration settings and systems. A configuration database contains the mapping information for all the HCAL components and subsystems and logs and downloads the relevant parameters to the various specific subsystems and components. From the photoelectric transducer (either an HPD or PMT), the analog signal is converted into a digital one via a charge integrating ASIC which four capacitors that are connected to the input in turn during each 25 ns bunch crossing period. The integrated charge is converted to a seven bit non-linear scale. The digital outputs of three such devices are combined with monitoring information to make a 32 bit data word, which is read out via a gigabit optical link chip. The optical output from this chip is transmitted via fiber into the service cavern, where it is received by an HCAL Trigger and Readout (HTR) board. The HTR board consists of the Level One pipeline and trigger primitive generation system. Trigger primitives are transmitted to the Regional Calorimeter Trigger via Serial Link Boards. Upon receipt of an L1A signal, a data packet for the DAQ is prepared by the HTR which is then transmitted to the an HCAL Front-End Driver to be concentrated and passed on to the DAQ [4].

3.2.5 The Muon System
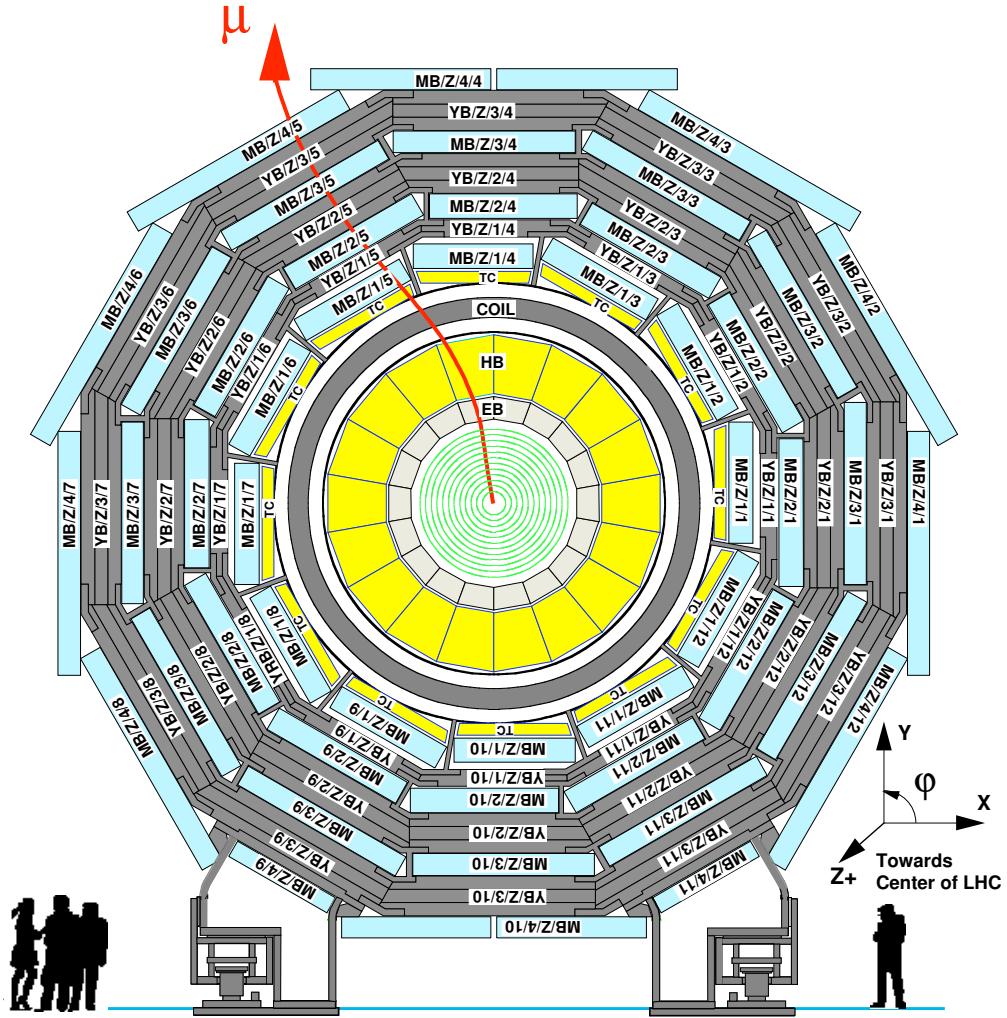
Figure 23: CMS muon drift tube configuration [4]

The muon system, as referenced in the experiment's name, is a critical component for a multitude of physics analysis (though it is not, in fact, used in the analysis upon which this dissertation is focused). It is particularly useful for Higgs searches (the four lepton decay channel, specifically), which are among

the most important of the long term analyses for the LHC experiments. The muon system serves three purposes: muon identification, momentum measurement, and triggering. High momentum resolution is enabled by the high magnetic field of the superconducting solenoid and the iron return yoke. Three types of gas-based muon detectors are used in the CMS muon system: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). The muon system has a cylindrical barrel region and two planar endcap regions , totaling about 25,000 m^2 of detection plane, which necessitates robust and inexpensive detection technology.

In the barrel region, the muon rate is low and the magnetic field changes slowly and is mostly contained within the solenoid and return yoke structure; thus, rectangular-cell DTs are used (see figure 23). A pseudorapidity range of $|\eta| < 1.2$ is covered, and four radial layers are incorporated into the return yoke superstructure. The first three layers each contain two groups of four chambers per layer, one group of four for measuring the r-ϕ position and one for measuring the z-position. The fourth and outermost layer does not contain the z-position-measuring group. Simple and effective meantimer circuits are used to provide excellent time resolution in the drift tubes, allowing for efficient bunch crossing identification. In the endcap regions, the muon rate is higher and the magnetic field is much more non-uniform, so a different technology is more optimal. Due to their radiation hardness, fine segmentation and fast response time, CSCs were selected as the primary detection device for this region. In the direction along the beamline, four layers of CSCs are utilized per endcap and provide a pseudorapidity coverage of $0.9 < |\eta| < 2.4$.

The chambers are perpendicular to the beamline and installed between the wheels of the iron flux-return superstructure (see figure 24). RPCs are placed throughout the barrel and endcap ($|\eta| < 1.6$) as a dedicated trigger element, as they are fast and highly segmented with a sharp transverse momentum threshold.
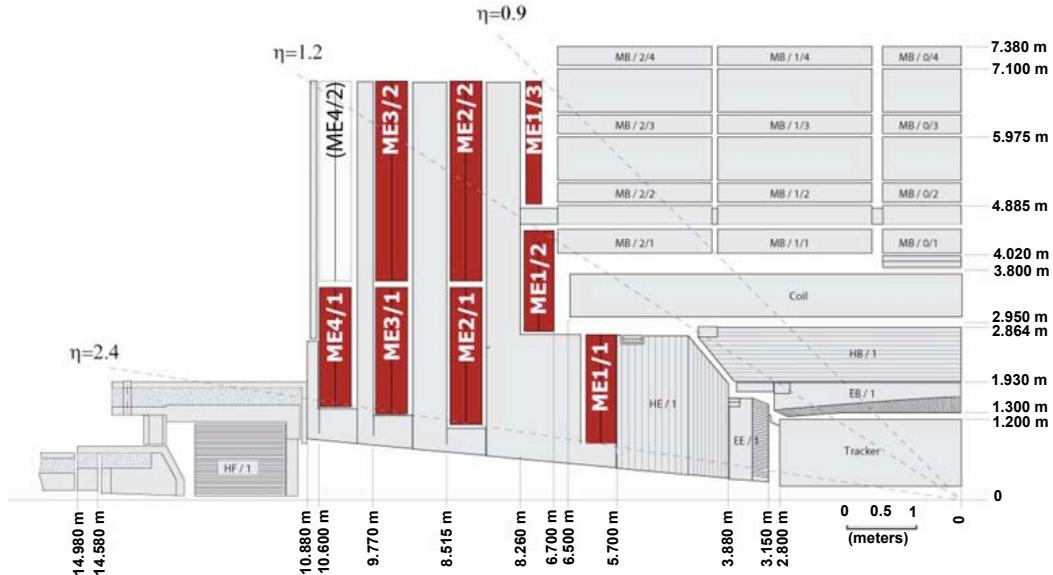
Figure 24: CMS endcap muon CSC configuration [4]

3.2.6 The Trigger and DAQ

At full design lumionsity, the LHC will collide 40,000,000 proton bunches per second with as many as 20 collisions per bunch crossing. To store and process each of these events would require unimaginable computing capacity, so the trigger system reduces this to a more manageable rate. This is achieved via a two stage process. First, a customized hardware system known as the Level

One (L1) trigger reduces the rate to the order of 10-100 kHz while pipelining and buffering the full readout data, using coarsely segmented muon and calorimeter data to make its decisions. Then, a custom software-based system called the High Level Trigger (HLT), which has access to the full readout data and is run on a dedicated processor farm, reduces the rate to approximate 100 Hz and sends the readout data on to the reconstruction farm.
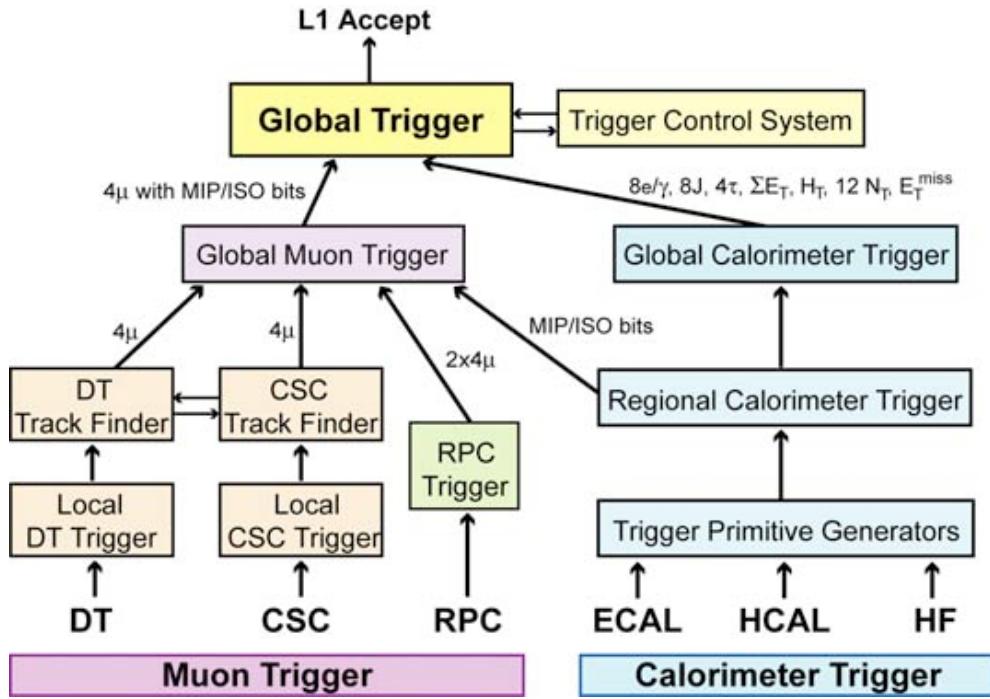

Figure 25: CMS trigger architecture [4]

The L1 system operates on a three-tiered system: local, regional and global. The local triggers, or Trigger Primitives (TPs), are based on calorimeter tower energy deposits and hit patterns in the muon chambers. Regional triggers combine TPs and use pattern logic to localize, categorize and rank various trigger

objects. The Global Calorimeter Trigger (GCT) and Global Muon Trigger (GMT) take the highest ranked objects and transfers them to the Global Trigger (GT), which is the final entity determining the L1 acceptance (see figure 25). Upon transmission of an L1 Accept (L1A) signal to the subdetectors via the Timing and Trigger Control (TTC) and to the HLT, the readout data is transferred to the HLT. The HLT bases its decision on complex and programmable pattern recognition and coarse analysis algorithms, as well as the readiness of the subdetectors and DAQ (which is determined by the Trigger Control System (TCS)). A decision from the HLT transfers the readout data from the DAQ to reconstruction and analysis farms for processing. The DAQ operates synchronously with the HLT, and must therefore be capable of accepting an input event rate of up to 100 kHz. The subdetector front-ends store their data continuously in buffered pipelines. Upon receipt of an L1A signal via the TTC, the front-ends push the pipelined data to the DAQ. The Event Builder then assembles all the relevant event fragments and transmits the assemblage to a single Filter Unit in the Event Filter for additional processing. The DAQ operates in eight slices, each of which is a nearly autonomous system for processing single events at up to 12.5 kHz per slice. Back-pressure is built into the DAQ, and the Trigger Throttling System (TTS) is used to prevent buffer overflows resulting from this. It provides fast feedback from the subdetector front-ends in order to manage the throttling before any overflow can occur. Though trigger thresholds and prescales are set in order to ensure optimal use of the DAQ, transient fluctuations may result in the activation of the TTS, resulting in brief periods of latency or dead-time. During each

Luminosity Section (a fundamental measure of time for the LHC consisting of 2^{20} LHC orbits or 93 s), trigger thresholds and prescales are held constant. A combination of custom hardware and high node count PC processing farms allows the Event Filters and Builders to achieve the desired performance. Finally, the DAQ passes the event information on to full reconstruction and processing farms [4].

The particular triggers used for this study are simple p_T threshold-based jet triggers - i.e. triggers that accept events with one or more jet with a p_T greater than a given threshold. The threshold values selected and the manner in which the streams are combined is described in section 5.

4 Event Reconstruction

In particle physics phenomenology, a jet describes a tightly-focused energetic cone of particles produced by the hadronization of a quark or gluon following a particle collision. Hadronization is the process by which, due to color confinement (which does not allow the existence of a "free" quark or gluon), hadrons are spontaneously produced as energetic "free" quarks and gluons "pull" additional quarks and gluons out of the vacuum and recombine with them. Though the exact theoretical dynamics behind them are not entirely understood, these jets are among the most common and phenomenologically well-understood signatures of quantum chromodynamic processes in particle colliders, and form the basis of numerous tests of QCD and searches for new physics.

In a typical calorimeter system, energy deposits are read out from pseudo-cylindrically projective "towers" centered around the nominal interaction vertex. Energy deposits in neighboring towers are clustered and summed to make a single calorimeter jet. Multiple algorithms exist for doing this. For our study, we select the anti-k_t algorithm, which produces a roughly conical jet shape. The algorithm proceeds in the following fashion.

A distance measure d_{ij} is introduced between entities i and j (particles in generator-level Monte Carlo or calorimeter towers in the case of fully-simulated Monte Carlo or actual data) and d_{iB} between entity i and the beam. We define the measures as follows:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{10}$$

$$d_{iB} = k_{ti}^{2p} \tag{11}$$

Here $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and k_{ti}, y_i and ϕ_i are the transverse (with respect to the beam direction) momentum, rapidity and azimuth, respectively, of entity i. The rapidity, which is identical to the pseudorapidity η in the limit in which the object in question is traveling at the speed of light, is defined as $y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$, where p_z is the component of momentum along the beam axis. We have two variable parameters, R, the cone radius ($R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$), and p, a power scale. Choices of $p = 1$ and $p = 0$ return the well-known k_t [3] and Cambridge / Aachen [14] algorithms, respectively. The anti-k_t algorithm used in this study has $p = -1$ [2]. Because of the choice of p, control of the jet

shape is vastly dominated by hard entities, around which softer entities will cluster in a cone-like shape of radius R. The clustering process proceeds by identifying the smallest d. If it is a d_{ij}, entities i and j are recombined. If it is a d_{iB}, entity i is considered to be a jet and is removed from the entity list. Following each iteration, the distance measures are recalculated and the procedure is repeated until no entities remain. Comparative algorithmic performance (jet shaping tendencies, specifically) can be seen in figure 26. The plane coordinates represent the unfolded η-ϕ space of the detector, the tower heights correspond to tower p_T (transverse momentum), and the shaded regions correspond to the shapes of the jet cones produced by the different jet reconstruction algorithms, with each color being a different jet.
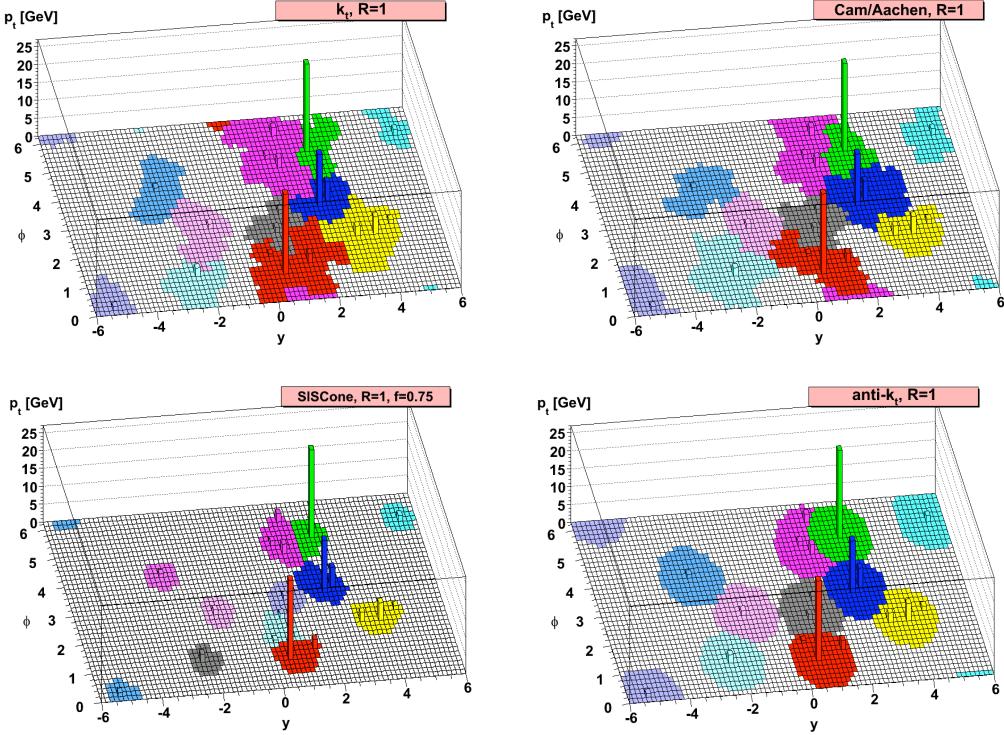
Figure 26: Comparison between jet shapes for a simulated sample event utilizing four different jet clustering algorithms; each color is a different jet [2]

Notice that the anti-k_t jet cones are the most consistently shaped, and that the strength of that shaping is dominated by the hardness of the central entity. This consistency is apparent as well in a comparison of average jet area as a function of jet transverse momentum (see figure 27) [2].
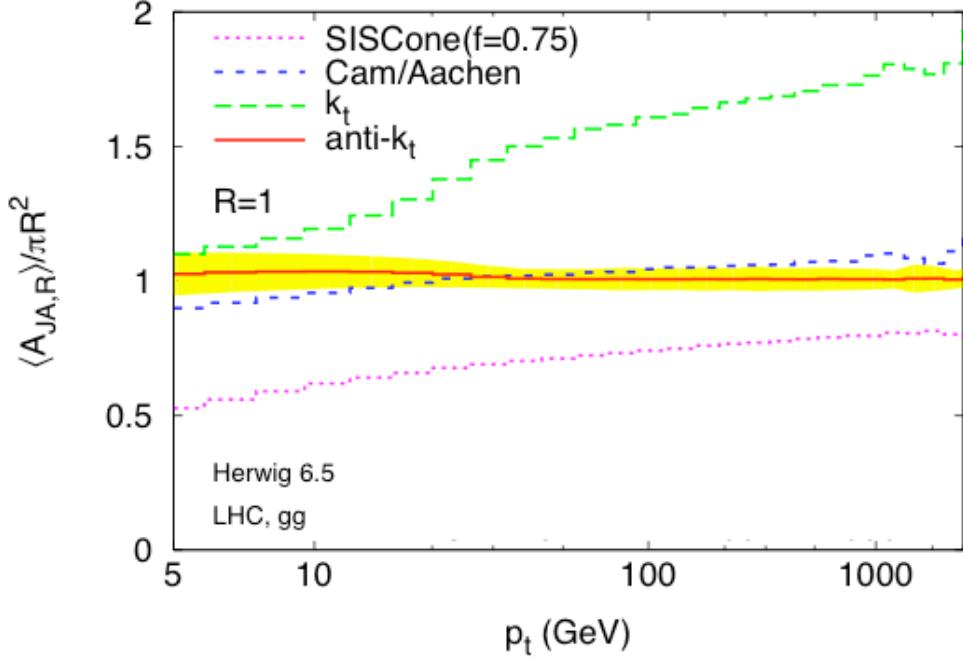
Figure 27: Average jet areas in simulated dijet events as a function of jet p_T for four different jet clustering algorithms; the yellow band around the anti-k_t line corresponds to jet area fluctuations; this is not included for the other three algorithms [2]

Our data is reconstructed in this fashion using the towers from the barrel regions of both the hadronic and electromagnetic calorimeters. The energy deposits are summed per projective tower and then clustered using the described algorithm and parametrized as four-vectors in the detector coordinate system. Following this the two leading (highest p_T) jets are selected, their positions in η-ϕ-space noted and their invariant mass $M_{jj} = \sqrt{(E_1 + E_2)^2 - (\overrightarrow{p}_1 + \overrightarrow{p}_2)^2}$ (allowing $c = 1$) reconstructed. The data can then be analyzed.

5 Observable, Data and Event Selection

We seek a simple jet observable to search for contact interactions. With this in mind, we describe the dijet centrality ratio. We define the inner region of the barrel calorimeter as having a pseudorapidity $|\eta| < \eta_{in}$, and the outer region of the barrel calorimeter as having a pseudorapidity $\eta_{in} < |\eta| < \eta_{out}$. We define inner dijets as events in which the two leading jets fall within the inner region of the barrel, and outer dijets as events in which the two leading jets fall within the outer region of the calorimeter. We define the dijet centrality ratio as the ratio of inner dijets to outer dijets:

$$R_\eta = \frac{N(|\eta| < 0.7)}{N(0.7 < |\eta| < 1.3)} \tag{12}$$

The choice of $\eta = 1.3$ as the outer boundary was motivated by the fact that this is the point at which the barrel calorimeter begins to transition into the endcap calorimeter, thus leading to additional instrumental systematic uncertainties. The inner boundary pseudorapidity of $\eta = 0.7$ was optimized in the following manner [22]. The test signal was taken to be a $\Lambda = 5$ TeV contact interaction in 100 pb^{-1} of LHC design energy data. The test observable is a simple statistical χ^2 between QCD and QCD plus the contact interaction. The results are shown in table 1. As expected, the largest barrel-only outer boundary value of $\eta_{out} = 1.3$ (corresponding to maximal barrel acceptance) is the optimal choice. $\eta_{in} = 0.7$ is seen to be the optimal inner boundary value.

This choice of observable results in a wide variety of sensitivities, and many systematics cancel out in the ratio. R_η can then be studied as a function of

η_{in},η_{out}	0.9	1.0	1.1	1.2	1.3
0.3	4.6	9.8	19.8	32.0	44.9
0.4	7.0	16.6	34.5	56.3	80.6
0.5	9.1	20.4	55.1	91.6	128.9
0.6	9.1	21.9	63.6	129.6	182.3
0.7	4.2	13.7	54.8	116.1	**199.9**
0.8		12.7	50.1	101.8	170.8
0.9			35.7	86.4	145.3

Table 1: χ^2 between QCD and QCD $+ \Lambda = 5$ TeV contact interactions at 100 pb^{-1} for different η_{in} and η_{out} values; optimal value in bold

M_{jj}. The bin widths of the mass histograms which are divided to study the dijet centrality ratio are chosen to roughly correspond with the dijet mass resolution at each bin center (see table 2). These bins were determined in CMS's dijet mass spectrum study [6], which is run in parallel to the dijet centrality ratio analysis. The first mass bin at 156 GeV is selected because this is where the lowest threshold trigger stream becomes fully efficient. In practice, the data is taken from a collection of six single jet triggers, having transverse momentum thresholds of 15, 30, 50, 70, 100 and 140 GeV. Each trigger covers the mass range from the point at which it becomes fully efficient to the point at which the next highest stream becomes fully efficient. The six streams are combined as such to form the full data set.

5.1 Simulation

It is necessary to have an effective simulation of both background and signal processes in order to effectively study one's data and search it for evidence of signals (or to rule out signals). To this end we utilize Monte Carlo simulations,

lower edge (GeV)	upper edge (GeV)	lower edge (GeV)	upper edge (GeV)
156	176	1181	1246
176	197	1246	1313
197	220	1313	1383
220	**244**	1383	1455
244	270	1455	1530
270	296	1530	1607
296	325	1607	1687
325	**354**	1687	1770
354	386	1770	1856
386	419	1856	1945
419	**453**	1945	2037
453	489	2037	2132
489	526	2132	2332
526	565	2332	2438
565	606	2438	2546
606	649	2546	2659
649	**693**	2659	2775
693	740	2775	2895
740	**788**	2895	3019
788	838	3019	3147
838	890	3147	3279
890	944	3279	3416
944	1000	3416	3558
1000	1058	3557	3704
1058	1118	3704	3854
1118	1181	3854	4010

Table 2: Mass bins used in study; trigger turnover points in bold

so named because events are seeded randomly (within the appropriate parameter space). The matrix elements necessary to produce the results of scattering interactions are calculated in perturbative QCD. Following the simulation of the initial scattering interactions, the detector response to the resulting propagating particles must be simulated as well. This is done as part of the CMSSW software framework utilized by CMS, the relevant part of the framework being based on the GEANT4 simulation package [11].

For the QCD background, two slightly different sets of background simulations are used - one for generalized data quality checks and one for the specific observable being examined. For the data quality checks, the PYTHIA leading-order calculation and Monte Carlo simulation software package [28] and the full CMSSW detector simulation are utilized. For the observable, the Herwig++ leading-order [18] and NLOJet++ next-to-leading-order [24, 25] packages are utilized as well. Its application is discussed in section 6.1.1. For the simulation of the contact interaction signal, only PYTHIA and the associated pre-detector data is utilized. This is justified in section 6.1.2.

5.2   Data
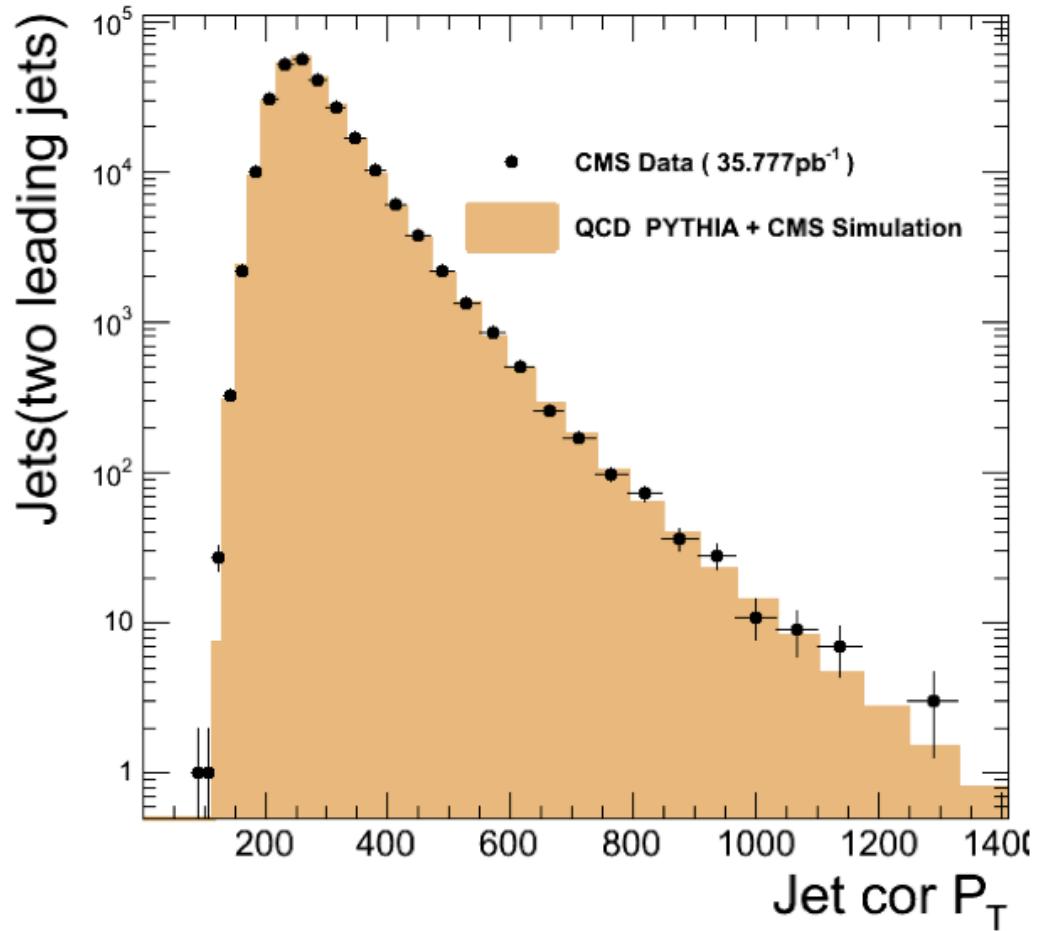
Figure 28: Comparison of data to normalized simulation of p_T of two leading jets; agreement verifies understanding of jet energy resolution and jet energy scale [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
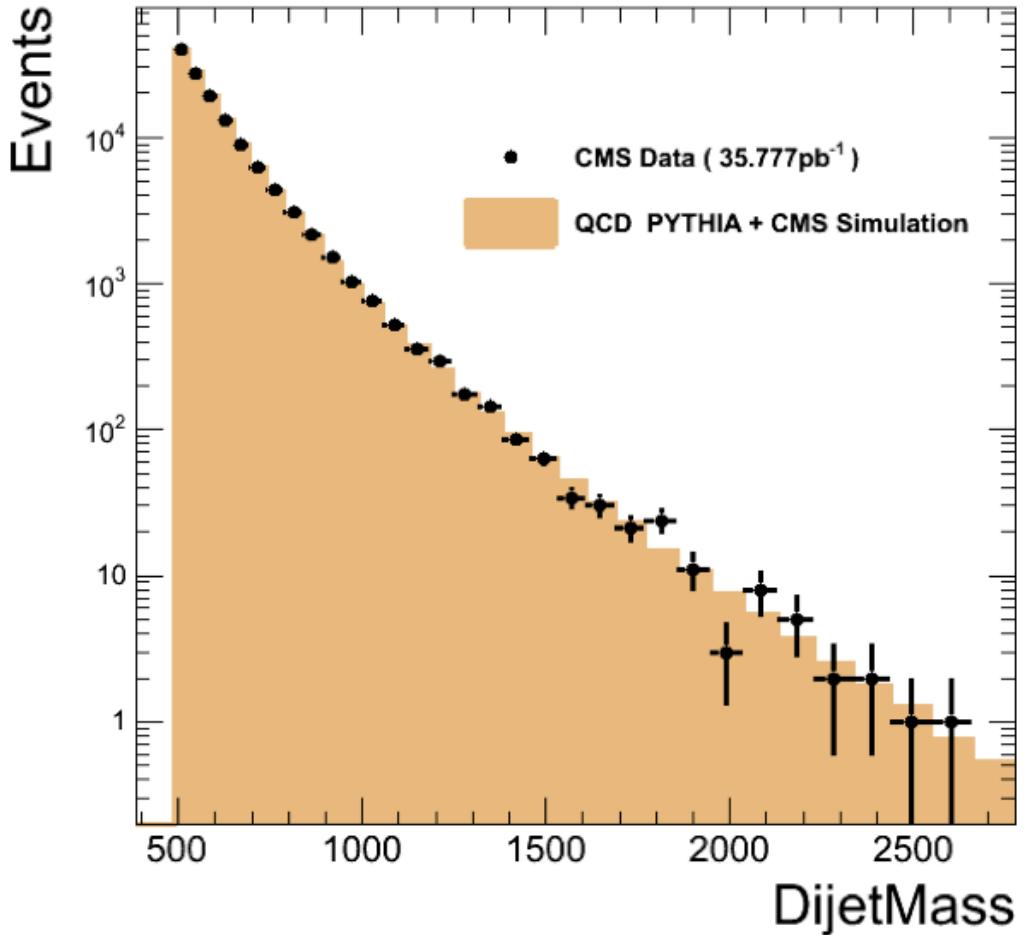
Figure 29: Comparison of data to normalized simulation of dijet invariant mass; agreement here verifies understanding of jet energy resolution and jet energy scale [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
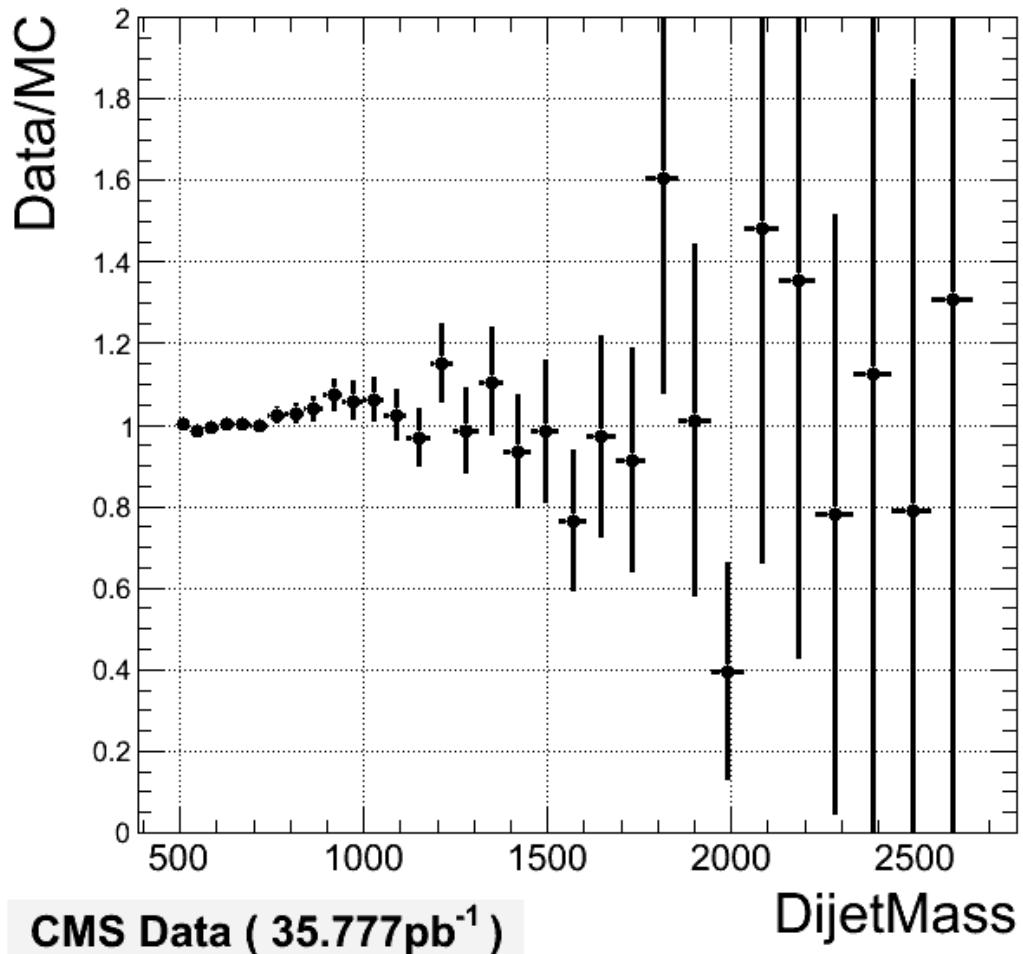
Figure 30: Pull in comparison of data to normalized simulation of dijet invariant mass; agreement here verifies understanding of jet energy resolution and jet energy scale [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
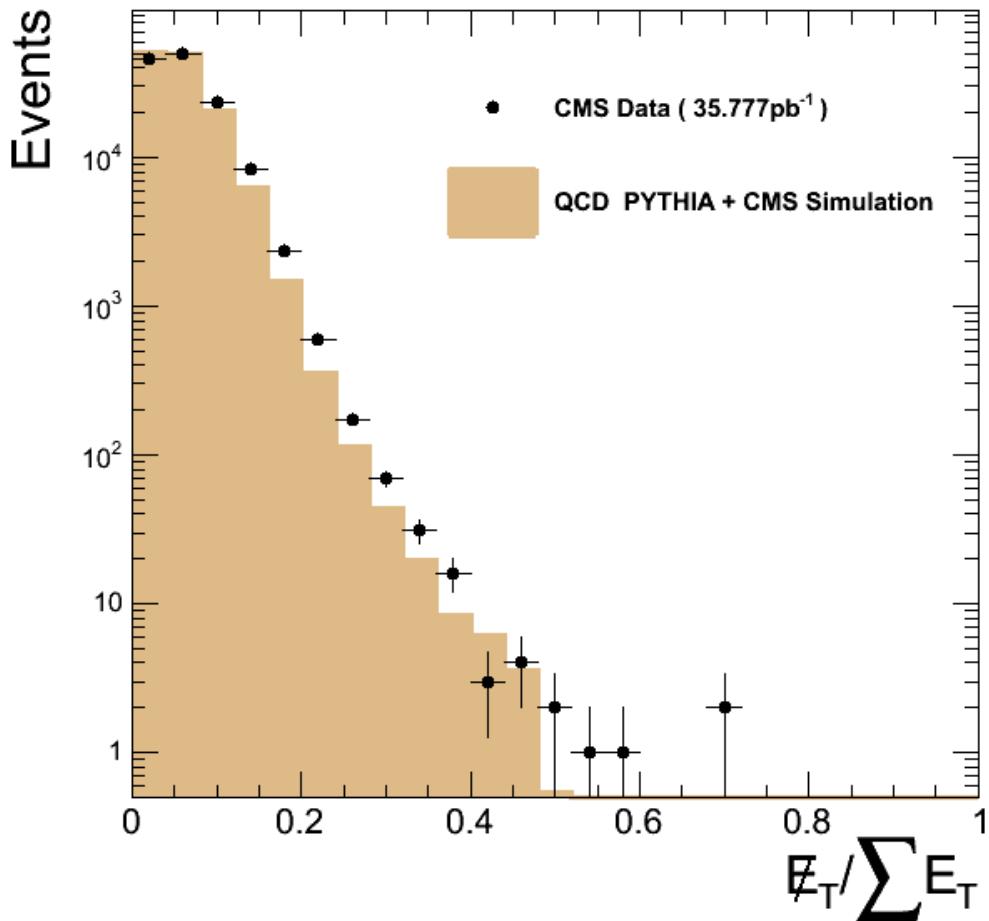
Figure 31: Comparison of data to normalized simulation of $\not{E}_T/\Sigma E_T$; agreement here verifies detector hermiticity [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
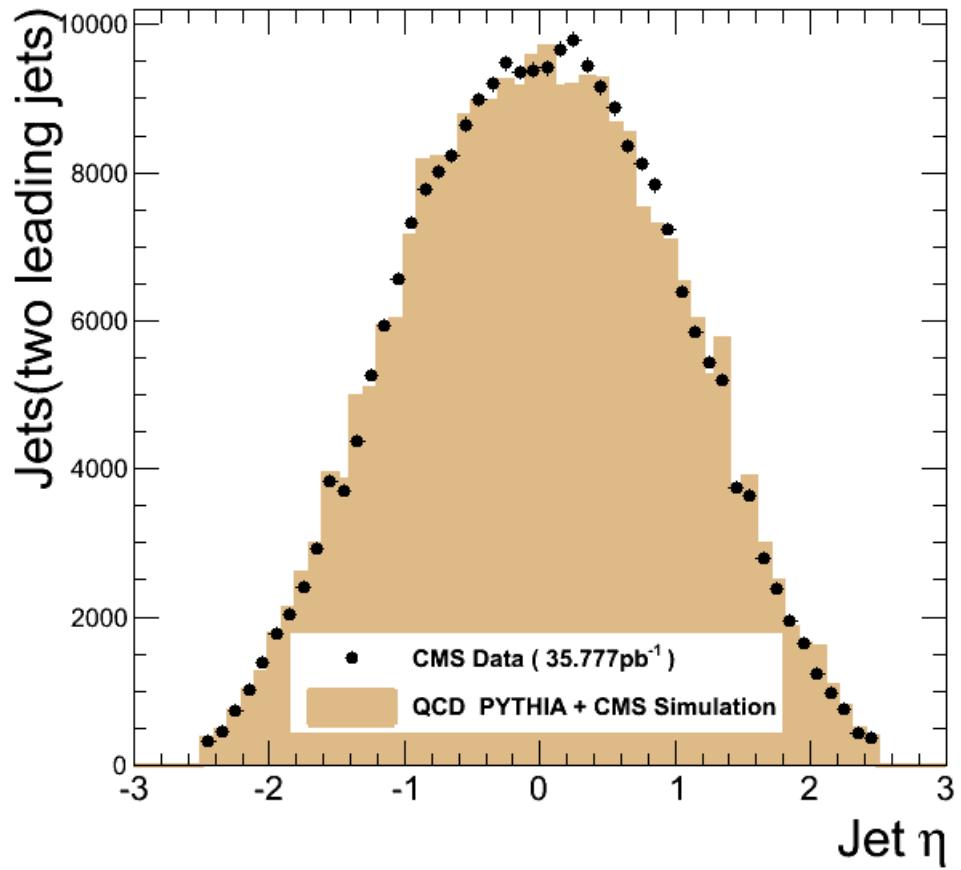
Figure 32: Comparison of data to normalized simulation of η of two leading jets; agreement here verifies uniformity of detector response in η [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
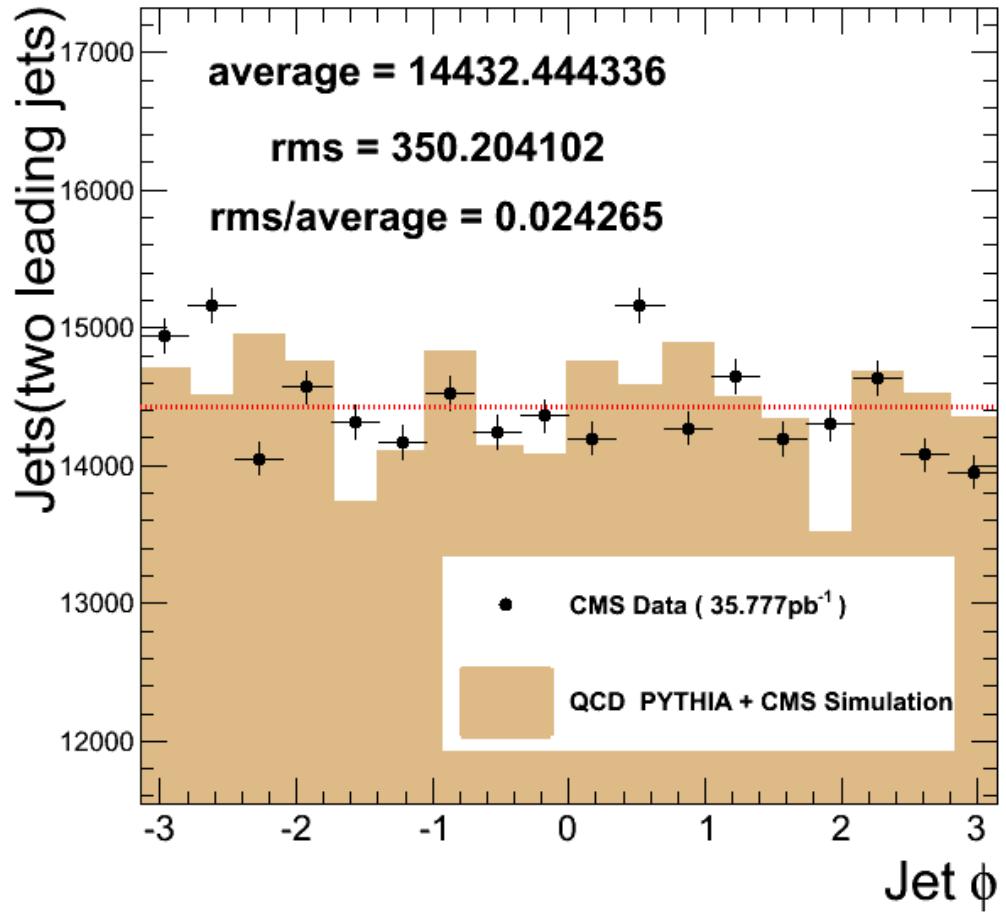
Figure 33: Comparison of data to normalized simulation of ϕ of two leading jets; agreement here verifies uniformity of detector response in ϕ [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
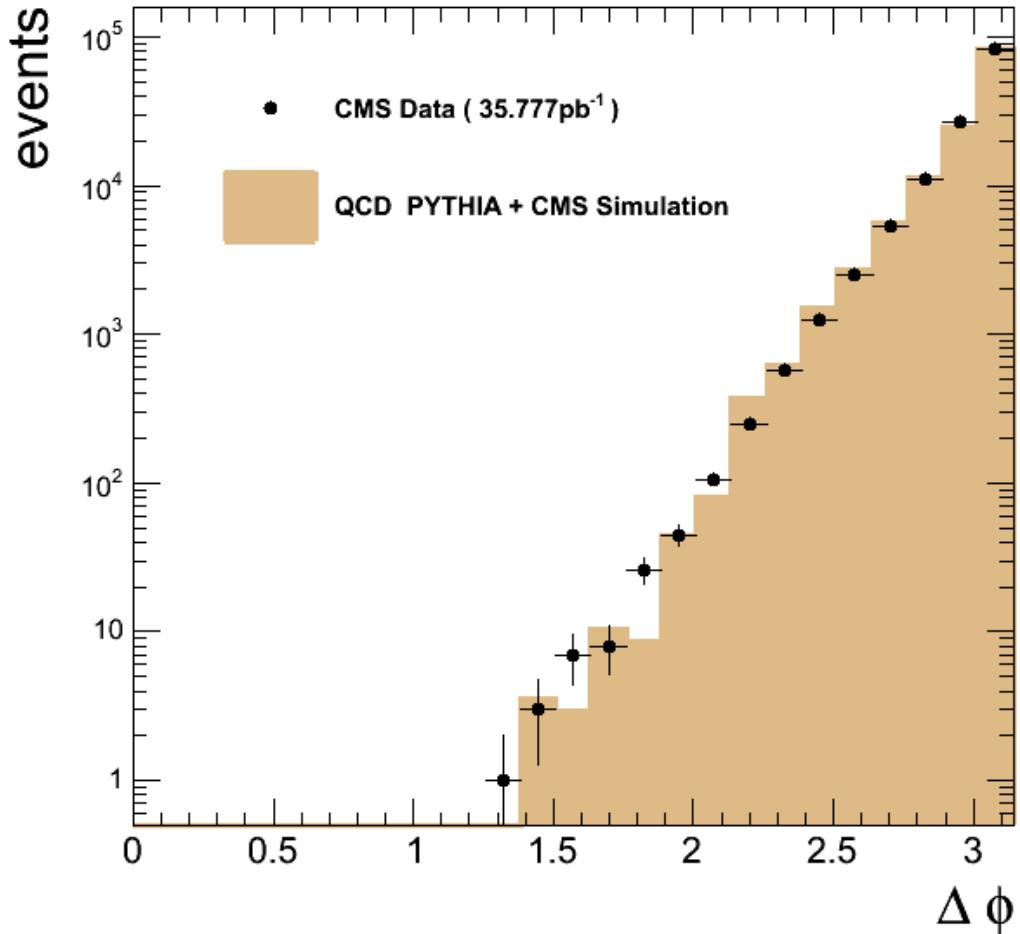
Figure 34: Comparison of data to normalized simulation in $\Delta\phi$ of two leading jets; agreement here verifies balance and uniformity of detector response in ϕ [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
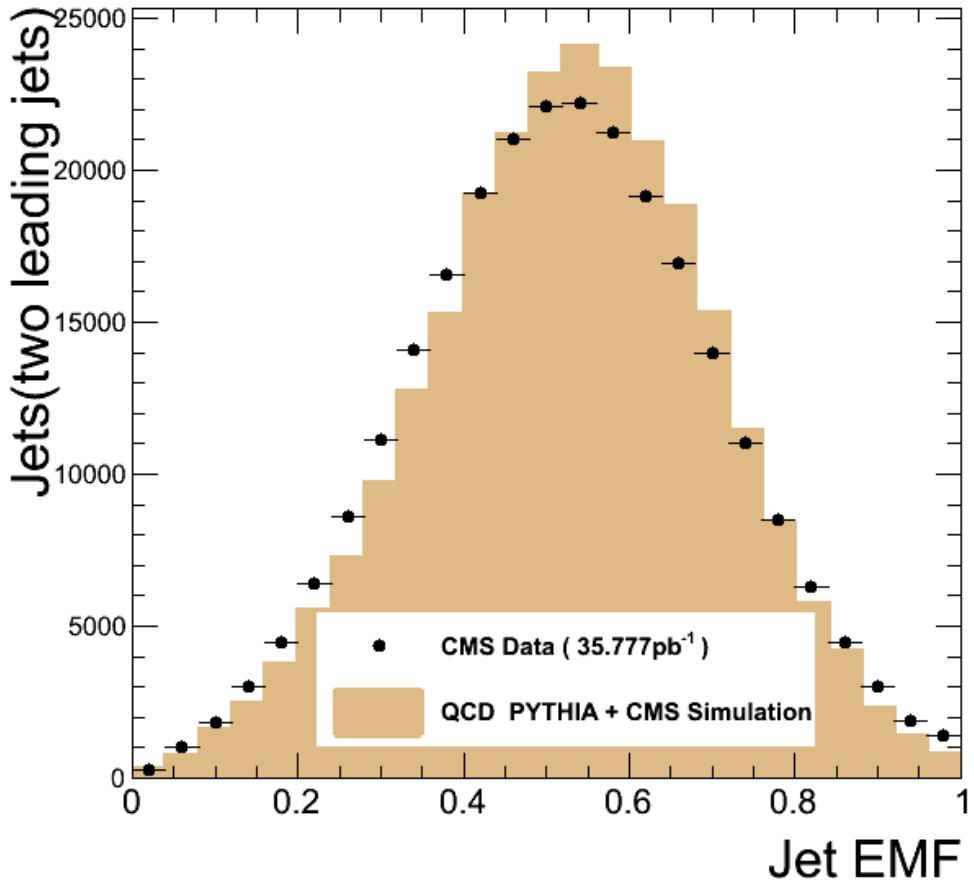
Figure 35: Comparison of data to normalized simulation of fraction of electromagnetic energy in two leading jets; agreement here verifies understanding of calorimeter functioning [CMS Internal, Dijet Resonance Search and Dijet Centrality Ratio groups]
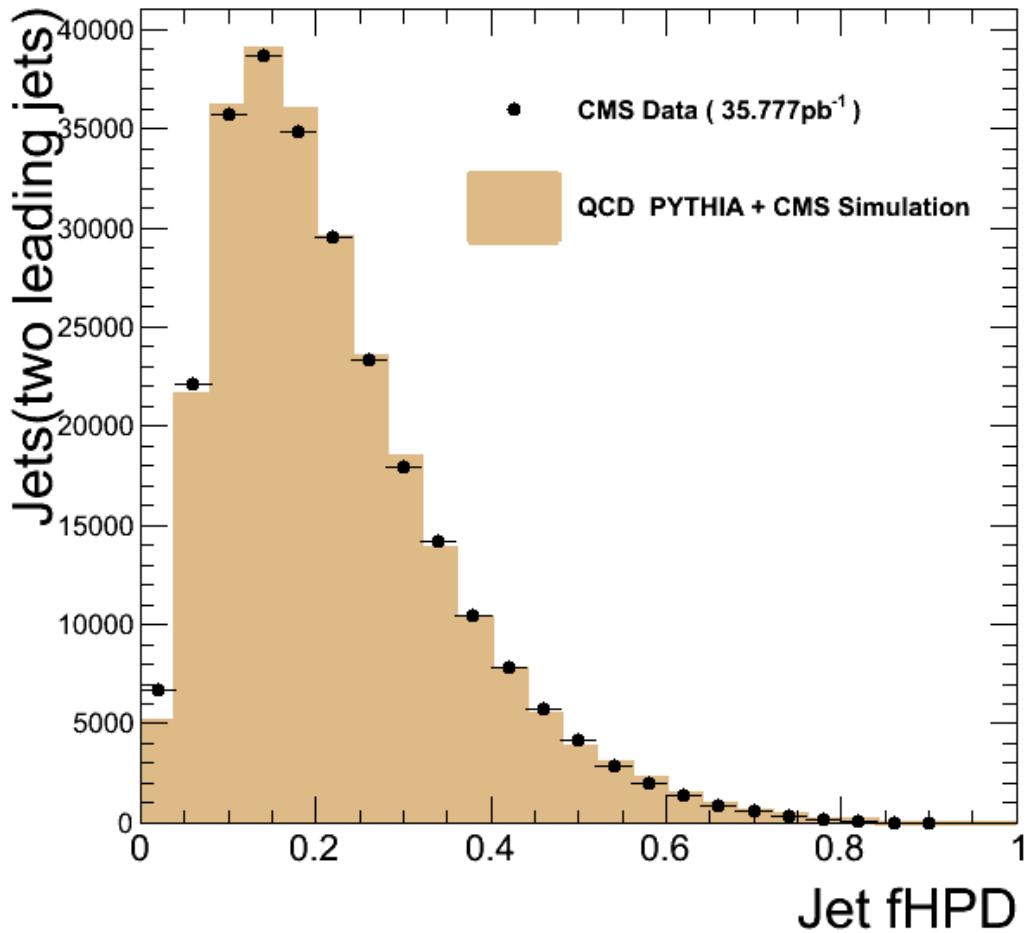
Figure 36: Comparison of data to normalized simulation of fraction of hybrid photodiodes fired in two leading jets; agreement here verifies understanding of hadronic calorimeter functioning [Dijet Resonance Search and Dijet Centrality Ratio groups]
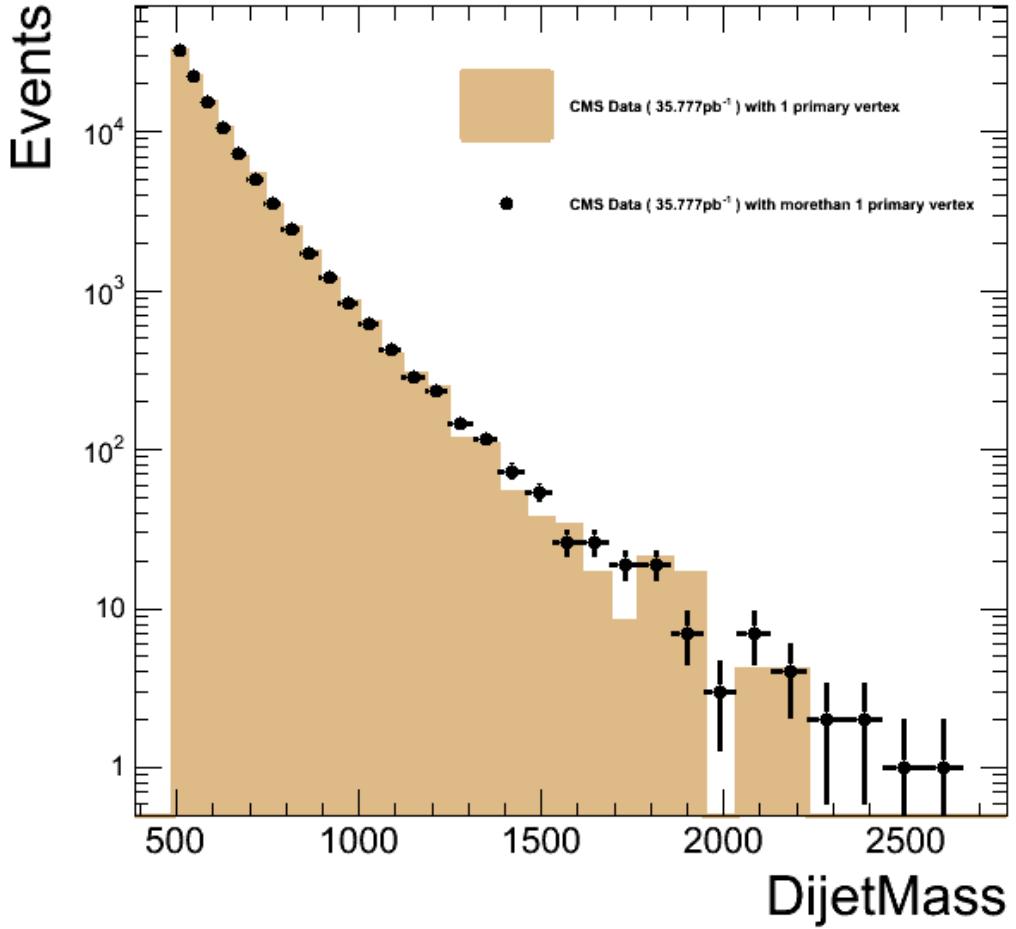
Figure 37: Comparison of data with single primary vertex to normalized data with multiple primary vertices of dijet invariant masse; agreement here verifies that the effect of pileup is negligible for this dijet study [Dijet Resonance Search and Dijet Centrality Ratio groups]

Once the data set was prepared, the procedure established for the mass spectrum analysis [19] is followed initially and detailed data quality checks were performed. These checks included several generalized comparisons to Monte

Carlo data predictions. First, the transverse momentum (p_T) spectrum (see figure 28) of the jets was compared with predictions, and found to be in good agreement. The dijet invariant mass spectrum (see figure 29) was found to be in good agreement as well. These two checks ensure that the energy resolution and scale are well-understood. Comparison of the data to the predicted fraction of missing transverse energy (out of total transverse energy) ($\not{E}_T/\Sigma E_T$) (see figure 31) ensures (in the absence of excessive amounts of new physics which could distort the results) the reliable hermeticity of the detector. The η and ϕ (see figures 32, 33, 34) jet distributions are checked, ensuring the uniformity of the detector response. Additional jet properties (the fraction of electromagnetic energy in the jet and the fraction of HPDs fired, see figures 35, 36) are also checked and serve to confirm that the calorimeters are functioning as expected. Finally, to show that the effect of pileup is negligible in this case, the mass spectrum of events with a single primary vertex was compared to the normalized mass spectrum of events with one or more primary vertex (see figure 37), demonstrating excellent agreement.

6 Predictions and Limit-Setting Methodology

6.1 Predictions

In order to evaluate our data, we must generate simulations of both our null hypothesis (QCD) and our signal hypothesis (contact interactions at various energy scales). Simulated datasets are generated using Monte Carlo methods. To generate our QCD predictions, we use PYTHIA plus a full detector simula-

tion and NLOJet++ [24, 25, 28]. Additional QCD studies are carried out using PYTHIA without the full detector simulation and using Herwig++ [18, 28]. The contact interaction signals are generated using PYTHIA without a full detector simulation. We justify this by comparing the dijet centrality ratio for a sample of fully simulated jets (corrected calojets, i.e. calojets adjusted for discrepancies in the absolute and relative (η-dependent) jet energy scale calibrations) and generator-level jets (genjets) (see figures 38 and 39). We find the difference between the two to be mass-independent and generally negligible.
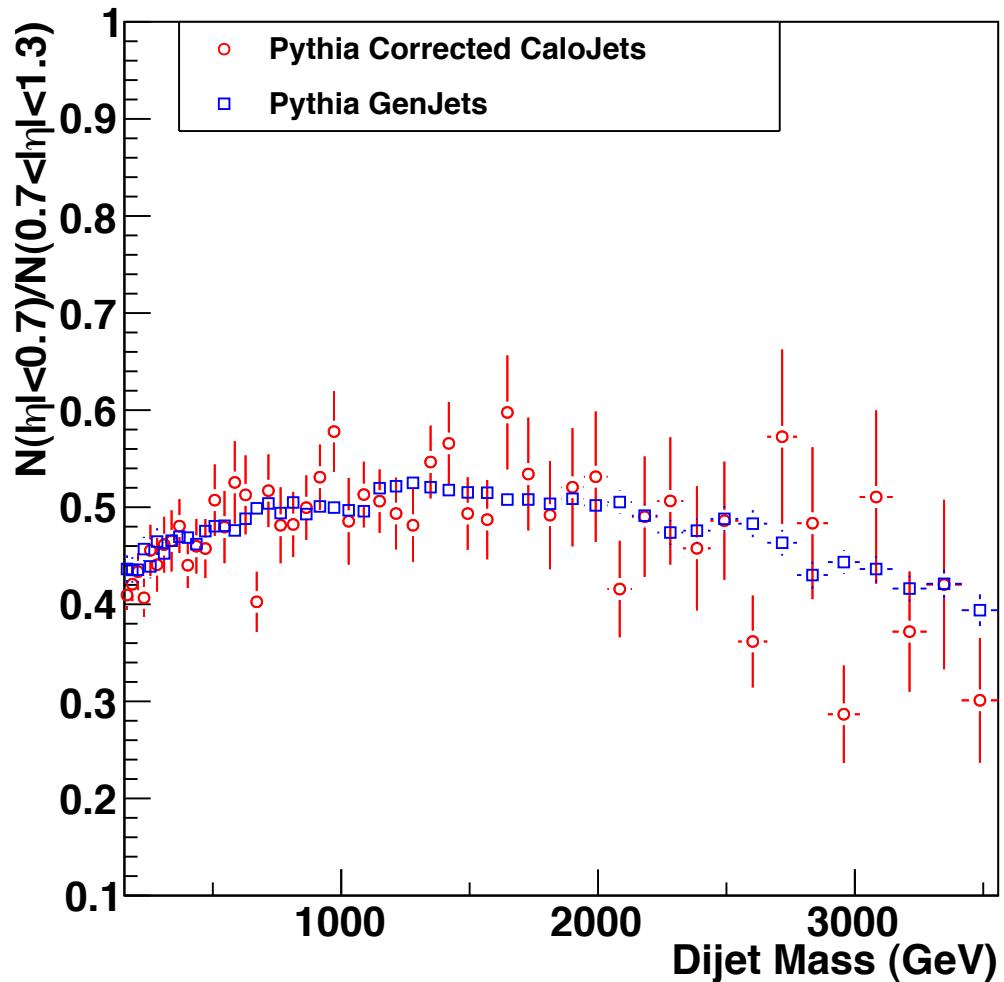
Figure 38: Comparison of generator level simulated jets and corrected calorimetry level simulated jets in the dijet centrality ratio [author-generated]
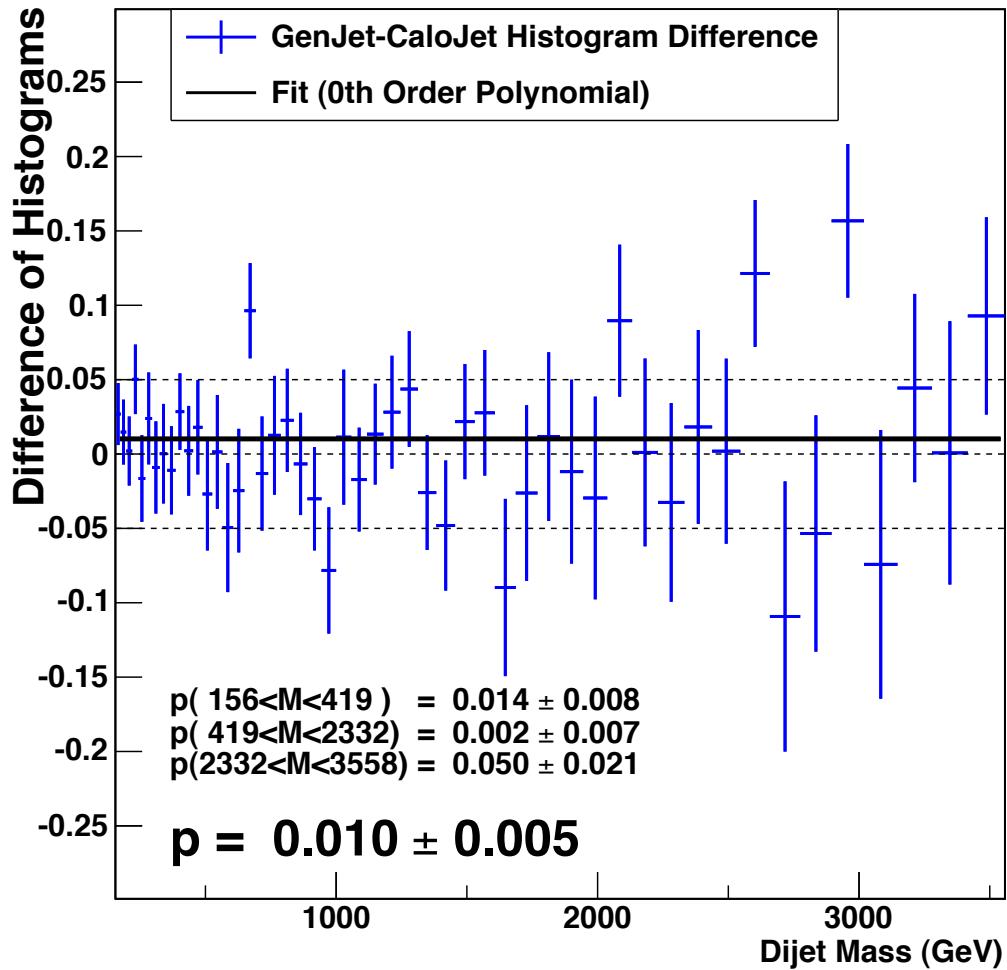
Figure 39: Difference between generator level simulated jets and corrected calorimetry level simulated jets in the dijet centrality ratio; fit close to zero shows negligibility of difference [author-generated]

6.1.1   QCD Predictions
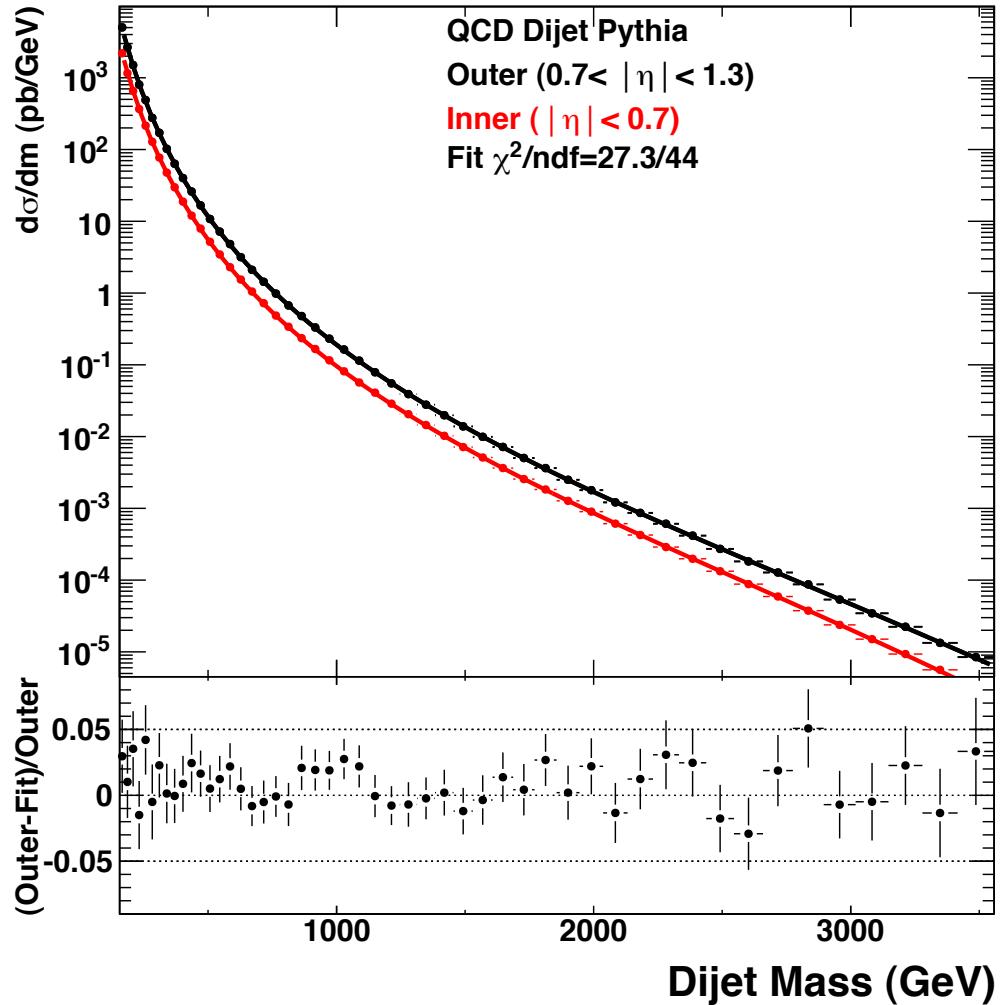
Figure 40: PYTHIA prediction for inner and outer dijet mass spectra; the bottom plot demonstrates the quality of the fit to the outer mass spectrum [Dijet Centrality Ratio group]
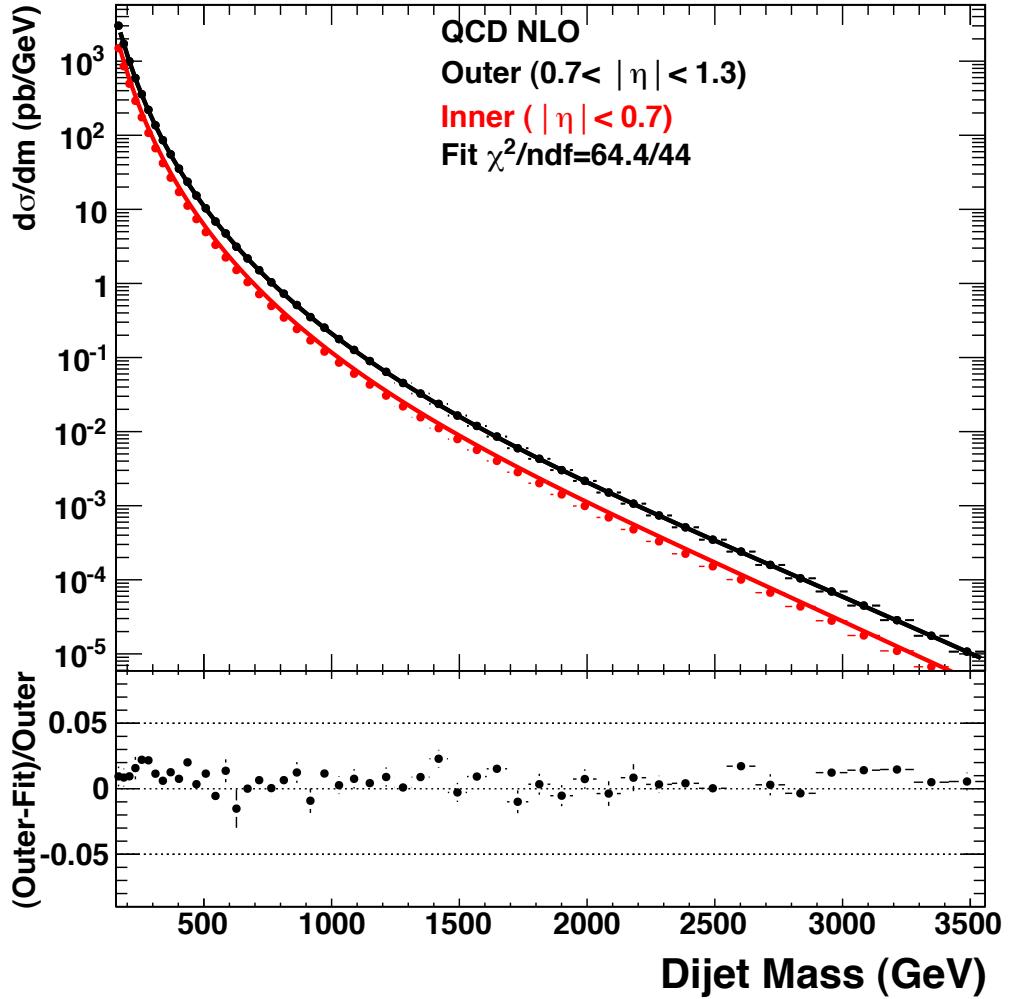
Figure 41: NLO prediction for inner and outer dijet mass spectra; the bottom plot demonstrates the quality of the fit to the outer mass spectrum (later used in the limit-setting procedure) [Dijet Centrality Ratio group]

We use a large PYTHIA plus detector simulation dataset to generate idealized inner and outer mass spectra for QCD (see figures 40 and 41 (NLO-corrected)). To generate a smooth functional form for the limit-setting procedure (described

in section 6.2), we fit the outer mass spectrum to the following analytical form (where the variable M_{jj} is the dijet invariant mass and p_n are the fit parameters):

$$f(M_{jj}) = p_0(1 - \frac{M_{jj}}{\sqrt{s}} - p_1\frac{M_{jj}^2}{s})^{p_3}(\frac{M_{jj}}{\sqrt{s}})^{-p_4} \tag{13}$$
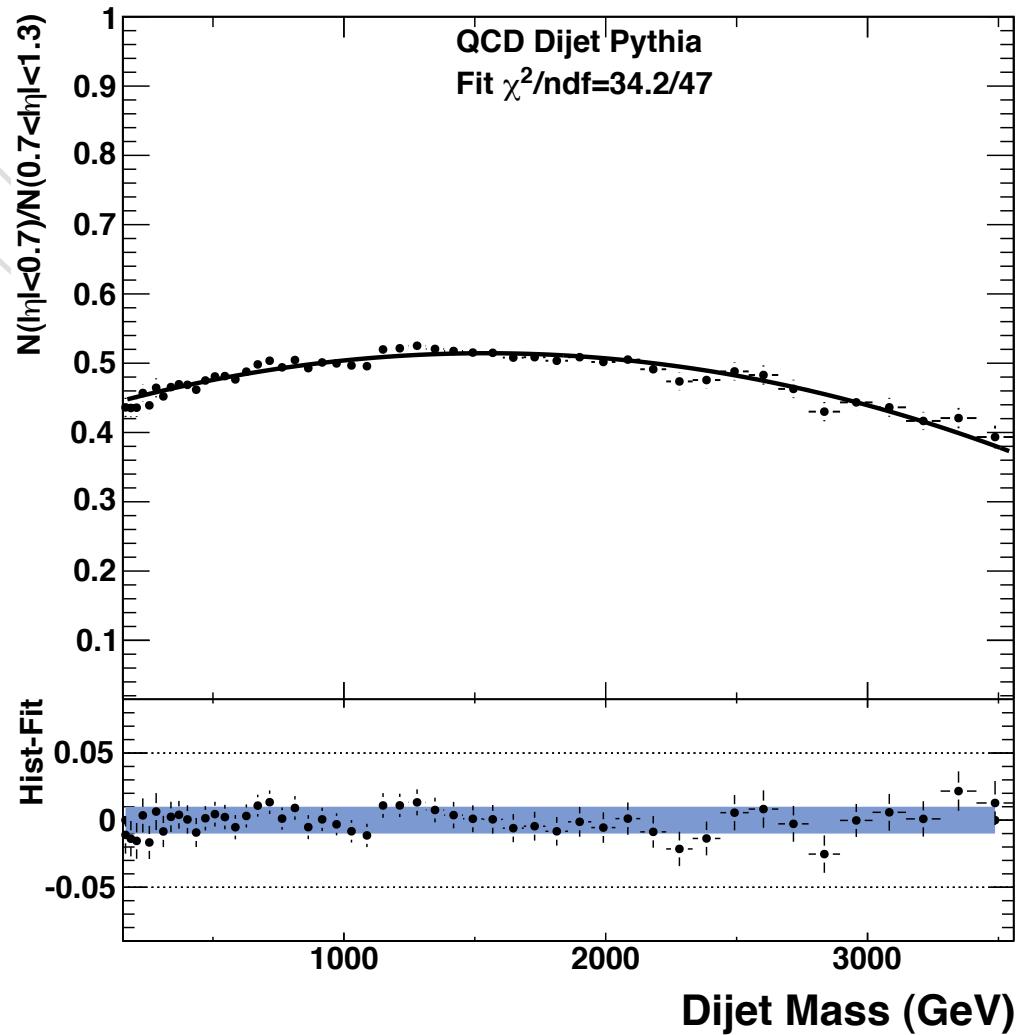
Figure 42: PYTHIA prediction for dijet centrality ratio; the bottom plot demonstrates the quality of the fit to the dijet centrality ratio, blue band is fit uncertainty [Dijet Centrality Ratio group]
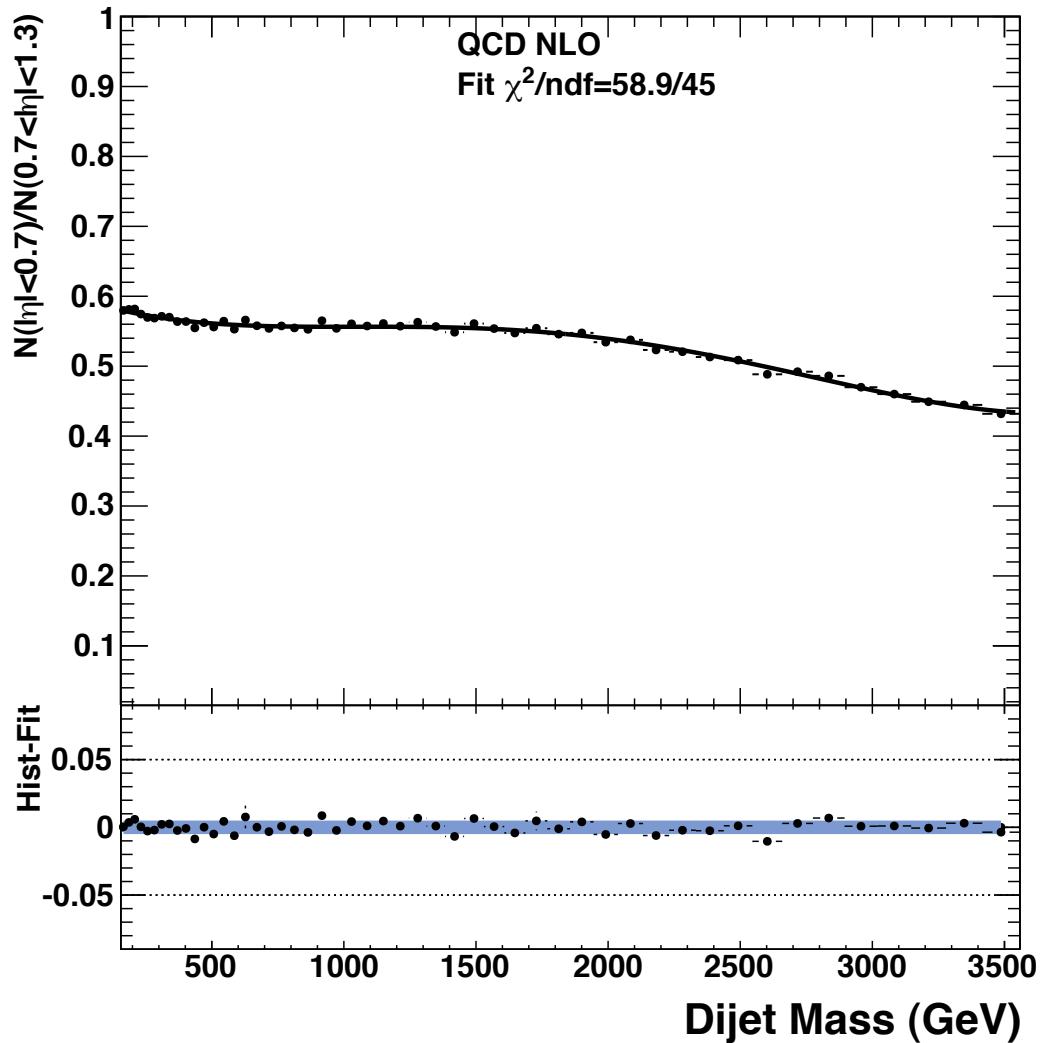
Figure 43: NLO Prediction for Dijet Centrality Ratio

NLO prediction for dijet centrality; the bottom plot demonstrates the quality
of the fit to the dijet centrality ratio (later used in the limit-setting
procedure), blue band is fit uncertainty [Dijet Centrality Ratio group]

The dijet centrality ratio for QCD is ~ 0.5 and is fit to a quartic function.

The functional form of the inner mass spectrum is then taken to be the product of the outer mass fit and the quartic fit (see figures 42 and 43; the blue band is the fit uncertainty). Only the dijet centrality ratio curve is used in the limit-setting procedure; the mass spectra are used only for obtaining the expected number of events for limit predictions.
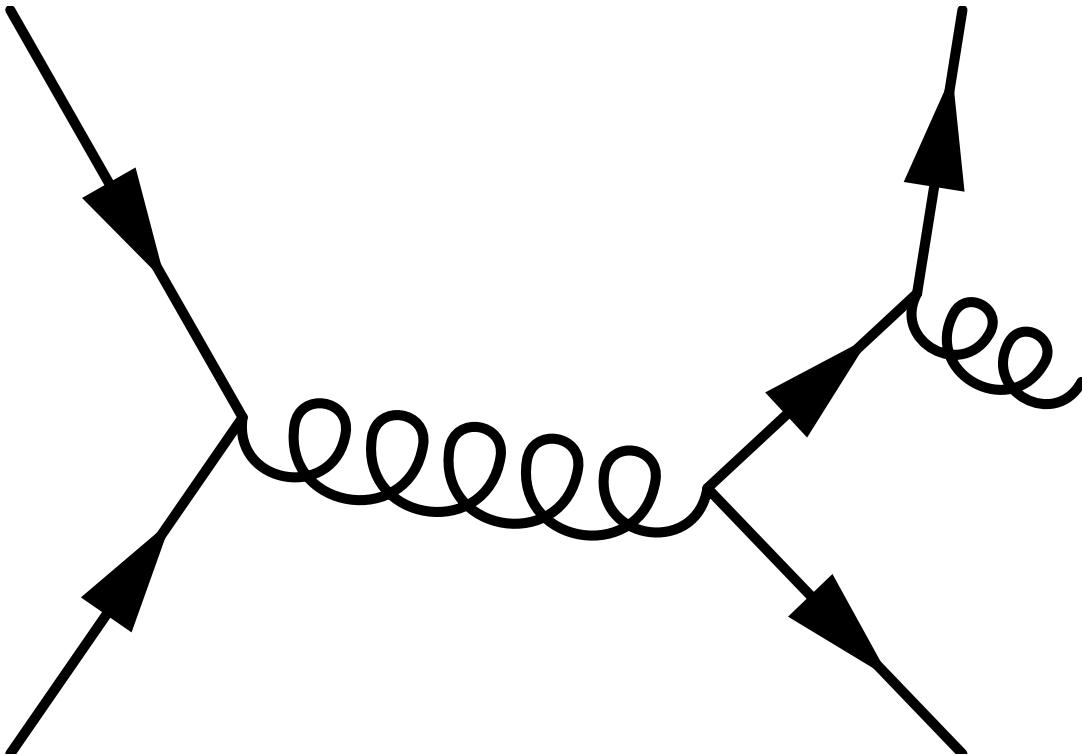
Figure 44: Example diagram of final state gluon radiation [author-generated]

In order to effectively understand any comparisons between our observable in data and Monte Carlo, we require a thorough examination of the various modes and models that we can use to generate our Monte Carlo data. Theses include Herwig++ samples, leading order (LO) jets from NLOJet++, next-to-leading order (NLO) jets from NLOJet++, and PYTHIA samples produced in

nominal, parton showering (PS), QCD and QED initial state radiation (ISR), final state radiation (FSR), multiple parton interaction (MPI), and hadronization (HAD) modes, and various combinations thereof. Parton showering is a simulation of the interactions of the the partons forming the jets (as opposed to simple reliance on the kinematics of the jet's seed particle to approximate the jet's properties). Initial and final state radiation involve the emission of a particle either immediately before or immediately after the primary interaction. This can be either electroweak or strong (i.e. gluon or photon) radiation (an example is shown in figure 44). Multiple parton interactions are softer interactions between the proton constituents not involved in the primary interactions. Hadronization was described in section 4, and here simply refers to a full and extensive simulation of this process.
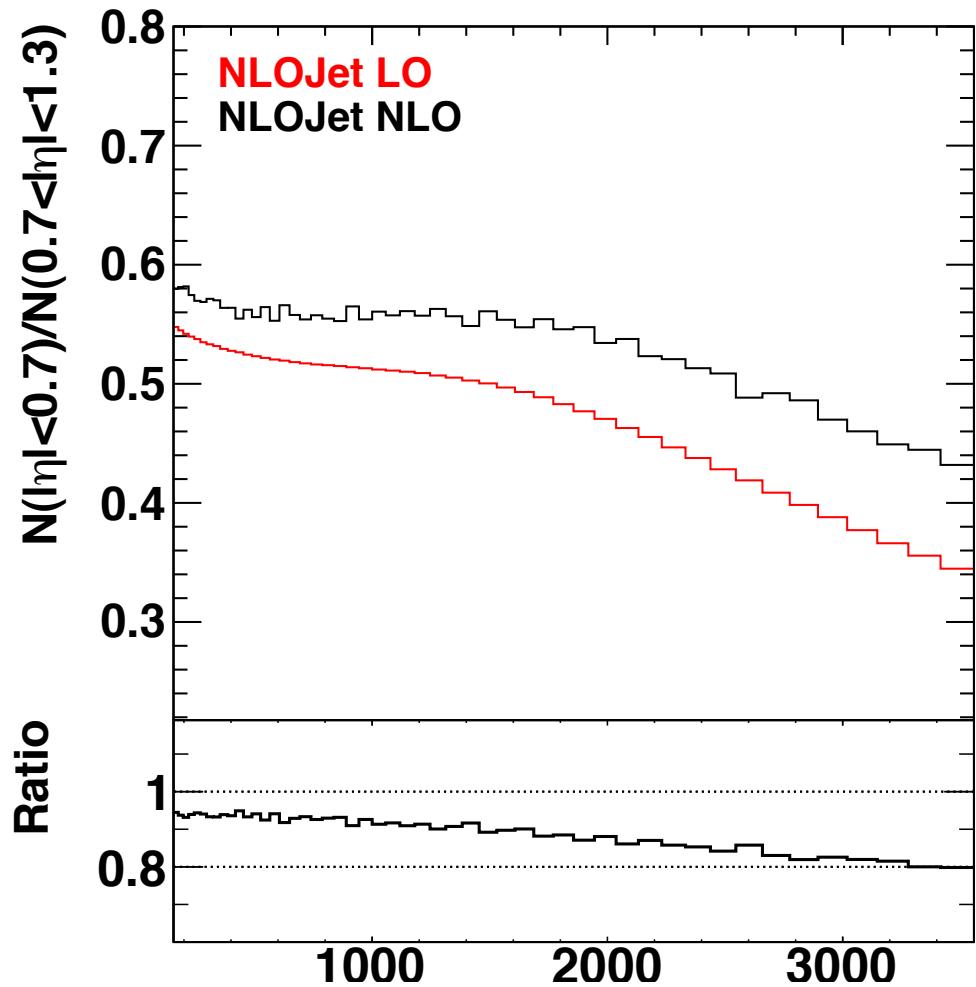
Figure 45: Dijet centrality ratio in NLOJet++ leading order simulation versus NLOJet++ next to leading order simulation; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the effect of next to leading order corrections [Dijet Centrality Ratio group]
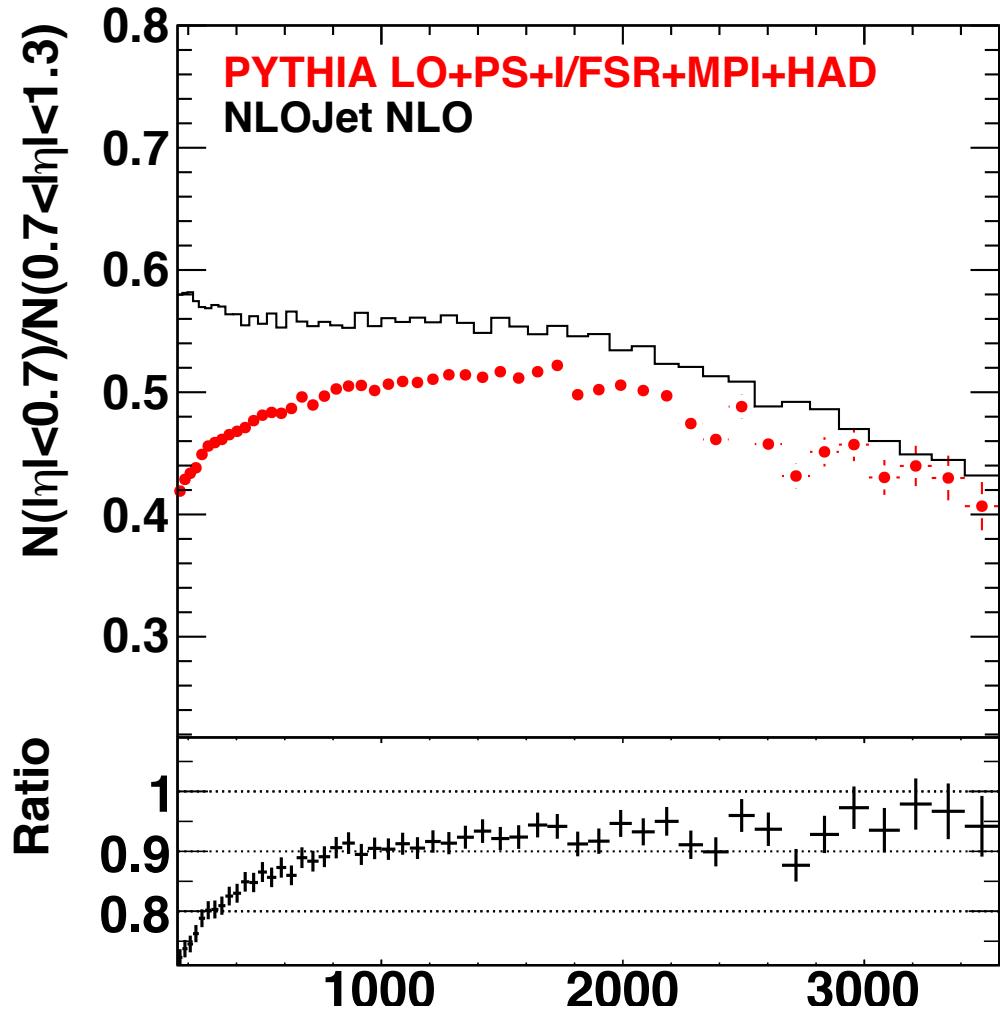
Figure 46: Dijet centrality ratio in PYTHIA leading order simulation with parton showering, initial state radiation, final state radiation, and multiple parton interaction versus NLOJet++ next to leading order simulation; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the effect of next to leading order corrections [Dijet Centrality Ratio group]
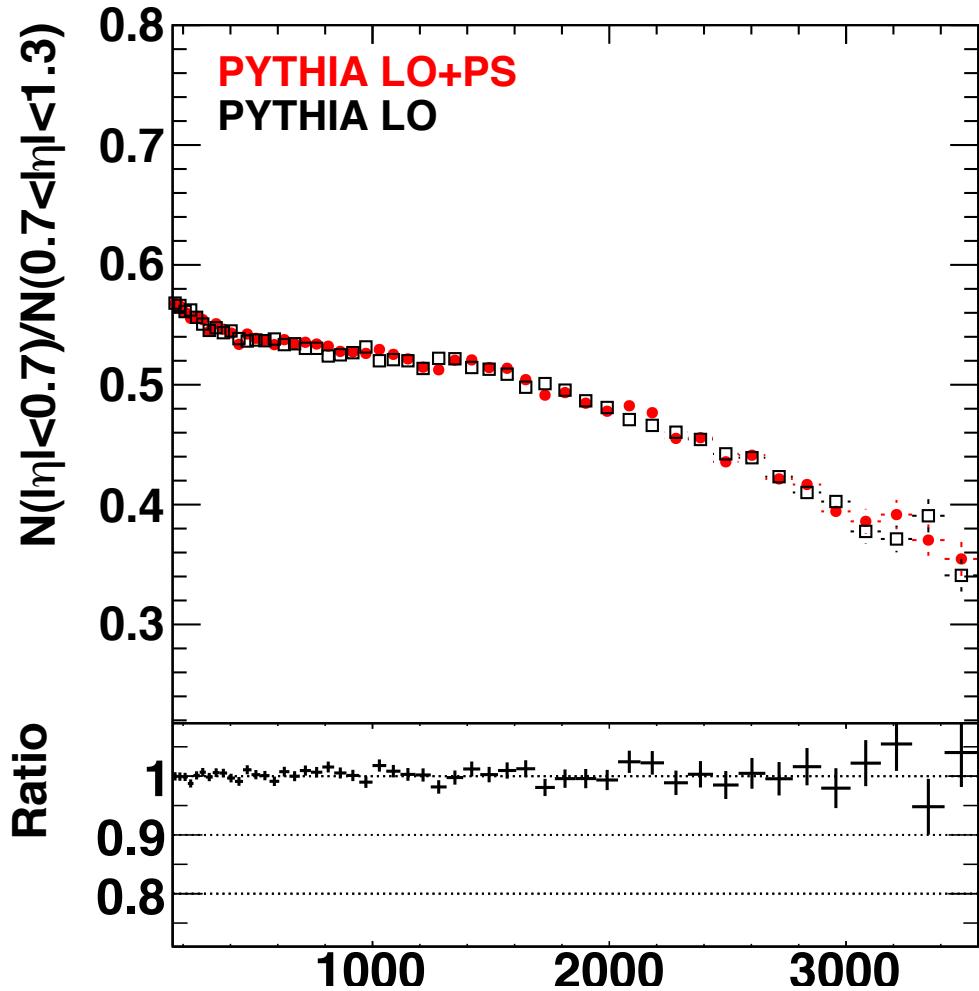
Figure 47: Dijet centrality ratio in PYTHIA leading order simulation with parton showering versus PYTHIA leading order simulation; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the negligible effect of parton showering [Dijet Centrality Ratio group]
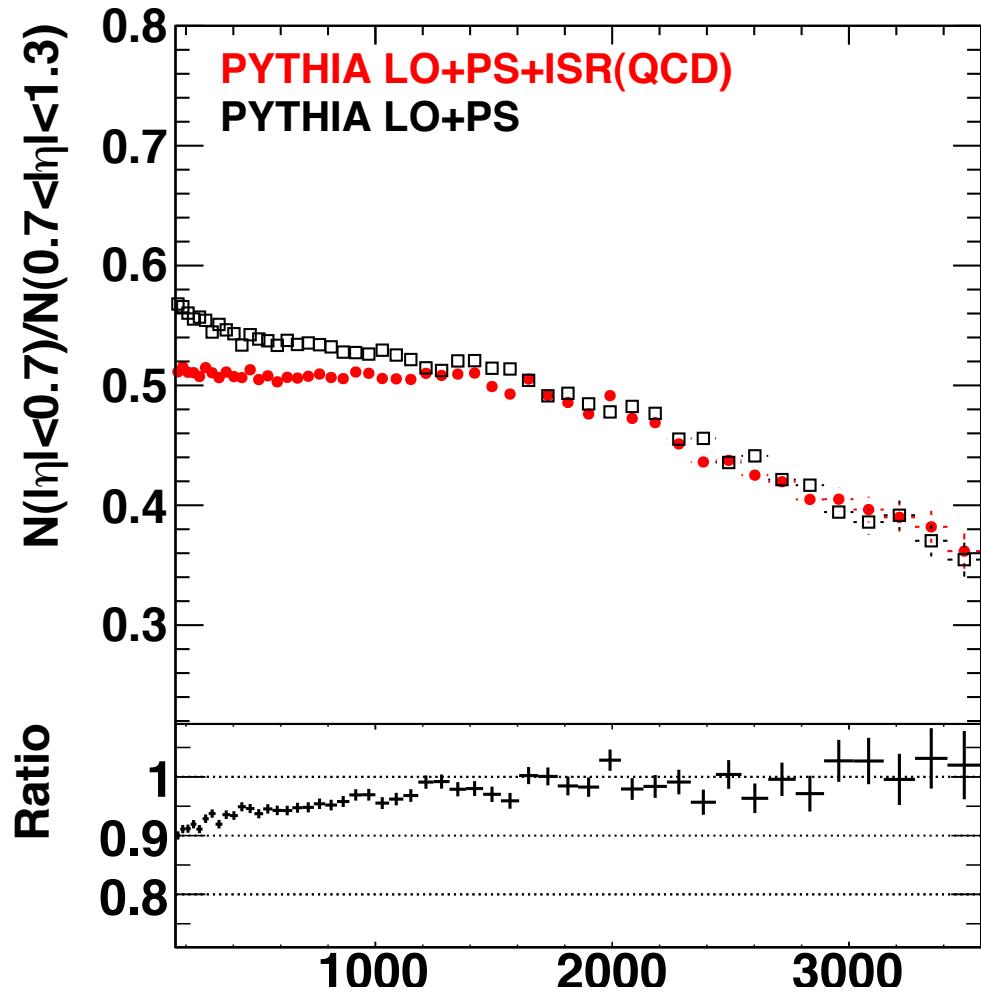
Figure 48: Dijet centrality ratio in PYTHIA leading order simulation with parton showering and quantum chromodynamic initial state radiation versus PYTHIA leading order simulation with parton showering; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the effect of quantum chromodynamic initial state radiation [Dijet Centrality Ratio group]
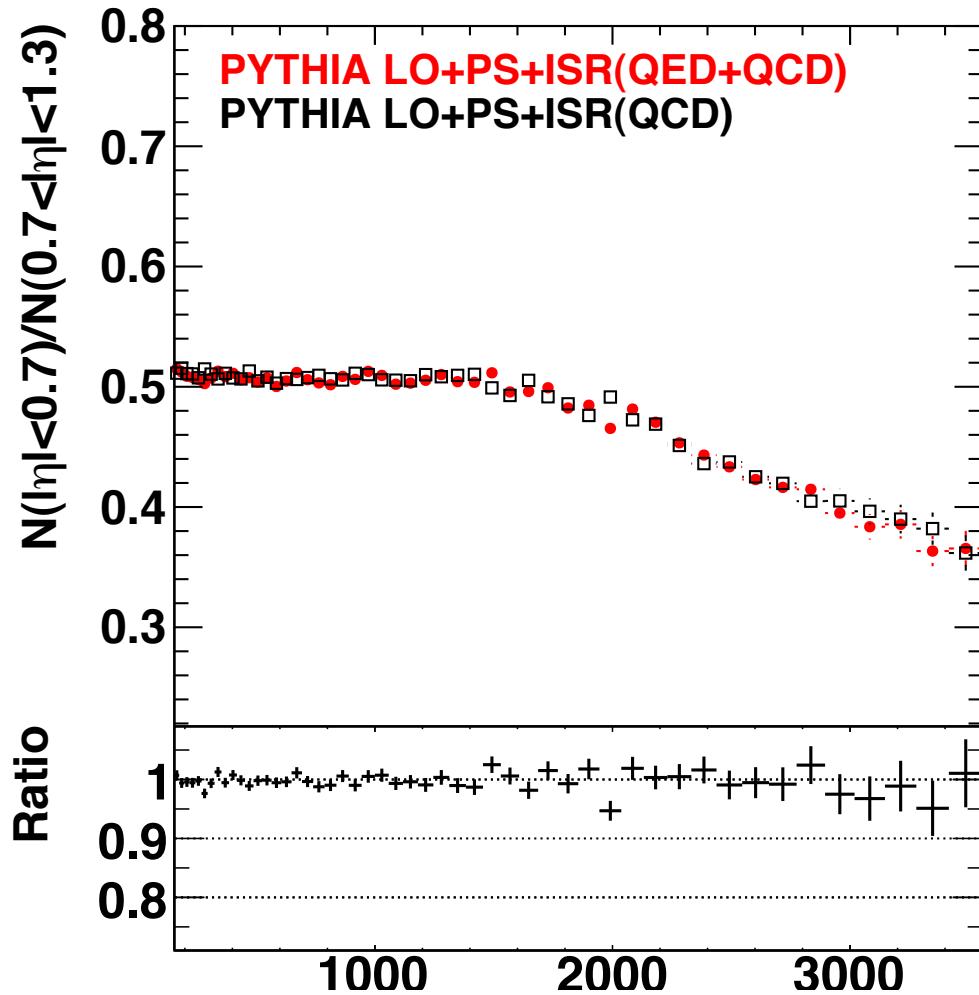
Figure 49: Dijet centrality ratio in PYTHIA leading order simulation with parton showering and quantum chromodynamic and electrodynamic initial state radiation versus PYTHIA leading order simulation with parton showering and quantum chromodynamic initial state radiation; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the negligible effect of quantum electrodynamic initial state radiation [Dijet Centrality Ratio group]

We begin by comparing the NLOJet++ LO, NLO, and PYTHIA LO with ISR, FSR, MPI and HAD predictions (see figures 45 and 46). The effect on the dijet centrality ratio from the difference between LO and NLO varies from 5% to 20% with the dijet invariant mass. The PYTHIA results with ISR, FSR, MPI and HAD processes come close to reproducing the NLO at high mass, but differ by as much as 30% at low mass. With this in mind we seek to determine which of the features of the full-featured PYTHIA results are necessary. We first examine parton showering (see figure 47) and see that its effects are negligible. Similarly, we see that while PYTHIA LO with PS and QCD ISR shows a 10% difference at low mass against PYTHIA LO with PS but no ISR (see figure 48), QED ISR has no notable effect (see figure 49). We follow this with a check of non-perturbative corrections - that is to say MPI and HAD corrections. We check this in both PYTHIA (see figure 50) and Herwig++ (see figure 51), fitting the correction to the sum of a constant and a Gaussian:

$$f_{corr}(m) = p_0 + \exp{\left(-\frac{1}{2}(\frac{m - p_1}{p_2})^2\right)} \tag{14}$$
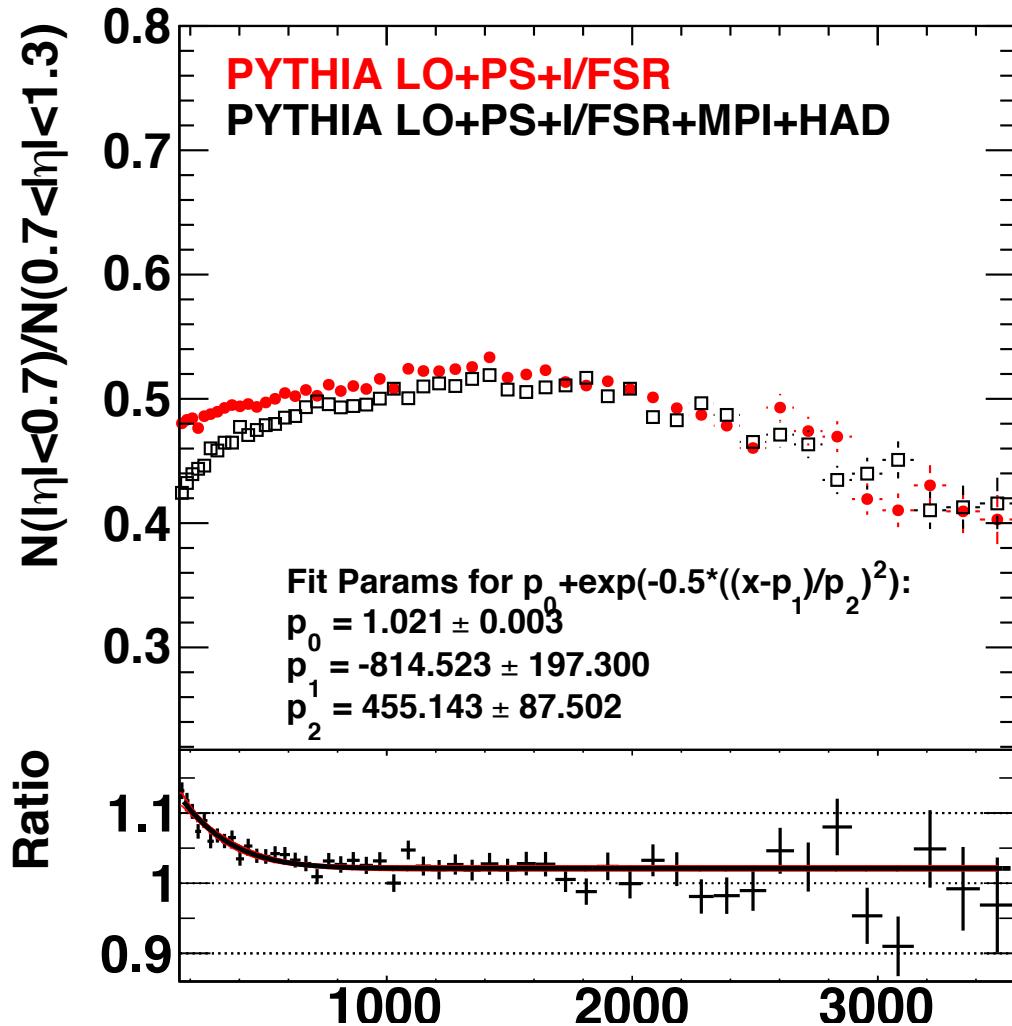
Figure 50: Dijet centrality ratio in PYTHIA leading order simulation with parton showering, initial state radiation and final state radiation versus PYTHIA leading order simulation with parton showering, initial state radiation, final state radiation, multiple parton interaction and hadronization; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the combined effect of multiple parton interaction and hadronization; the fit and fit parameters for the resultant corrections in PYTHIA are shown [Dijet Centrality Ratio group]
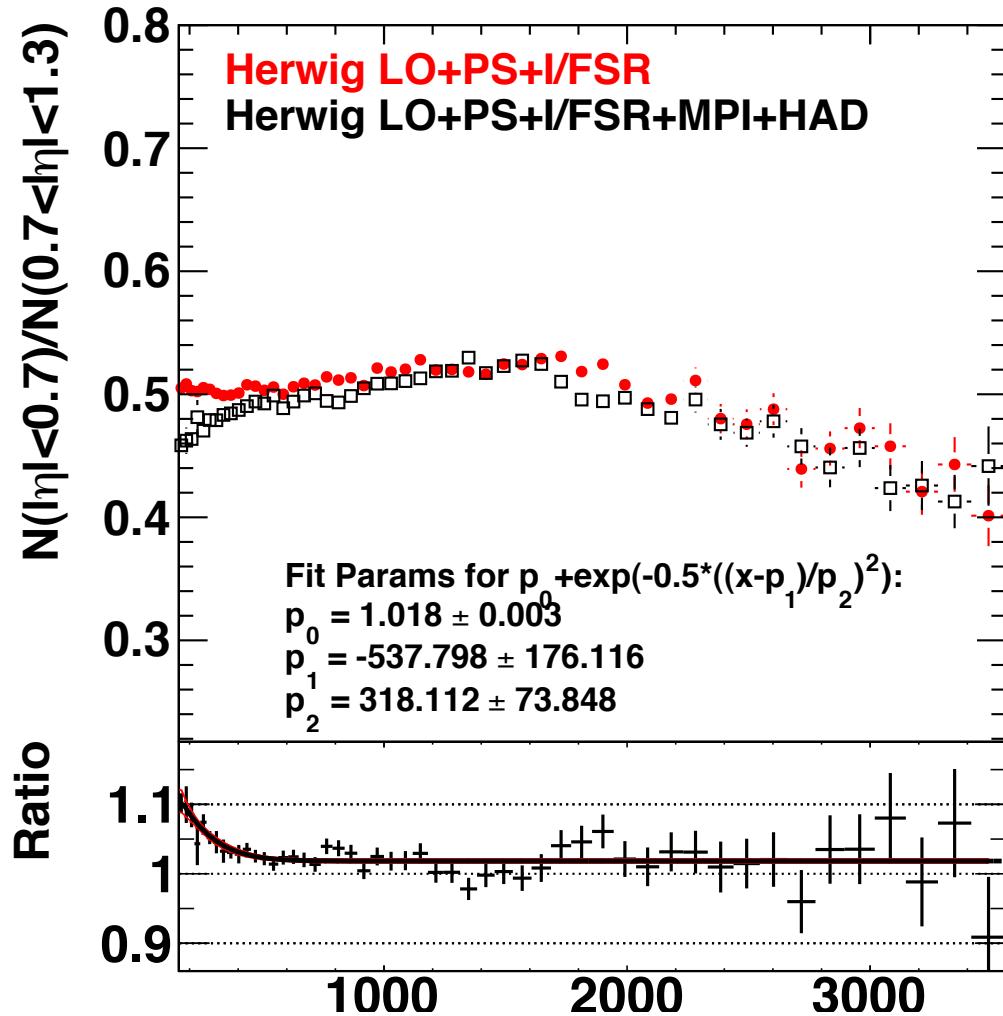
Figure 51: Dijet centrality ratio in Herwig++ leading order simulation with parton showering, initial state radiation and final state radiation versus Herwig++ leading order simulation with parton showering, initial state radiation, final state radiation, multiple parton interaction and hadronization; the bottom plot shows the agreement between the two simulation modes in ratio form; this demonstrates the combined effect of multiple parton interaction and hadronization; the fit and fit parameters for the resultant corrections in Herwig++ are shown [Dijet Centrality Ratio group]

Here, p_n are the fit parameters. Examining both the PYTHIA and the Herwig++ corrections (see figure 52), both of which vary from 10% to 2% with dijet mass, we see that they agree within 1.5%. We choose to assign a general systematic uncertainty of 20% of the correction to remain conservative. We choose the correction from PYTHIA to apply to our NLO model for use in our new physics search.
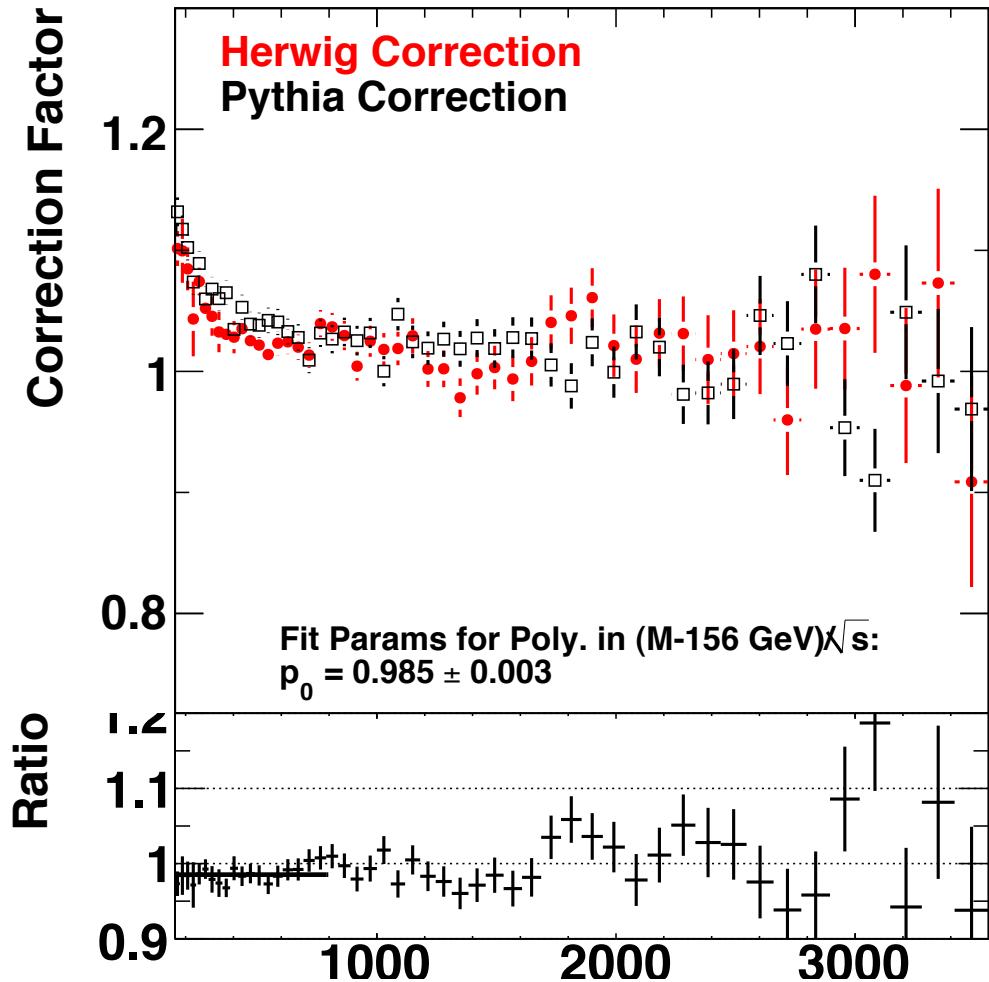
Figure 52: Non-perturbative correction in Herwig++ versus non-perturbative correction in PYTHIA; the bottom plot shows the agreement between the two corrections in ratio form; this demonstrates the comparable parametrization in each mode; the agreement-demonstrating fit and fit parameters are shown [Dijet Centrality Ratio group]

In the predictions used in our limit-setting procedure, we take as the QCD

model the shape of the NLO dijet centrality ratio prediction and apply the non-perturbative corrections derived from PYTHIA. Additionally, we allow for an overall offset of the shape which we fit within a selected normalization region of the data. The region is well-determined by Tevatron experiments to agree with QCD, and several factors go into the specific choice of its mass range. While, in general, more bins lead to a lower statistical uncertainty in the offset, allowing the lower end of the range to drop too low leads to higher systematic uncertainties due to the size of the non-perturbative corrections at low mass. It should also be noted that regardless of corrections normalizing at higher mass inevitably yields more conservative limits. This must, of course, conversely, be limited in order to maintain experimental sensitivity. Given that the statistical uncertainties are relatively small compared to the non-perturbative correction uncertainties at low mass, we prefer for the majority of the total uncertainty to be statistical. To this end, following an evaluation of the fractional uncertainties (see figures 53 and 54), we select a normalization region between bin numbers 13 and 19, which corresponds to a mass window of 489 - 788 GeV. Within this range corrections will range from 3.8% to 2.3%, with a minimum correction of 2.1% regardless of mass. Aditionally, after fitting, we take the PYTHIA-NLO difference (about 10% above the normalization region) as a mass-dependent systematic uncertainty to be entered into our limit-setting procedure.
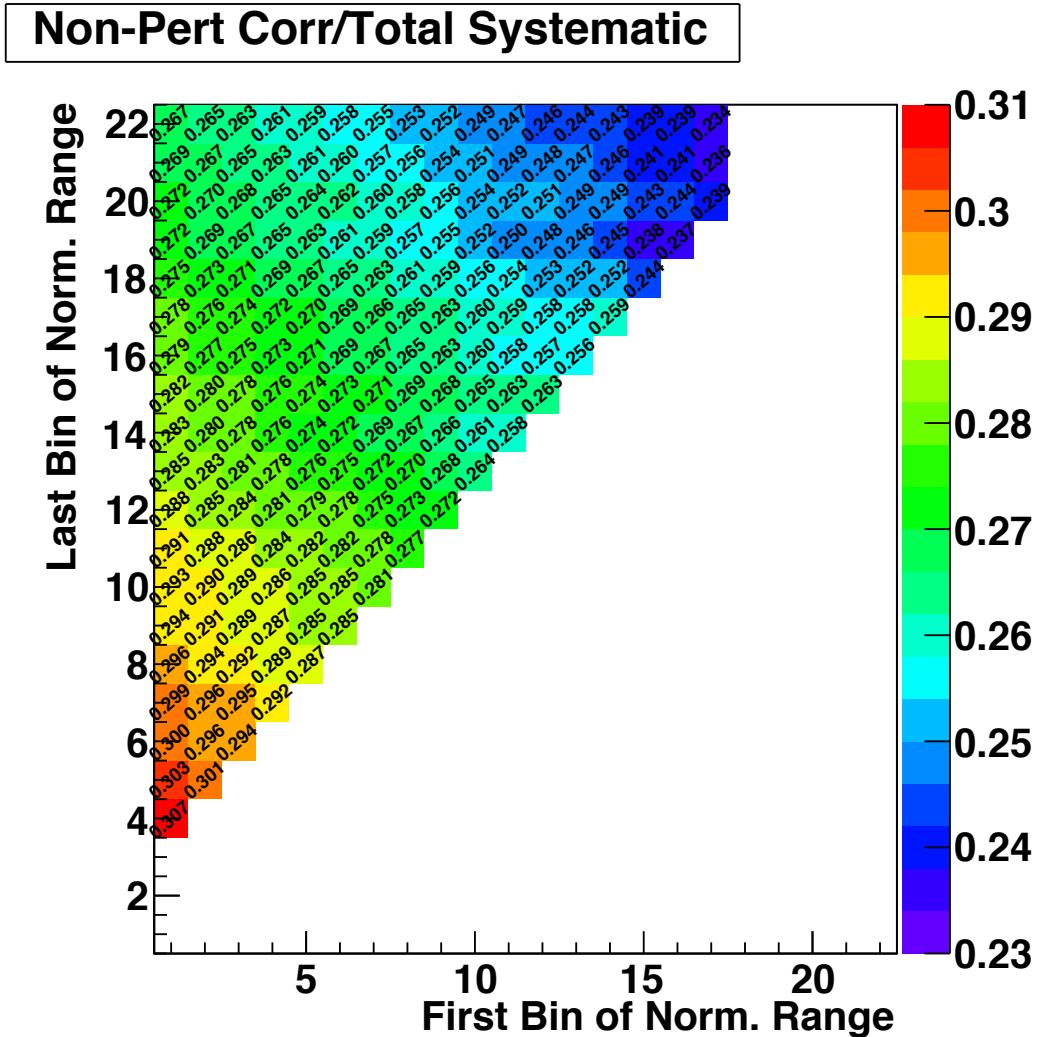
Figure 53: The fraction of the total systematic uncertainty in the normalization procedure from non-perturbative corrections as a function of normalization range [Dijet Centrality Ratio group]
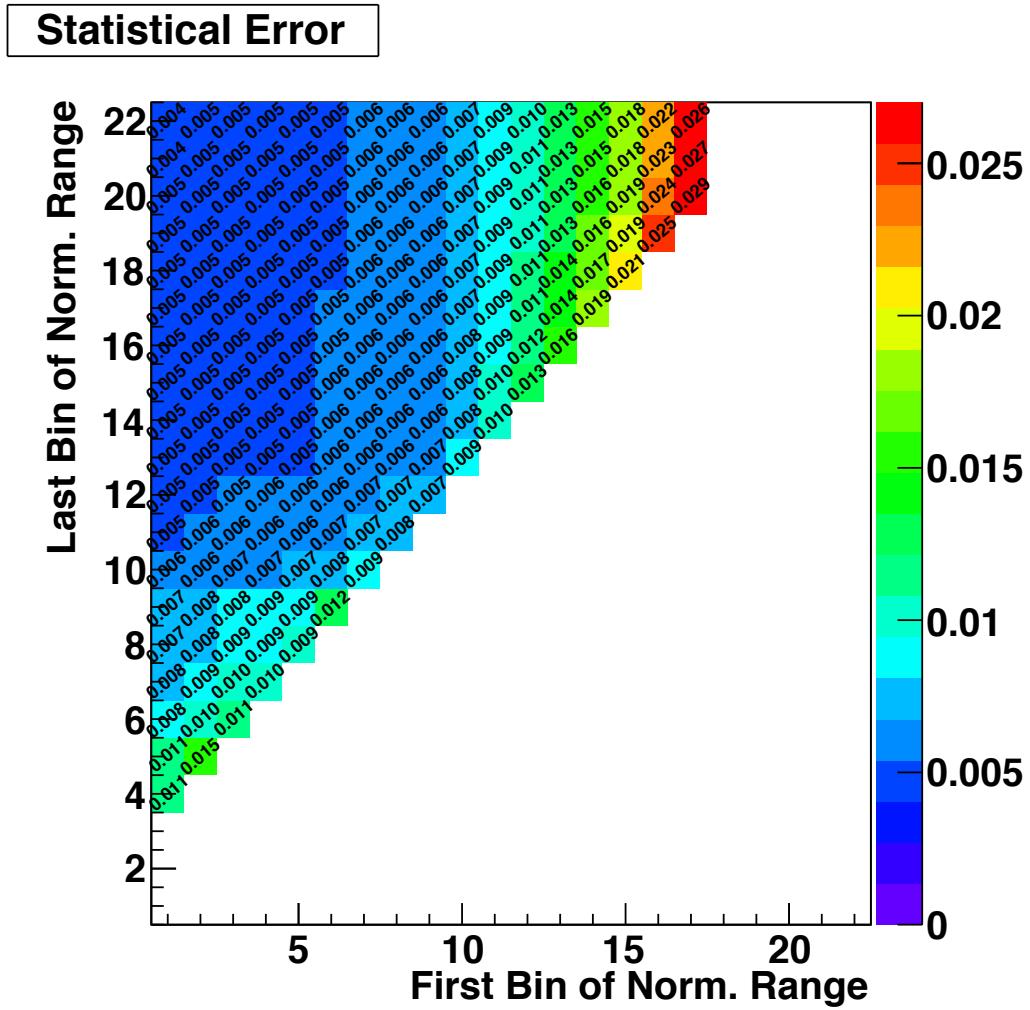
Figure 54: The fraction of the total systematic uncertainty in the normalization procedure from statistical uncertainty as a function of normalization range; a balance between this plot and the previous plot (figure 53) results in a normalization range choice of bins 13 to 19 [Dijet Centrality Ratio group]

6.1.2 Contact Interaction Predictions

Using PYTHIA, we generate a large collection of contact interaction samples for various values of Λ. In order to save computing time and storage space, we use only the generator-level simulation and exclude the full detector simulation for these samples (this is justified at the beginning of section 6.1).
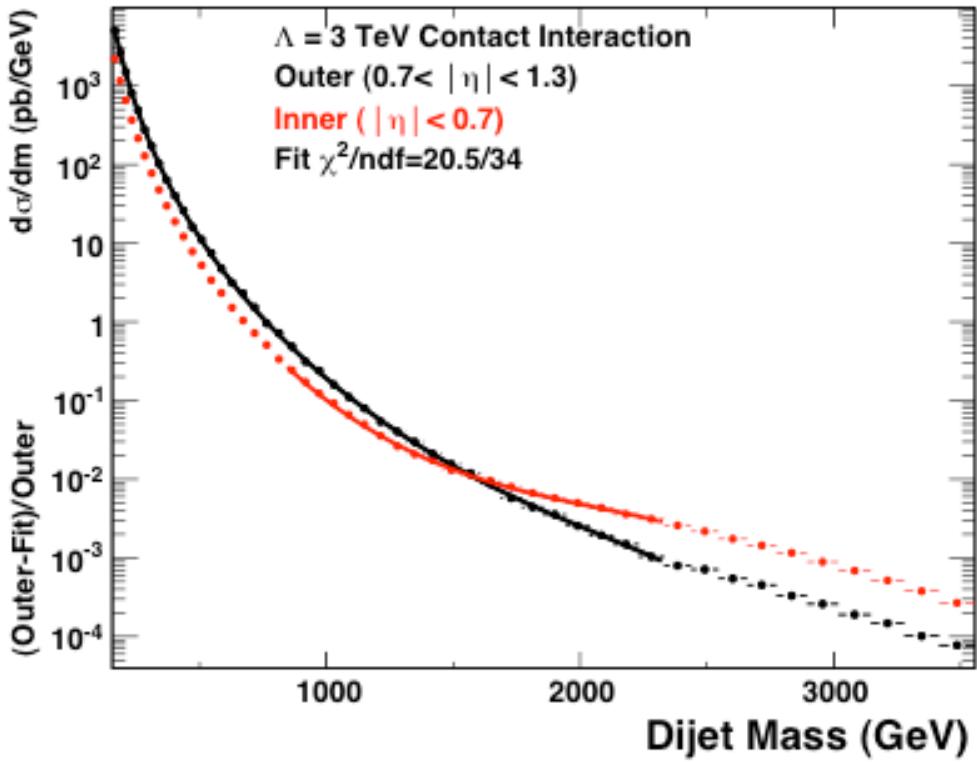
Figure 55: PYTHIA prediction for inner and outer dijet mass spectra with $\Lambda = 3$ TeV contact interaction; the bottom plot demonstrates the quality of the fit to the outer mass spectrum (later used in the limit-setting procedure) [author-generated]
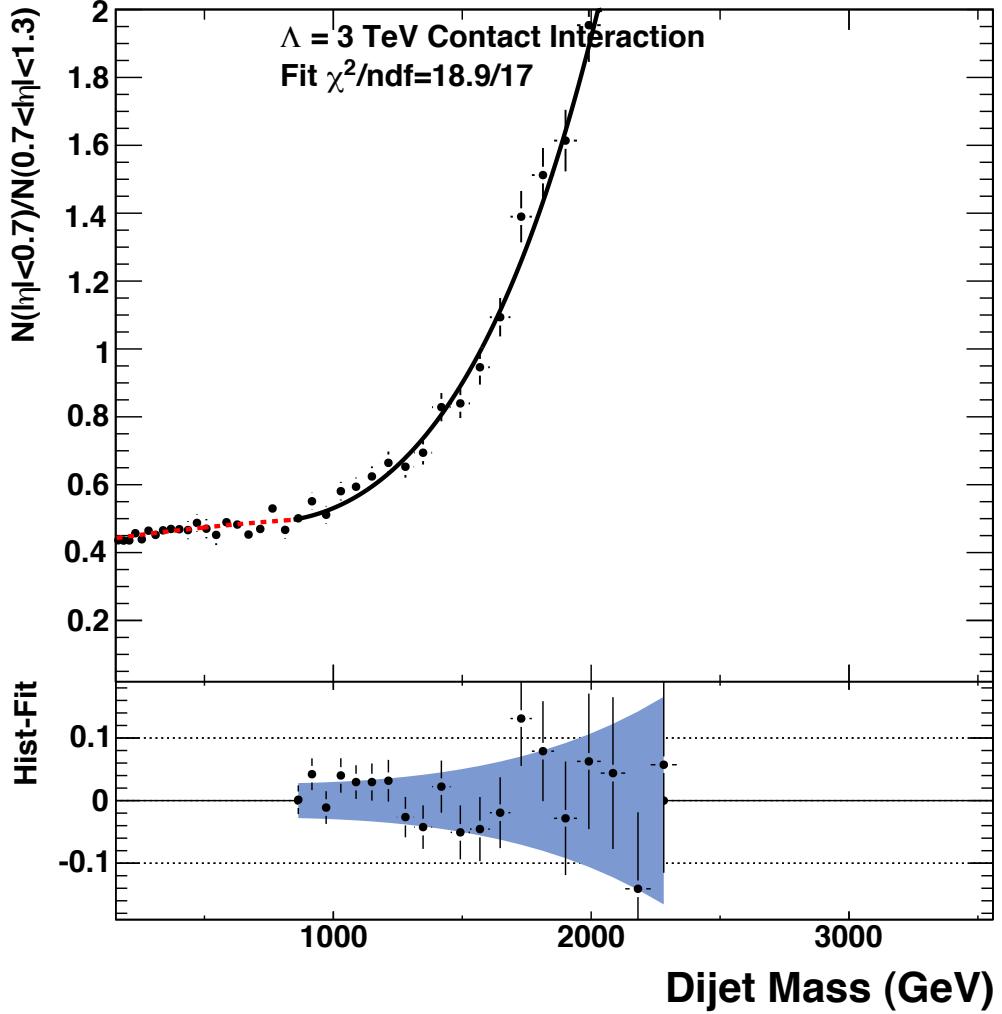
Figure 56: PYTHIA prediction for dijet centrality with $\Lambda = 3$ TeV contact interaction; the bottom plot demonstrates the quality of the fit to the dijet centrality ratio (later used in the limit-setting procedure), blue band is fit uncertainty [author-generated]

An example of such a generated sample – in this case for $\Lambda = 3$ TeV – is shown in figures 55 and 56. Note how the rise over the inner count over

the outer count, whuch is not seen in QCD. We define the turn-on of the contact interaction (in this case 838 GeV) to be the point at which a steady rise above QCD begins to occur in the dijet centrality ratio due to the contact interaction. Below this turn-on we constrain our fit to be identical to the QCD prediction. Above the turn-on we fit the contact interaction curve to a fourth order polynomial, which is terminated where the curve begins to turn over and flatten as the definition of the contact interaction breaks down near its energy scale. We call this termination point the turn-off mass. Masses above the turn-off are not used in the limit-setting procedure. This is a conservative approach, as it only serves to exclude data from the limit-setting procedure. The turn-on and turn-off values are determined directly, and are listed in table 6.1.2.

Λ (GeV)	turn-on (GeV)	turn-off (GeV)
2000	693	1950
2200	740	2000
2400	740	2200
2600	788	2300
2800	788	2400
3000	838	2500
3200	838	2500
3400	890	2700
3600	944	2800
3800	1000	2900
4000	1058	3000
4200	1118	3100
4400	1181	3200
4600	1246	3300
4800	1313	3400
5000	1383	3500

Table 3: Contact Interaction Turn-Ons and Turn-Offs

6.2   Limit-Setting Methodology

In the absence of obvious and compelling evidence for new physics in the dijet centrality ratio, we seek to set a limit on the energy scale Λ of any quark contact interactions, as described. We do this using a CL_S method [27] for a

log likelihood ratio (LLR) test statistic. Systematic uncertainties, which are not intrinsically included in this, are incorporated in ensemble testing via the Cousins-Highland method [12]. These methods are described in the following sections.

6.2.1 The Log Likelihood Ratio

We begin by defining the likelihood for each mass bin j as the probability to observe n_{in} and n_{out} events with respective Poisson expectations of μ_{in} and μ_{out} (*in* referring to events in which the two leading jets have $|\eta| < 0.7$ and *out* referring to events in which the two leading jets have $0.7 < |\eta| < 1.3$):

$$L_j = P(n_{in}, n_{out}|\mu_{in}, \mu_{out}) \tag{15}$$

$$L_j = P_{Poisson}(n_{in}|\mu_{in})P_{Poisson}(n_{out}|\mu_{out}) \tag{16}$$

We can rewrite this in terms of a Poisson probability of the total counts and a binomial probability of the inner count given the total count $n_{tot} = n_{in} + n_{out}$ and the ratio $\rho = \mu_{in}/(\mu_{in} + \mu_{out})$ of expected values.

$$L_j = P_{Poisson}(n_{tot}|\mu_{tot})P_{Binomial}(n_{in}|n_{tot}, \rho) \tag{17}$$

Here, $\mu_{tot} = \mu_{in} + \mu_{out}$. Since the first term contains no information on the dijet centrality ratio, we eliminate it:

$$L_j = P_{Binomial}(n_{in}|n_{tot}, \rho) \tag{18}$$

As the total count is operant only as an auxiliary statistic, we condition on it. We compute the binomial probability with a Gaussian approximation when $n_{tot} > 10^8$ and the incomplete beta function $B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ (x being the variable and a and b being input parameters) for $n_{tot} \leq 10^8$ [13]. The total likelihood is the product of L_j for all bins. We switch to the natural logarithm so that we can we can take the log likelihood ratio LLR to be a simple difference between the the natural log of the likelihood for QCD and for an alternative hypothesis:

$$\ln L = \Sigma \ln L_j \tag{19}$$

We can then define the log likelihood ratio:

$$LLR = \ln L_{alt} - \ln L_{QCD} \tag{20}$$

6.2.2   The CL_S Method

The CL_S method [27] is a modified frequentist method for setting confidence limits on the presence of signal over background in a dataset. We start with a test statistic Q (in our study, Q is the LLR). We take the confidence in the signal plus background hypothesis to be given by the probability that the statistic is less than or equal to its observed value:

$$CL_{S+B} = P_{S+B}(Q \leq Q_{obs}) \tag{21}$$

Where:

$$P_{S+B}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{S+B}}{dQ} dQ \qquad (22)$$

Here, dP_{S+B}/dQ is the probability distribution function of the test statis-tic in a signal plus background experiment. If CL_{S+B} is small, the signal plus background hypothesis is unlikely and the background is favored. The confidence in the background-only hypothesis is analogous:

$$CL_B = P_B(Q \leq Q_{obs}) \qquad (23)$$

Where:

$$P_B(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_B}{dQ} dQ \qquad (24)$$

Here, dP_B/dQ is the probability distribution function of the test statis-tic in a background-only experiment. High values of CL_B indicate that the background hypothesis is favored.

At extremely low signal rates or in the presence of downward fluctuations, CL_{S+B} can produce ambiguous results which can potentially be unphysical. A solution to eliminate this possibility is to normalize CL_{S+B} to CL_B. This leads to our CL_S limit:

$$CL_S = CL_{S+B}/CL_B \qquad (25)$$

Although this leads to conservative coverage, it eliminates unphysical re-sults, and excludes the signal hypothesis when the confidence level CL satisfies

the following condition:

$$CL \geq 1 - CL_S \tag{26}$$

We now have a conservative but guaranteed-physical method of setting a confidence limit. In practice, we calculate this using ensemble testing of various signal plus background hypotheses (contact interactions at various energy scales) and a single background hypothesis (QCD). Ensemble testing is a process by which simulations of potential outcomes of an experiment using different sets of input hypotheses are produced in large numbers and then used to numerically calculate the difficult-to-determine probability distributions relevant to the experiment.

6.2.3   Limit-Setting
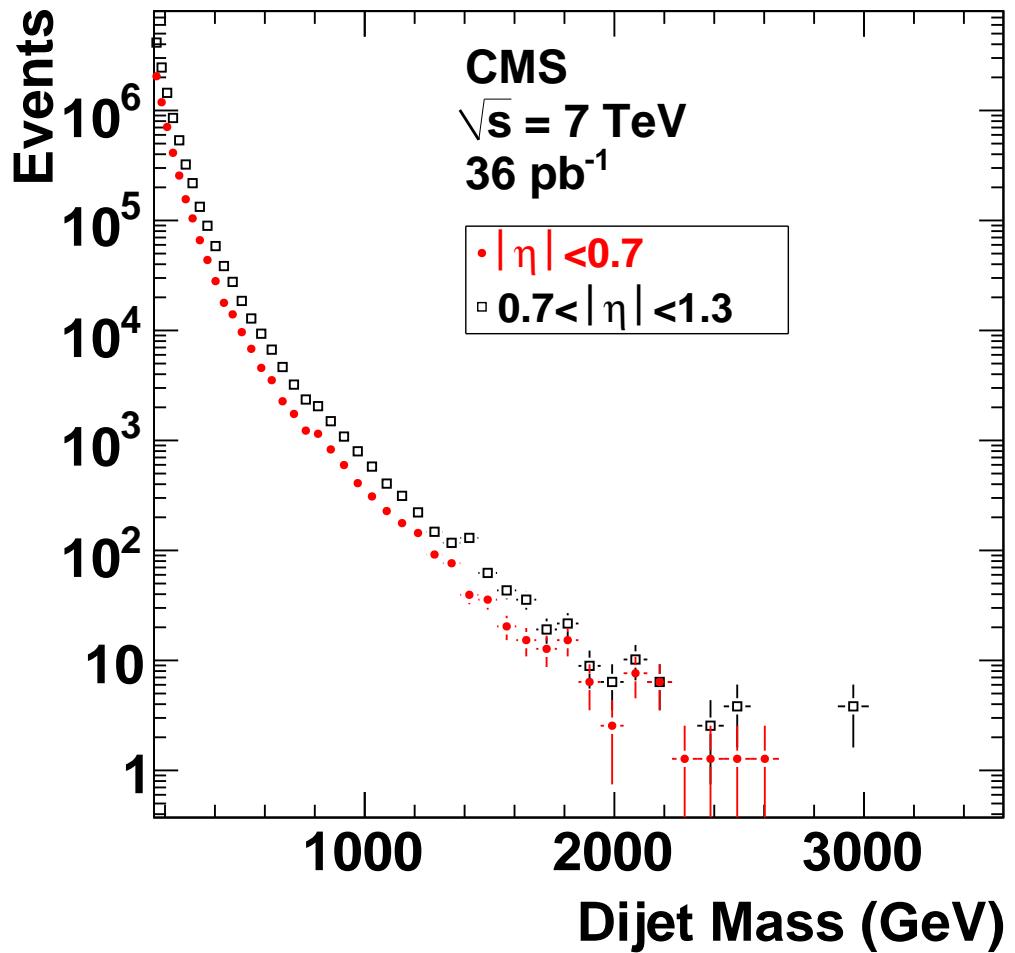
Figure 57: Inner and outer dijet mass spectra in 35.7 pb^{-1} of $\sqrt{s} = 7$ TeV data from the LHC [author-generated]
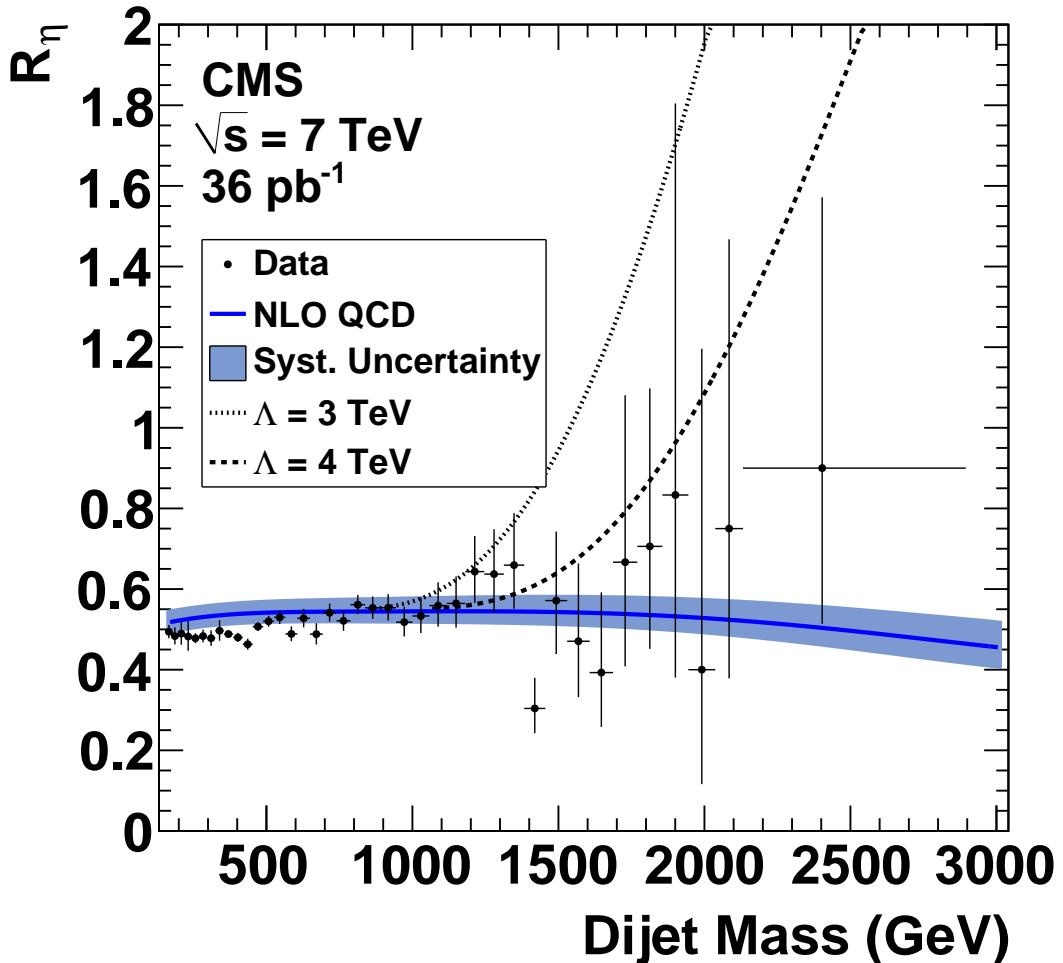
Figure 58: Dijet centrality ratio in 35.7 pb^{-1} of $\sqrt{s} = 7$ TeV data from the LHC with NLO QCD (with systematic uncertainties) prediction and contact interaction predictions at $\Lambda = 3$ TeV and $\Lambda = 4$ TeV [author-generated]

We have, as a background hypothesis, the QCD model described in section 6.1.1. For our signal hypothesis, we have contact interaction models at various energy scales as described in section 6.1.2. The dijet ratio data and

example hypotheses can be seen in figures 57 and 58. For each hypothesis, several thousand pseudo-datasets are generated. These pseudo-datasets are conditioned on the number of events in the actual data. We use the distributions of pseudo-datasets to determine the LLR expectation, the σ bands for our background (QCD) hypothesis and the 95% CL_S on all our hypotheses. Any contact interaction for which the 95% CL_S lies on the positive side of the data LLR value is excluded (see figures 59 and 60). Between generated Λ points, interpolation is used to obtain an exact value for the limit.
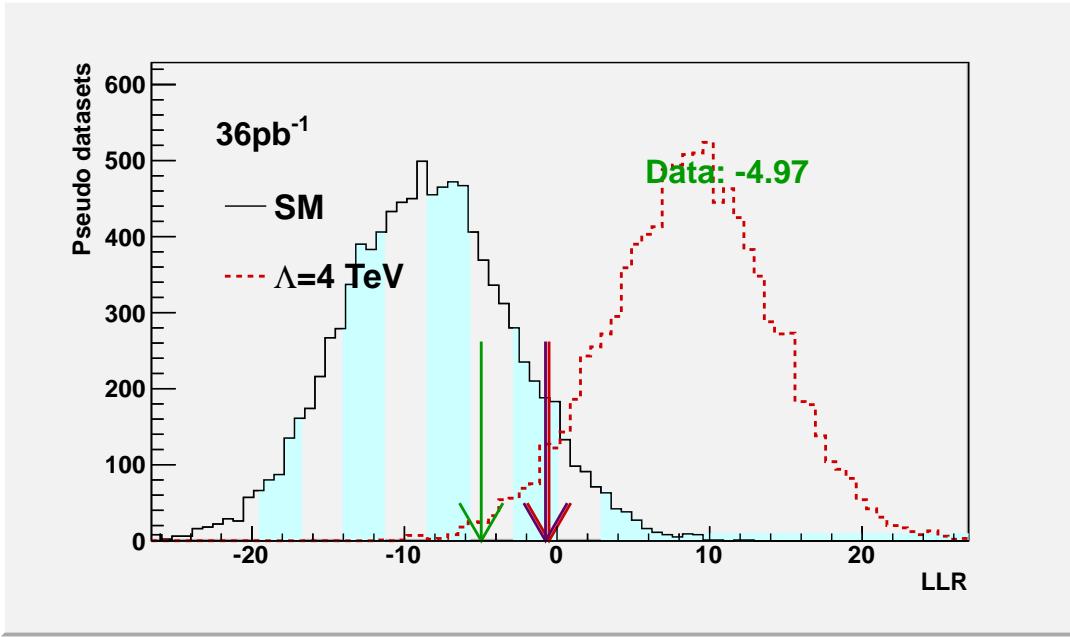
Figure 59: LLR ensemble testing results for $\Lambda = 4$ TeV, with the green arrow Indicating the data LLR Value, the red arrow indicating the 95% limiting CL_S LLR value and the purple arrow indicating the 95% limiting CL_{S+B} or pure frequentist LLR value [author-generated]
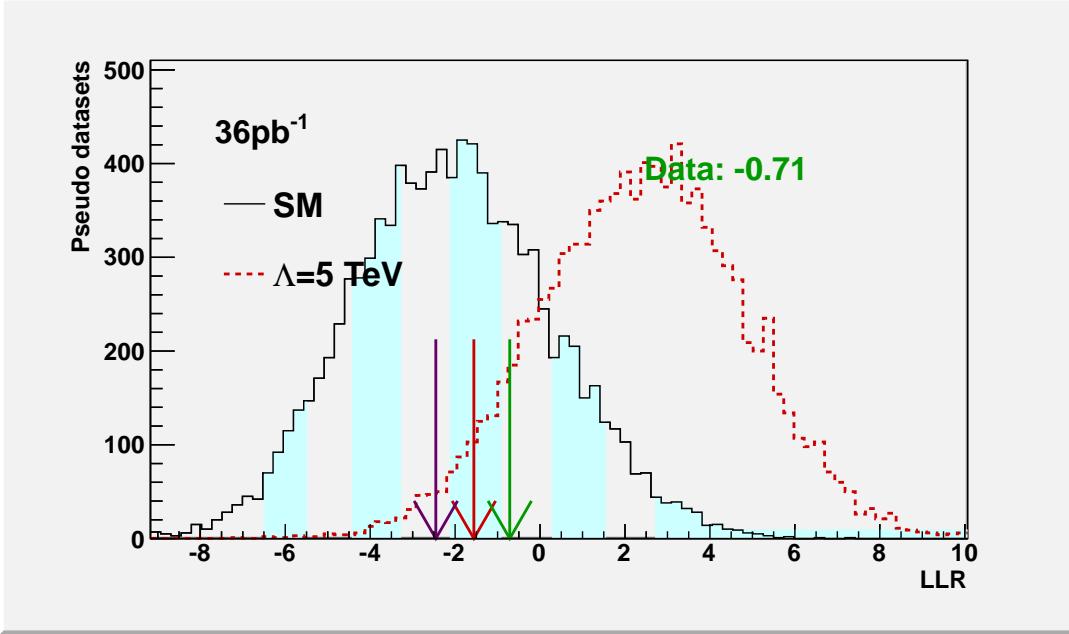
Figure 60: LLR ensemble testing results for $\Lambda = 5$ TeV, with the green arrow Indicating the data LLR Value, the red arrow indicating the 95% limiting CL_S LLR value and the purple arrow indicating the 95% limiting CL_{S+B} or pure frequentist LLR value [author-generated]

6.2.4   Systematic Uncertainties

The systematic uncertainties in the measurement are taken into account via the Cousins-Highland method [12] in the ensemble testing procedure. This essentially involves generating a random value from a Gaussian distribution of width equal to the systematic uncertainty and mean equal to the value of the entry without systematic uncertainties. The entry value is then replaced with the randomized value. The known systematic uncertainties are added in quadrature to produce a single nuisance parameter to use as the width of the

Gaussian. The systematic uncertainties used to generate this final nuisance parameter are considered in two categories: model uncertainties and detector uncertainties.
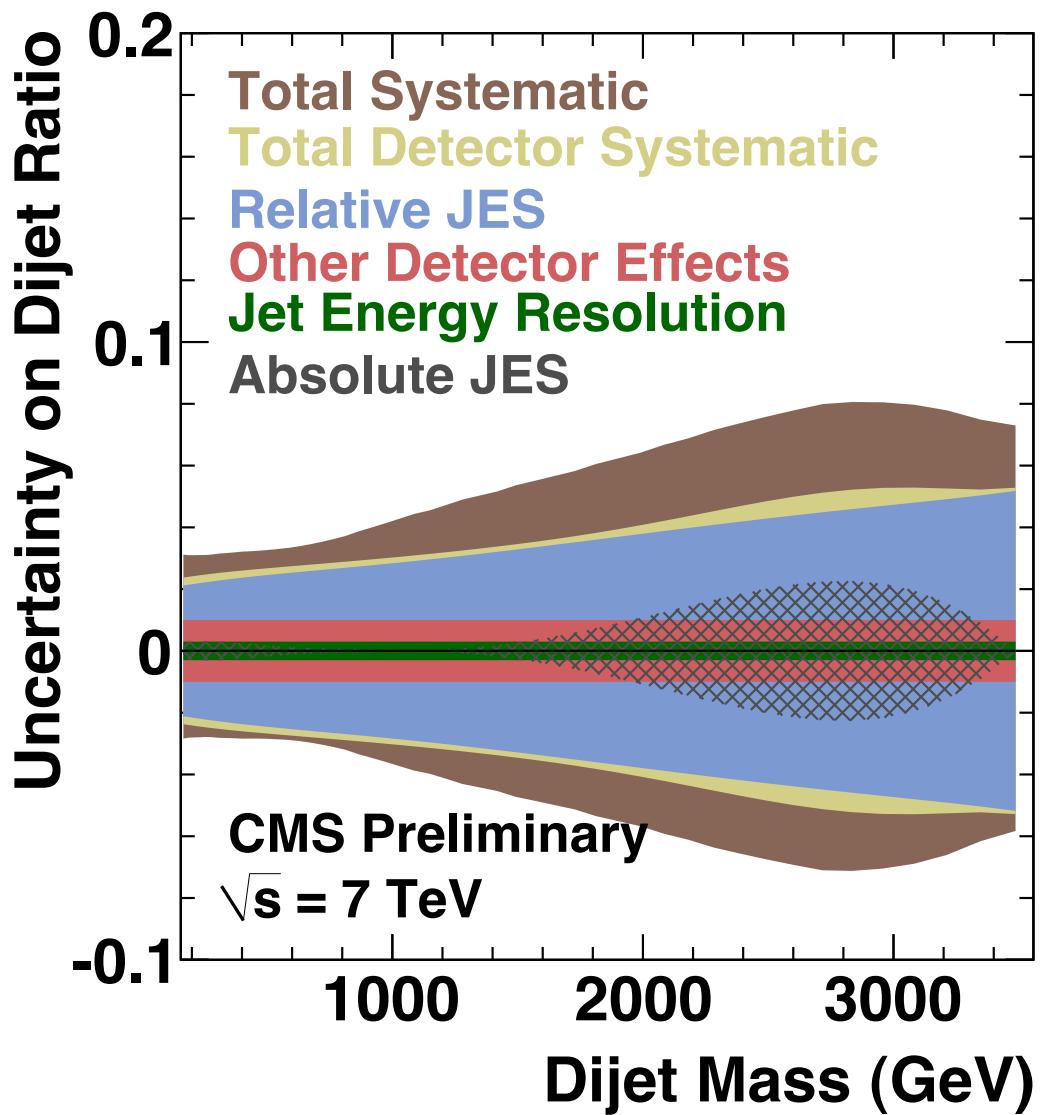
Figure 61: Detector systematic uncertainties in the dijet centrality ratio, including relative jet energy scale uncertainty, miscellaneous detector effects, jet energy resolution uncertainty and absolute jet energy scale uncertainty [Dijet Centrality Ratio group]
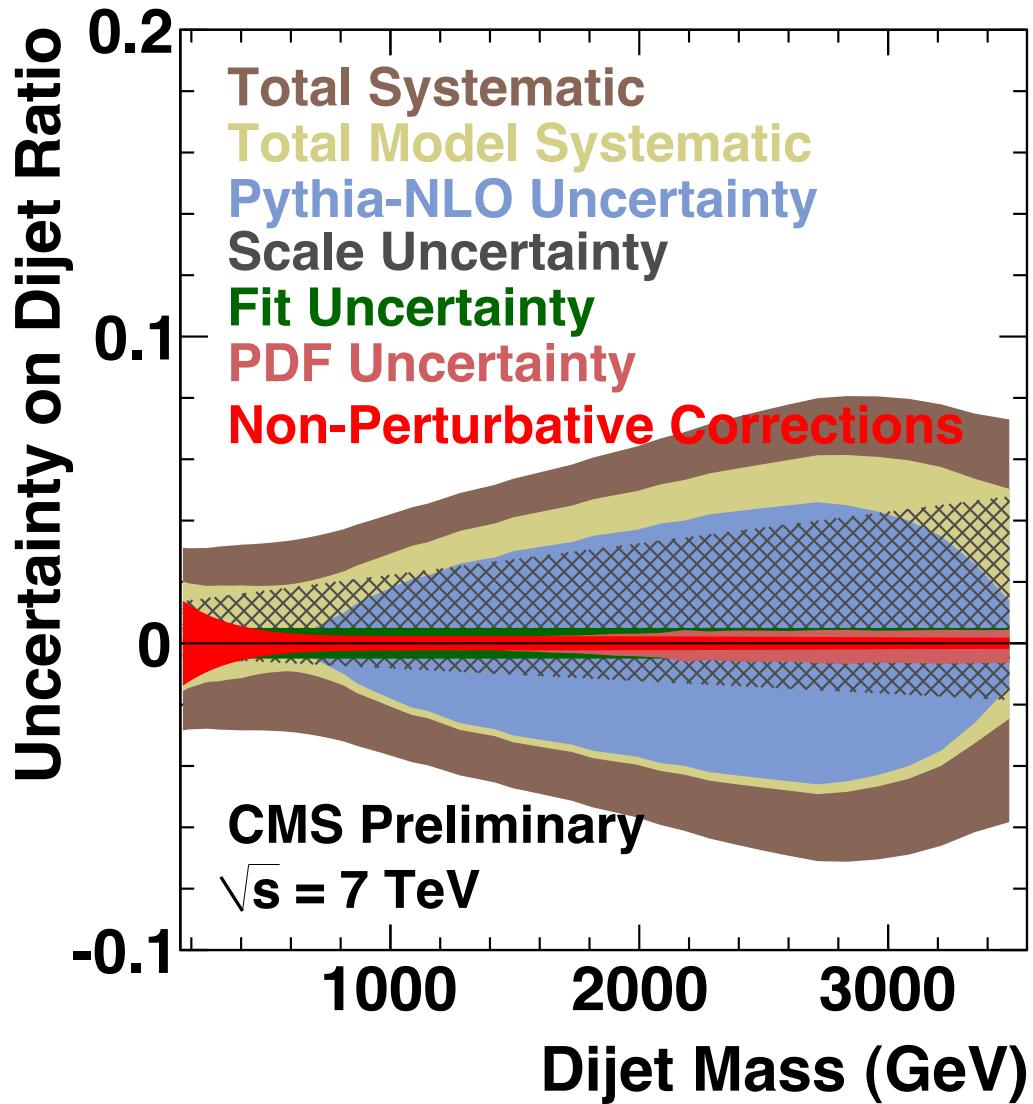
Figure 62: Model systematic uncertainties in the dijet centrality ratio, including PYTHIA - next to leading order uncertainty, scale uncertainty, fit uncertainty, parton distribution function uncertainty and uncertainty in non-perturbative corrections [Dijet Centrality Ratio group]

Detector uncertainties include absolute jet energy scale (JES), relative (i.e.

η-dependent) JES, jet energy resolution (JER) and generalized detector effects (see figure 61). Model uncertainties include renormalization / factorization scale, non-perturbative corrections, choice of parton distribution function (PDF), choice of QCD model, statistical uncertainty on offset (see section 6.1.1), and fit uncertainty (see figure 62).

Source	QCD (%)	Contact Interaction (%)
Detector	6.0 - 14.1	-
Relative JES	5.4 - 12.6	-
Absolute JES	0.0 - 5.8	-
JER	0.8	-
Other	2.6	-
Model	$+(3.8-17.0)$ $-(1.9-13.3)$	6.0 - 30.0
Absolute JES	-	5.0 - 30.0
PYTHIA-NLO	0.0 - 11.8	-
Offset	5.2	-
Scale	$+(3.5-11.5)$ $-(1.2-4.4)$	-
PDF	0.4	-
Fit	1.3	2.0 - 5.0
NPC	0.5 - 3.5	-
Total	$+(7.1-22.1)$ $-(6.3-19.4)$	6.0 - 30.0

Table 4: Systematic Uncertainties

In order to determine the systematic uncertainties on the JES, two tests are performed. Firstly, we re-calculate the dijet centrality ratio having shifted the

jet energies in both the inner and outer dijet mass spectra by 10% to test the uncertainty on absolute JES. Secondly, we re-calculate the dijet centrality ratio having shifted only the inner dijet energies by 1% to test the uncertainty on relative JES. As expected, to the first order, the uncertainties in the absolute JES cancel in the ratio and have a minimal effect (being the ratio between two mass spectra, the dijet centrality ratio is unaffected by simultaneous identical shifts in both spectra). The effect of the uncertainty in the relative JES is clearly implicated as the dominant detector systematic uncertainty.

In order to determine the effects of a poorly-simulated JER, we generate a simply parametrized η-dependent modification to the jet energy resolution and determine its effect. This is done using a sample of truth-matched (i.e. generator level and fully reconstructed jets are indexed and matched; unmatched sets are discarded) fully-simulated calojets and genjets. The four-momentum p of each inner genjet is inflated by 10% of the difference between the calojet and genjet energy: $p_{inflated} = 1.1(p_{calo} - p_{gen}) + p_{gen}$. The effect of this on the dijet centrality ratio is determined to be mass-independent and negligible (a straight line fit yields a value of 0.003 ± 0.001 in units of the ratio).

Other generalized detector effects are quantified in terms of the differences between genjets and corrected calojets in the simulation. Referring back to figures 38 and 39, we can see that the fit to the difference between the two, 0.010 ± 0.005 in units of ratio is nearly zero, and take 0.01 in units of the ratio (2.6%) to be the systematic uncertainty on this.

The biggest model systematic uncertainty comes from the choice of QCD model. The maximal difference in models is taken to be the mass-dependent

systematic uncertainty in this case, and it can be as much as 11.8% of the value of the ratio. The fit uncertainty in the offset, described in section 6.1.1, provides an additional related systematic, and this is statistically determined to be 0.02 in units of the ratio (5.2%).

The next biggest contribution to the model uncertainty comes from the renormalization / factorization scales, which are scales used as cut-offs in the perturbative calculations in order to prevent divergent integrals. In the NLO calculations used in the simulations, the p_T of the jets is the relevant scale. This scale is varied between one half and twice its nominal value and the dijet centrality ratio is recalculated accordingly. The difference between these ratios and the nominal ratio are taken to be the upper and lower values for the systematic uncertainty arising from the renormalization / factorization scale. This varies with mass from 3.5% - 11.5% on the upper side and 1.2% - 4.4% on the lower side.

The non-perterbutive corrections provide an additional source of systematic uncertainties in the model. This is taken from the difference in the PYTHIA and Herwig++ models described in section 6.1.1. The models agree closely enough to prevent this uncertainty (even taken as a highly conservative overestimate) from growing above 3.5%. The final source of uncertainty in the QCD model comes from the choice of parton distribution function (PDF), which describes the way that the momenta of the partons (quarks and gluons) are distributed in each proton. The vast majority of these uncertainties cancel to the first order in the ratio and are less than 0.4% at all dijet masses.

In the contact interaction model, systematic uncertainties arise from the

absolute jet energy scale, which has a much greater effect than in the QCD model due to the flatness of the dijet centrality ratio in the QCD model versus the sharply rising slope of the dijet centrality ratio in a contact interaction model. There are additional uncertainties from the fits to the Monte Carlo data used to generate the idealized contact interaction models. Systematic uncertainties are summarized in table 6.2.4.

7 Results and Conclusions
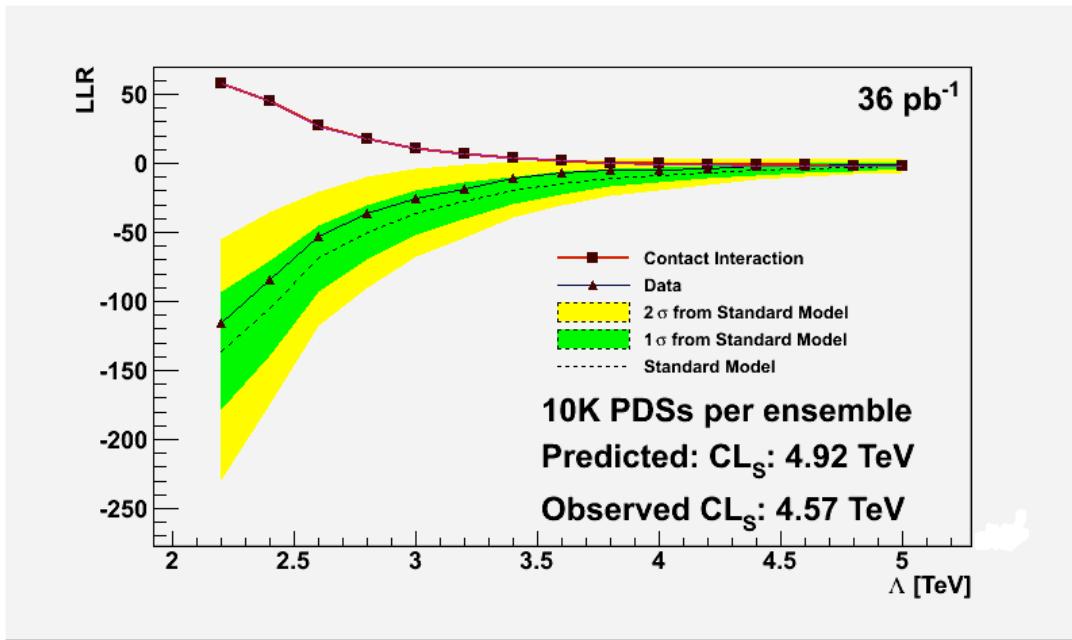
Figure 63: CL_S limit on contact Interaction scale; red line is contact interactions; dashed black line is predicted standard model; solid black line is data [author-generated]

Utilizing the methodology described (evaluation of the dijet centrality ratio using the CL_S method with systematics accounted for via the Cousins-Highland method) and the data (36 pb^{-1} gathered in 2010 in proton-proton collisions at $\sqrt{s} = 7$ TeV) at the LHC, we have been able to set a limit on a quark compositeness energy scale of $\Lambda < 4.57$ TeV with 95% confidence (see figure 63), and find our results to be generally consistent with the Standard Model as it is currently understood. We conclude that we have found no compelling evidence of composite structure in quarks below an energy scale of 4.57 TeV.

References

[1] I. Antoniadis. A possible new dimension at a few tev. Phys. Lett., B246, 1990.

[2] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. JHEP, April 2008.

[3] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant k(t) clustering algorithms for hadron hadron collisions. Nucl. Phys. B, 406, 1993.

[4] The CMS Collaboration. The cms experiment at the cern lhc. Jinst, 3(S08004), August 2008.

[5] The CMS Collaboration. Jet performance in pp collisions at $\sqrt{s} = 7$ tev. CMS PAS JME-10-003, 2010.

[6] The CMS Collaboration. Search for dijet resonances in 7 tev pp collisions at cms. Phys. Rev. Lett., 105, December 2010.

[7] The CMS Collaboration. Search for quark compositeness with the dijet centrality ratio in pp collisions at $\sqrt{s} = 7$ tev. Phys. Rev. Lett., 105, December 2010.

[8] The CMS HCAL Collaboration. Design, performance, and calibration of cms hadron barrel calorimeter wedges. CMS-NOTE-2006-138, 2007.

[9] The DZero Collaboration. Dijet mass spectrum and a search for quarks compositeness in $\overline{p}p$ collisions at $\sqrt{s} = 1.8$ tev. Phys. Rev. Lett., 82, March 1999.

[10] The DZero Collaboration. Measurement of dijet angular distributions at $\sqrt{s} = 1.96$ tev and searches for quark compositeness and extra spatial dimensions. Phys. Rev. Lett., 103, 2009.

[11] The GEANT4 Collaboration. Geant4: A simulation toolkit. Nucl. Inst. Meth. A, 506, 2003.

[12] R.D. Cousins and V.L. Highland. Incorporating systematic uncertainties into an upper limit. Nucl. Instrum. Meth., A320, 1992.

[13] R.D. Cousins, J.T. Linneman, and J. Tucker. Evaluation of three methods for calculating statistical signiÞcance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a poisson process. Nucl. Instrum. Meth., A595, 2008.

[14] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. JHEP, 1997.

[15] E. Eichten, I. Hinchliffe, K. D. Lane, and C. Quigg. Super collider physics. Rev. Mod. Phys, 56, 1984.

[16] E. Eichten, K. D. Lane, and M. E. Peskin. New tests for quark and lepton substructure. Phys. Rev. Lett., 50, March 1983.

[17] S. Esen and R. Harris. Cms sensitivity to quark contact interactions using dijets. CMS NOTE 2006/071, 2006.

[18] M. Bahr et al. Herwig++ physics and manual. Eur. Phys. J, C58, 2008.

[19] R. Harris et al. Search for dijet resonances in the dijet mass distribution in pp collisions at $\sqrt{s} = 7$ tev. CMS AN-2010/108, 2010.

[20] L. Evans and P. Bryant (editors). Lhc machine. Jinst, 3(S08001), August 2008.

[21] David Griffiths. Introduction to Elementary Particles. Wiley, 1987.

[22] M. Jha, R. Harris, and M. Zielinski. Dijet ratio from qcd and contact interactions. CMS AN-2007/015, 2007.

[23] S. P. Martin. Perspectives in Supersymmetry, chapter A Supersymmetry Primer. World Scientific, 1997.

[24] Z. Nagy. Three-jet cross sections in hadron hadron collisions at next-to-leading order. Phys. Rev. Lett., 88, March 2002.

[25] Z. Nagy. Next-to-leading order calculation of three-jet cross observables in hadron hadron collisions. Phys. Rev. D, 68, November 2003.

[26] L. Randall and R. Sundrum. A large mass hierarchy from a small extra dimension. Phys. Rev. Lett., 83, October 1999.

[27] A. L. Read. Modified frequentist analysis of search results (the cl(s) method). Prepared for Workshop on ConÞdence Limits, Geneva, Switzerland, 17-18 Jan. 2000.

[28] T. Sjostrand, S. Mrenna, and P. Z. Skands. Pythia 6.4 physics and manual. <u>JHEP</u>, 2006.