Search for Low-Scale String Resonances with ATLAS at the Large Hadron Collider

by

Fairhurst Lyons

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics

University of Alberta

Abstract

String theory requires the existence of extra spatial dimensions. They must be compactified since they have not yet been detected. If the size of these extra dimensions is large then it is possible that the string scale M_s could be on the order of TeV. In the Dp-brane model of string theory two-parton scattering in proton-proton collisions can create massive string resonances that could be detected at colliders such as the LHC. We simulate proton-proton collisions producing string resonances and compare with data taken with the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. We search the dijet invariant mass distribution for a significant signal over the expected background. We detect no excess in the data distribution over background and set a lower limit on the string scale of the first string resonance. We find with 95% confidence level that string resonances with an observed upper limit $M_s < 8.30$ TeV do not exist in ATLAS data with an integrated luminosity of 139 fb^{-1} .

Acknowledgements

There are many people that I would like to thank for their help along the way in making this research possible. I am extremely grateful to Dr. Douglas Gingrich for allowing me the opportunity to complete this project. I am very thankful for his supervision, support, and feedback. Thank you for being so patient in answering my questions, and for providing me with guidance and wisdom.

I very much appreciate the other members of our research group for their willingness to help whenever I needed it. I would like to acknowledge Haoji Ni, Jyoti Prakash Biswal, Nishu, Jack Lindon, and Dengfeng Zhang.

Lastly, I would like to thank my friends and family. They provided me with much-needed moral support throughout the duration of this project. I am forever grateful for the love, friendship, and endless words of encouragement.

Contents

List of Tables

List of Figures

List of Abbreviations

ATLAS	A Toroidal LHC apparatuS
CL	Confidence Level
CMS	Compact Muon Solenoid
EM	ElectroMagnetic
g	Gluon
GR	General Relativity
GRL	Good Run List
HL-LHC	High Luminosity LHC
HLT	High Level Trigger
JER	Jet Energy Resolution
JES	Jet Energy Scale
LCW	Local Cluster Weighting
LHC	Large Hadron Collider
LHCb	LHC beauty
LHCf	LHC forward
LHE	Les Houches Event
LO	Leading Order
MC	Monte Carlo
NLO	Next-to-Leading Order
NN	Neural Network
NP	Nuisance Parameter
p	Proton
Pb	Lead
PDF	Parton Distribution Function
q	Quark
QCD	Quantum ChromoDynamics
ROI	Region Of Interest
SM	Standard Model
TDAQ	Trigger and Data AQuisition

Statement of Originality

The figures and tables in this thesis are the author's own, unless denoted by citation. Several contributions to this work were made by members of the ATLAS collabpration. The STRINGS-1.00 [42] generator was written by Pourya Vakilipourtakalo and Douglas Gingrich, and modified by the author. Data and QCD MC samples including jet reconstruction were generated by Dengfeng Zhang. Signal samples including uncertainties were generated by Douglas Gingrich and Katherine Rybacki. The background fit was produced using the SWiFt [65] software package. The resonance search was performed using the BumpHunter [70] package. Limit setting was performed using the HistFitter [72] package.

1 Introduction

The Standard Model (SM) of particle physics is our most complete theory that describes the particles that we observe and the forces that they interact with. The SM describes our universe in terms of zero dimensional particles that define all matter and interactions. This is a framework that describes the observable properties of particles, which in turn govern what states they can inhabit. There are some outstanding questions about the model and problems that it has not yet answered, including the hierarchy problem [1]. String theory is a candidate to replace the SM that includes all of the fundamental forces we observe and answers some of the questions left within the SM. In string theory we describe the universe in terms of one dimensional strings that move around in our spacetime and can either be open or closed. These strings are observed as the SM particles depending on the string energy and configuration. This theory reproduces what we observe, simplifying to the SM in certain approximations. String theory requires the existence of extra spatial dimensions that we have not yet detected. Objects called Dp-branes are also required to exist in spacetime, which open strings must be attached to. In this theory fermions are observed from open strings with their ends attached to stacks of Dp-branes while bosons stretch between Dp-branes [2]. If the size of the extra spatial dimensions is large then the energy scale of these strings can be on the order of TeV. In this case the effects of string theory would be detectable at current Large Hadron Collider (LHC) energies [3]. String theory predicts different interactions between protons than the SM at energies around the fundamental string theory energy scale. By studying collisions between protons we should be able to determine whether interactions are caused by the SM or string theory. Theoretical string theory and SM predictions can be simulated and compared to data obtained at colliders to determine which theory best describes our observations. In this analysis we will study data taken from the ATLAS experiment at the LHC and search for evidence of this theory. Previous resonance searches have been performed with dijet events looking for various new physics signals such as W', excited quark $q*$, and quantum black holes [57]. The full Run 2 ATLAS data has not yet been analysed for string resonances.

Section 2 provides a theoretical background on the SM and string theory. This includes an overview of the SM framework, along with some questions that remain unanswered within the theory. This section also describes string theory, specifically the Dp-brane model that we are interested in. Section 3 discusses proton-proton (pp) collision such as those that occur within the ATLAS detector. Theoretical background on how we study these collisions is given, including how string resonances differ from SM predictions and could potentially be found by studying these collisions. Section 4 provides a description of the LHC and ATLAS detector, including hardware, and detection and reconstruction algorithms. Section 5 provides a description of the simulated string resonances, expected SM background, and ATLAS data used in this analysis. In Section 6 we compare our string theory and SM models to data from the ATLAS detector. A resonance search is conducted and we set limits on the string theory energy scale. Section 7 gives a summary of the analysis.

2 Review of Theory

2.1 Standard Model

The SM of particle physics describes three of the four fundamental forces that we have identified. These are electromagnetic (EM), strong, and weak forces, whose interactions are based on quantum field theory. The fourth interaction, gravity, is described by general relativity (GR). GR is a classical theory and it is not compatible with a quantum formulation. A consistent quantum theory of gravity has yet to be constructed so it remains separate from the SM. Nonetheless, the SM has been very successful in describing the phenomena that we observe. It has even predicted the existence of several particles before they were experimentally identified such as the top and charm quarks, and Higgs Boson.

The SM is based on interactions between the particles shown in Figure 2.1. Particles are

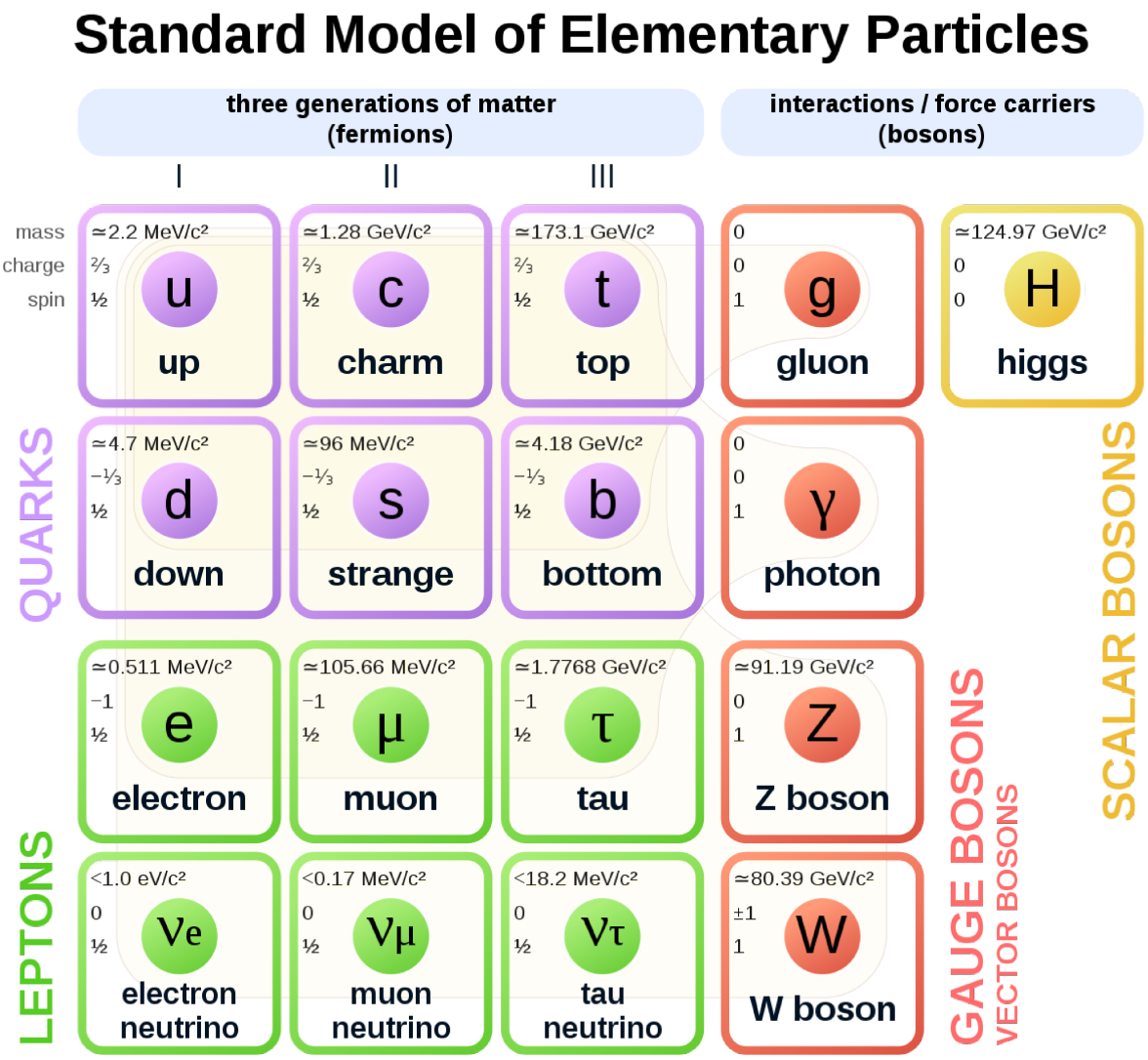


Figure 2.1: The particles that make up the standard model of particle physics. Each fermion also has a corresponding anti-particle with the same mass but opposite charges [4].

separated into two groups according to their spin. Particles with integer spin are called bosons, which obey Bose-Einstein statistics. Particles with half-integer spin are called fermions and follow Fermi-Dirac statistics. Fermions are divided into two categories, leptons and quarks, both of which have three sub-groups called generations. Within each generation of leptons is one flavour of lepton and its corresponding neutrino. There are six flavours of quarks, each generation contains one up-type and one down-type. Quark and lepton anti-particles also exist as fermions.

The SM is invariant under certain symmetries that arise from gauge fields, which then lead to the existence of gauge bosons. Gauge bosons are the mediators of the three forces within the theory. The symmetry groups of the SM are unitary U and special unitary SU groups $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where the subscripts represent colour, left-handedness, and weak hypercharge, respectively. These forces act on particles according to certain quantum numbers that they possess. Elementary fermions only interact with forces for which their corresponding quantum numbers are nonzero. These quantum numbers are associated with conservation laws. Spin is an example of a quantum number that governs properties of a particle, specifically what type of quantum states it can obey. Elementary particles can have spin with any integer or half-integer value and spin must be conserved in any interaction. Spin is related to the helicity of a particle, also known as its handedness. This can take one of two values: if a particle's spin is in the opposite direction as its momentum then the particle is left-handed, and if its spin is in the same direction as its momentum then it is right-handed. Charge is the quantum number associated with the EM force and can be either positive or negative. Particles can have integer or fractional values of both positive and negative charges. Charge must be conserved in EM interactions. Weak hypercharge is the quantum number associated with the weak force; particles can have many different weak hypercharges including integer and fractional values, both positive and negative. Weak hypercharge is conserved in all weak interactions. Colour is the quantum number associated with the strong force. There are three different colour charges along with corresponding anticolour charges. When all three charges are combined together the state is said to be colourless. Colour is conserved in all QCD interactions. Each force is described by a coupling, which is a unitless number that describes the strength of the interaction.

The EM force is described by quantum electrodynamics. This is a relativistic quantum field theory of classical electrodynamics that outlines the interactions between charged particles via gauge photons. The EM force can be unified with the weak force to a collective electroweak interaction, described by the left-handedness and hypercharge symmetries mentioned previously. This force is mediated by the massive gauge bosons $W^{\pm}$ and Z for weak, and the massless photon γ for EM interactions. The strong interaction is described by quantum chromodynamics (QCD), which corresponds to the $SU(3)$ colour group. The force particles in this interaction are gluons. Gluons come in eight colour combinations that must include at least one colour and one anticolour. The strong force has a small coupling at short distances and a constant coupling at large distances, leading to a phenomenon called confinement. Colourless states are made out of one particle with colour and another particle with its corresponding anticolour or three (anti)particles with one of each type of (anti)colour. Colourless glueballs formed of only gluons are also theoretically possible, although difficult to detect experimentally. Colourless

states are the lowest energy configurations so as particles with different colours are separated it eventually becomes energetically favourable to produce other coloured particles from vacuum to restore the colourless state. As a result particles that carry colour charge cannot exist on their own, they must combine to create colourless states. Gluons bind quarks to form hadrons in one of two categories, mesons or baryons. Mesons are formed out of one quark and one anti-quark, while baryons are made up of three quarks or anti-quarks [5].

2.2 Oustanding Problems

The SM is the most complete theory that we currently have to explain fundamental physics and has been successful in explaining and predicting experimental observations. There are, however, many questions that currently cannot be answered by this theory. For example, well-established measurements of neutrino oscillations cannot be explained by the SM. Conversions of one neutrino type to another have been observed experimentally, yet are forbidden by the SM due to lepton number conservation. This also requires nonzero neutrinos mass, whereas they must be massless in the SM. There is evidence in cosmology that the model is incomplete in the form of baryon asymmetry. In the early universe we would presume that an equal number of quarks and antiquarks were produced but there remain many more baryons (protons and neutrons) than antibaryons at present. This asymmetry over the course of the evolution of the universe cannot be explained within the SM. By studying astrophysical objects we can estimate the amount of mass in the universe and compare this with the amount of visible matter. Astrophysical observations lead to the conclusion that there must be a significant mass density in areas where visible matter is absent, called dark matter. The observed properties of these particles do not coincide with any particle within the SM so it must consist of some unknown particle. It turns out that a large number of observations related to the structure and evolution of the universe cannot be explained by the SM only, new particles and interactions have to be introduced [6].

One of the most puzzling issues with the SM is the hierarchy problem, which has been amongst the greatest motivations to construct a new theory beyond the SM. The hierarchy problem can be formulated in various ways but the operative dilemma is that there seem to be at least two different fundamental energy scales in nature, the electroweak scale $m_{EW} \sim 1$ TeV and the Planck scale $M_{Pl} \sim 10^{15}$ TeV. The Planck scale is where gravity becomes as strong as the three gauge interactions described by the SM in Section 2.1 [1]. Particle physicists have proposed new particles and forces to solve this problem, leading to theories such as supersymmetry and technicolor [7]. Another possibility is that our current model of gravity is flawed. We assume that the electroweak and Plank energy scales are both fundamental but electroweak interactions have been probed at distances $\sim m_{EW}^{-1}$ while gravity has only been accurately measured in the range ~ 0.01 m. The treatment of M_{Pl} as a fundamental energy scale is based on the assumption that gravity is completely unchanged over the entire range of magnitudes between where it has been measured at ~ 0.01 m and the Planck length $\sim 10^{-35}$ m. Given that the electroweak scale has been confirmed experimentally as a fundamental constant and the Planck scale has not, we are free to formulate a new theory beyond the SM in which we adjust the strength of gravitational interactions [1]. To answer this and other questions we must look for a new model.

One such candidate is string theory, which is described in Section 2.3.

2.3 String Theory

String theory is an attractive candidate for a unified theory of all fundamental forces. It is a quantum theory and includes gravity so it is a self-contained and consistent theory that describes all known interactions. In this theory instead of SM point particles the fundamental natural unit is strings that vibrate and move relativistically through spacetime. These strings correspond to the particles that we observe, depending on certain properties of a given string and the geometry of spacetime. In the SM there are several parameters that are used to formulate the theory whose values are not well motivated. By contrast, there are no adjustable dimensionless parameters in string theory. Quantities such as the number of spacetime dimensions must be input into the SM but can be derived from string theory.

There are multiple different string theories, most of which fall under either bosonic or superstring theories. As the name suggests, bosonic strings can only represent bosons not fermions so these theories cannot completely describe our observations and are useful only as a mathematical framework. Superstring theories include both bosons and fermions so all realistic string theories include superstrings. Superstring theories require the existence of 10 spacetime dimensions, nine spatial and one time. Since we observe only three spatial and one time dimension then the rest of the spatial dimensions must be compactified. It is possible that additional spatial dimensions exist but have remained undetected at low energy experiments if their volume is quite small. One way to compactify dimensions is through identification. For example, a circle can be seen as an open line where points are the same if they differ by $2\pi R$:

$$x(P_1) = x(P_2) + 2\pi R. \tag{2.1}$$

This identification turns the open line to a circle, making a non-compact dimension into a compact one if R is small. We can repeatedly compactify dimensions in this way, as illustrated with Figure 2.2. If we start with a plane in the x, y directions we can make the identification in Eq 2.1 to obtain a cylinder. If we then make a similar identification along the y-axis

$$y(Q_1) = y(Q_2) + 2\pi R, \tag{2.2}$$

we have compactified two dimensions and are left with a torus whose size is smaller than the plane we started out with. This is called torroidal compactification. It is not the only way that compactification of dimensions can be achieved but it is a simple mathematical framework that allows us to visualise and perform calculations with compactified dimensions.

There are multiple theories within the superstring theory category, we are interested in the D-brane model. In this theory strings move relativistically through spacetime and can either be open or closed, as shown in Figure 2.3. Closed strings can move freely but open string ends must be attached to objects called Dp-branes. D stands for Dirichlet - since the string ends must lie on the Dp-brane this specifies a Dirichlet boundary condition - and p is the number of spatial dimensions. In this theory closed strings represent particles called gravitons, tensor objects that mediate the force of gravity. Dp-branes can intersect and open strings attached to them result

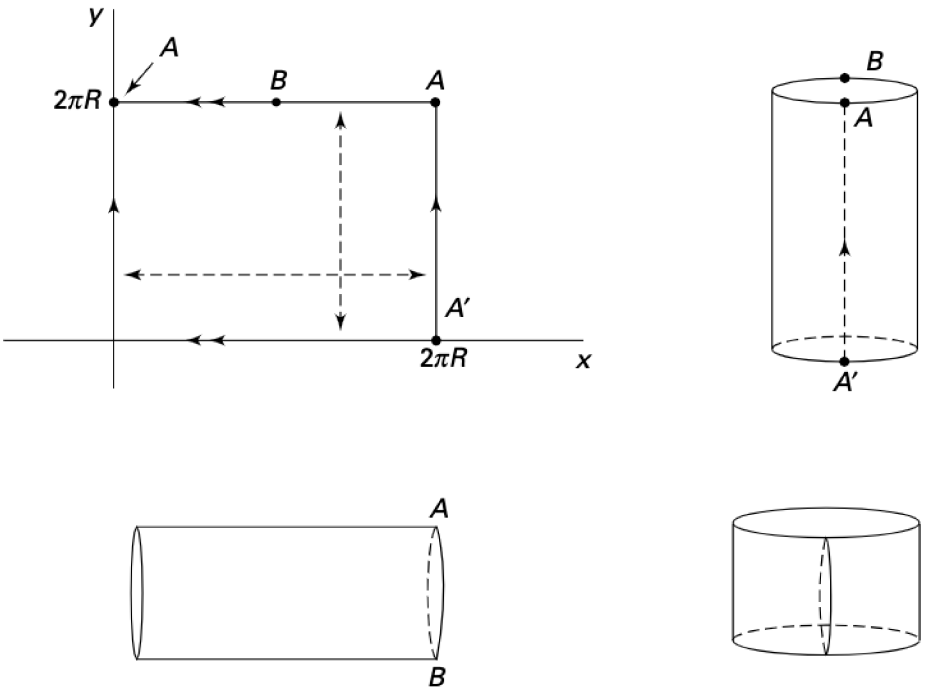


Figure 2.2: A rectangular region in the xy plane with identifications indicated by the dashed arrows. The identification along the x-axis results in a cylinder, when combined with the idenfication along the y-axis the result is a torus [2].

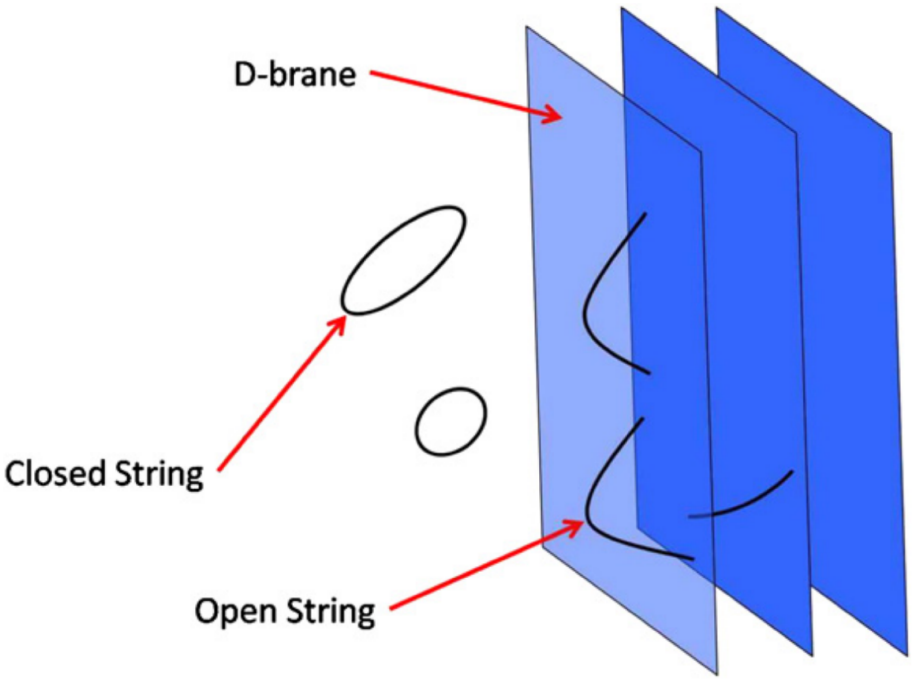


Figure 2.3: Both open string ends must be attached to a Dp-brane, this can be the same or different Dp-branes. Closed string ends are attached to each other, not Dp-branes [8].

in matter fields with the properties of different SM particles. The type of particle depends on the type of Dp-brane and the string orientation, as shown in Figure 2.4. N coincident Dp-branes

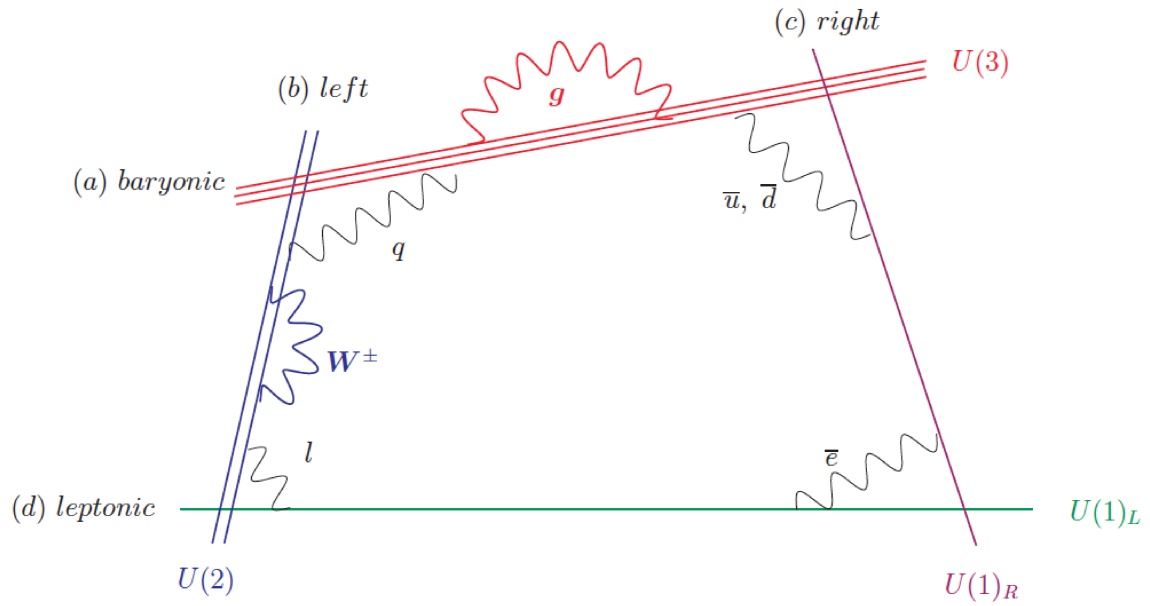


Figure 2.4: Intersecting Dp-brane stacks that give rise to SM particles [3].

give rise to $U(N)$ gauge bosons. In order to obtain the $SU(3)$ group that describes the colour force we then must have 3 coincident colour, or baryonic Dp-branes. The same goes for the $SU(2)$ weak force, which comes from two left-handed weak Dp-branes. Gluons interact only with the colour force so they correspond to open strings with both ends attached to a stack of baryonic Dp-branes. Quarks also have colour but they belong to the SM $SU(2)$ representation. They are chiral so a left-handed quark is made up of an open string with one end attached to a stack of three colour Dp-branes and the other attached to a stack of two left-handed weak Dp-branes. Electrons are open strings that stretch between a stack of two left-handed weak Dp-branes and a leptonic Dp-brane. All SM fermions and gauge bosons can be described by strings in the Dp-brane model. String theory is able to describe fundamental physics since it reproduces our observations.

There has yet to be any experimetal confirmation of string theory. This would require a unique and testable prediction, which has proven difficult to identify. Most string theory models predict deviations from the SM at higher energies than we can probe with current technology. We do not have the experimental capabilities required to measure the expected differences between the SM and most string theory predictions. There is a class of models within string theory that make predictions that are testable at current collider energies. The deviations from the SM predicted by these string theories are measurable with the technology that we possess at hadron colliders [2]. These theories will be discussed in detail in Section 2.4.

2.4 Large Extra Dimensions

String theory requires the existence of extra spatial dimensions that have not yet been observed. This implies that they must be compactified in order to evade our detection. We are interested in a superstring theory that requires 10 spacetime dimensions, leaving six additional spatial dimensions. The large extra dimension model states that these extra dimensions may be much larger than many had initially considered. Some string theories suggest that the size of the extra dimension is on the order of the Planck length $\ell_{Pl} \sim 10^{-35}$ m but this need not be the case. As discussed previously the Planck scale has not been confirmed experimentally as a fundamental scale. We assume that gravitational interactions at the Planck length obey the same physics that we can measure at larger scales but this has not been measured. In our string theory gravity arises due to closed strings that can propagate in the full 10 dimensional spacetime so we can take the 10 dimensional gravity scale as our fundamental value. If we assume a very simple scenario of 10 dimensional spacetime with six compactified dimensions of the same radius R then the effective four dimensional M_{Pl} can be related to the 10 dimensional Planck scale M_{Pl10} by

$$M_{Pl}^2 \sim M_{Pl10}^8 R^6. \tag{2.3}$$

We are free to set $M_{Pl10} \sim m_{EW}$ to remove the problematic hierarchy and we must have $M_{Pl} \sim 10^{15}$ TeV as observed. Inserting these values and rearranging for R gives [1]

$$R \sim 10^5 \,\text{TeV}^{-1}. \tag{2.4}$$

The four dimensional Planck scale can be expressed in terms of the volume $V_6 = (2\pi R)^6$ of the compactified dimensions [3]:

$$M_{Pl}^2 = \frac{8}{g_s^2} M_s^8 \frac{V_6}{(2\pi)^6}, \tag{2.5}$$

where g_s is a dimensionless constant called the string coupling and the string scale M_s is a parameter that describes the energy at which string effects become noticeable. The parameters on the right side of the equation have not been measured so we are free to assume any value for them that satisfies our observed value of M_{Pl}. Inserting the R value obtained from Equation 2.4 and $M_{Pl} \sim 10^{15}$ TeV into Equation 2.5 gives

$$M_s \sim g_s^{\frac{1}{4}} \,\text{TeV}. \tag{2.6}$$

If the coupling of the theory is on the order of unity then the energy scale of strings is on the order of TeV. In this theory the four dimensional Planck scale is not fundamental, its value is a consequence of the compactification of extra dimensions and the true gravity scale is $\sim m_{EW}$. This provides an elegant solution to the hierarchy problem described previously. In this theory all fundamental forces have the same energy scale but our observed gravity is much weaker since it interacts with the extra dimensions while SM fields do not. Limits on g_s [9] and the size of the extra dimensions [10] must be set experimentally. If the extra dimensions are large then it is possible for the string energy scale to be on the order of TeV, which is observable with current collider technology.

3 Proton-Proton Collisions

3.1 Collider Kinematics

To describe the interactions between particles that we aim to study we will first discuss two-particle scattering in the centre-of-mass frame. We will consider a collision between two incoming particles labelled a and b that interact with each other in some way, resulting in two outgoing particles labelled c and d. Each particle has has four-momentum

$$p = (E, p_x, p_y, p_z)\,, \tag{3.1}$$

and three-momentum denoted as $\vec{p}$, as shown in Figure 3.1. Interactions can be elastic, where

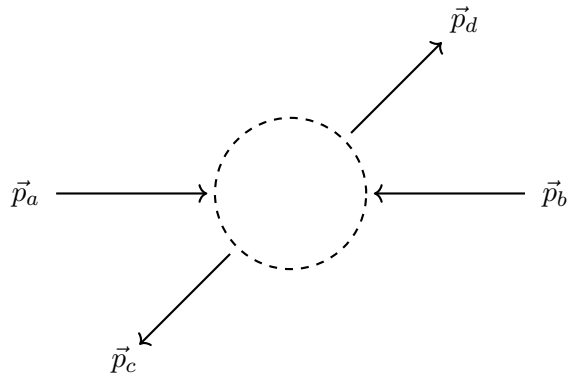


Figure 3.1: Interaction between two particles in momentum space in the centre-of-mass frame. Two incoming particles undergo some interaction resulting in two outgoing particles.

beam particles collide with each other and rebound, or inelastic, where particles are torn apart by the collisions. Collisions can also be categorised at either hard or soft, depending on if the momentum transfer between particles is large or small, respectively. Most inelastic events are soft interactions.

The interactions that occur in hadron colliders are recorded in the rest frame of the detector, or lab frame. In order to relate the kinematics that we measure in the lab frame to those in the centre-of-mass frame we can use variables that transform easily under Lorentz boosts in the beam direction. We define the Lorentz invariant Mandelstam variables as

$$s = (p_a + p_b)^2, \qquad t = (p_a - p_c)^2, \qquad u = (p_a - p_d)^2, \tag{3.2}$$

where the p_i are particle four-vectors. These variables are scalars so their value is the same regardless of our choice of reference frame. In a pp collider we can assume that each proton carries half of the centre-of-mass energy and that its momentum is only along the beam axis, which we define as the the z-axis. The momentum of the protons is $\sim$ 10 TeV while their mass is $\sim$ GeV so we can neglect their mass and consider only energy that comes from momentum. We will label the proton travelling in the positive z direction as a and the proton travelling in

the negative z direction as b:

$$p_a = \left(\frac{s}{2}, 0, 0, \frac{s}{2}\right), \qquad p_b = \left(\frac{s}{2}, 0, 0, -\frac{s}{2}\right). \tag{3.3}$$

Protons are not elementary particles but are made up of quarks bound together by gluons, these two types of particles are referred to as partons. Protons contain two up-type quarks and one down-type quark, known as valence quarks. These are embedded in a sea of virtual quark-antiquark pairs generated by the gluons that hold quarks together. At low energy protons can elastically rebound off of each other due to their EM charge repulsion. At high energies the protons come close enough to collide. Protons are not solid objects but are made up of partons separated by empty space. When protons collide at high energies the interactions are not between entire protons but rather between two-partons, one from each incoming proton. High energy pp collisions are inelastic since the incoming protons are destroyed in the collision. They are split into their constituent partons and these individual partons interact with each other. The collisions between partons may be elastic since they are elementary particles that can rebound off of each other at high energy. As the result of pp scattering we can thus have interactions between gluons of any type, and quarks and antiquarks of any colour and flavour. These are $2 \to 2$ scattering processes with two incident and two outgoing partons, each with four-momentum $k = (E, k_x, k_y, k_z)$ and three-momentum denoted by $\vec{k}$. We will consider elastic hard scattering events between two partons. Incoming partons are labelled 1 and 2, while outgoing partons are labelled 3 and 4. Incoming partons collide and rebound off of each other at some angle θ, as shown in Figure 3.2. Since the protons are made up of many partons each

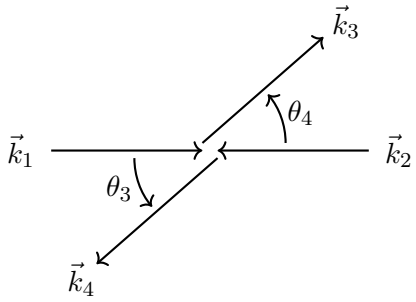


Figure 3.2: Two-parton scattering in momentum space in the centre-of-mass frame.

parton can only carry some fraction x of the proton's energy and momentum

$$x_a = \frac{|\vec{k}_1|}{|\vec{p}_a|}, \qquad x_b = \frac{|\vec{k}_2|}{|\vec{p}_b|}. \tag{3.4}$$

To describe interactions between partons we will use Mandelstam variables with carets

$$\hat{s} = (k_1 + k_2)^2, \qquad \hat{t} = (k_1 - k_3)^2, \qquad \hat{u} = (k_1 - k_4)^2. \tag{3.5}$$

Here k_i are the external four-momenta and $\hat{s} + \hat{t} + \hat{u} = m_1^2 + m_2^2 + m_3^2 + m_4^2$. We are interested in high energy collisions where the momenta of the particles involved are much larger than their masses. In this case the masses of partons can be neglected, giving $\hat{s} + \hat{t} + \hat{u} = 0$. The top quark

mass is approximately 173 GeV [11] and cannot be neglected in TeV collisions but since quarks of this flavour are not found within protons the approximation remains valid for our analysis.

Furthermore, we will define the invariant mass M of a collision as

$$\hat{s} = M^2. \tag{3.6}$$

The transverse momentum of outgoing particles is defined as

$$p_T = \sqrt{p_x^2 + p_y^2}, \tag{3.7}$$

which is invariant with respect to Lorentz boosts along the beam axis. Rapidity is defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \tag{3.8}$$

If the particle is massless then this is related to the angle θ that the particle's three-momentum makes with the beam axis by

$$y = \frac{1}{2} \ln \frac{1 + \cos\theta}{1 - \cos\theta} = -\ln \tan \frac{\theta}{2}. \tag{3.9}$$

We can then define pseudorapidity as

$$\eta = -\ln \tan \frac{\theta}{2}, \tag{3.10}$$

such that rapidity and pseudorapidity have the same value for massless particles. Particles close to the beam in the positive direction have a large positive rapidity and particles with three-momentum only transverse to the beam have zero rapidity. The rapidity of each particle shifts by a constant under boosts along the beam axis, making differences in rapidities boost-invariant observables [12].

We can express the transverse momentum of outgoing particles in terms of rapidity as

$$p_T = \frac{M}{2 \cosh y}. \tag{3.11}$$

Outgoing partons then have energies

$$E_3 = p_T \cosh y_3, \qquad E_4 = p_T \cosh y_4, \tag{3.12}$$

and three-momenta

$$\begin{aligned} p_{x_3} &= p_T \cos\theta_3, & p_{x_4} &= p_T \cos\theta_4, \\ p_{y_3} &= p_T \sin\theta_3, & p_{y_4} &= p_T \sin\theta_4, \\ p_{z_3} &= p_T \sinh y_3, & p_{z_4} &= p_T \sinh y_4. \end{aligned} \tag{3.13}$$

3.2 Parton Distribution Functions

In order to understand pp collisions at the LHC we must discuss parton distribution functions (PDFs). As mentioned in the previous section protons are made up of partons so when protons collide at high energies the interactions that occur are between these partons. Each parton carries some fraction x of the incident proton's momentum. The distribution of the momentum fraction of each parton is given by the PDF. The parton density function $f_i(x, Q^2)$ gives the probability of finding a parton of flavour i with a fraction x of the proton's momentum, Q is the energy scale of the hard interaction [13].

QCD does not predict the parton content of the proton so the shape of the PDFs must be determined experimentally. Different groups of PDFs are obtained from various sources, for example CTEQ PDFs [14] are obtained from analysis of hard scattering data, HERAPDFs [15] are based on electron-proton scattering data obtain by HERA, and Neural Net (NN) PDFs [16] are created by training a neural network with Monte Carlo (MC) replicas of experimental data. Depending on how the data is parametrised different PDFs can be constructed [17]. The CTEQ6L1 [14] PDF set that is typically used in ATLAS gravity analyses is shown in Figure 3.3. As seen there this PDF does not include top t or anti-top $\bar{t}$ quarks. Due to lack of data the

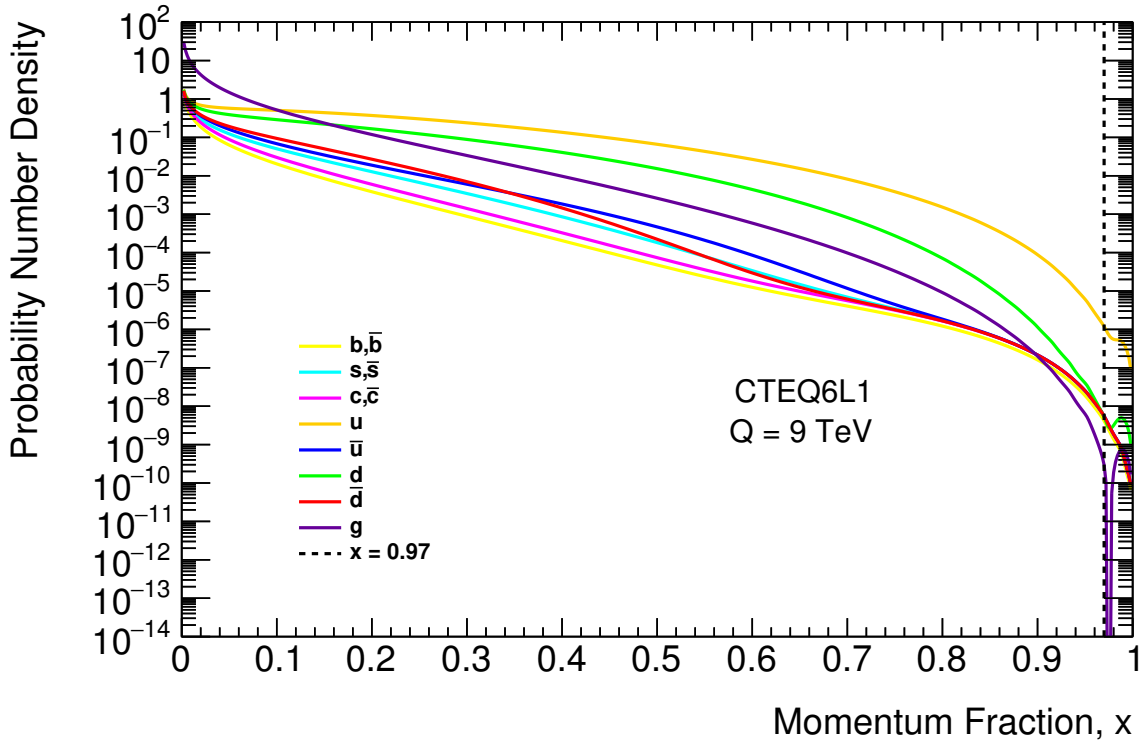


Figure 3.3: Parton distribution function set CTEQ6L1 [14] at the energy scale $Q = 9$ TeV. The dashed line at $x = 0.97$ indicates where the discontinuity begins.

PDFs in high energy regions must be extrapolated by fitting low energy regions. There are discontinuities in the CTEQ6L1 PDF set that begin at $x \approx 0.97$. The highest string scale that we simulate is $Q = M_s = 9.0$ TeV, compared to proton energies of 13 TeV generated by ATLAS. The highest value of x that we encounter in simulations is approximately 0.95 so the discontinuities in the very high x region of the CTEQ6L1 PDF set do not have any adverse

effect on our analysis. Other PDF sets exhibit similar behaviour and the CTEQ6L1 PDF set is smoother than other alternatives so it is the standard choice for many ATLAS analyses.

3.3 Scattering Amplitudes and Cross Sections

Within pp collisions there are many different interactions that can occur between the various partons. These involve different particles in the initial and final states, for example, a gluon and a quark can collide and rebound off of each other or two gluons can collide and create a quark and antiquark. Some interactions occur more frequently than others and the measure of likelihood of a certain interaction taking place is called the cross section σ. The cross section is the effective area for a collision, analogous to the size of a target that we are attempting to hit with a beam. For elementary particle collisions σ is of a similar order of magnitude as the size of particles involved. It is measured in units of area where 1 barn (b) is equal to 10^{-28} m^2. Cross sections for specific interactions depend on the properties of the particles involved such as their energy and charge. In order to determine the cross section of any given process we must first calculate its amplitude $\mathcal{M}$. This depends on the energies of incoming particles and is the probability amplitude of a certain interaction. The cross section for each interaction is then a function of its amplitude and the different states that the particles involved can be in. We will estimate scattering amplitudes using QCD below. They may be calculated using different methods in other theories such as string theory, but here we will follow the rules of the SM.

Here we will only consider scattering events between two partons. A simple example of this is two-gluon scattering, which is depicted in the Feynman diagram in Figure 3.4. Two gluons enter and they exchange some particle with momentum ξ, then two gluons exit. When we detect

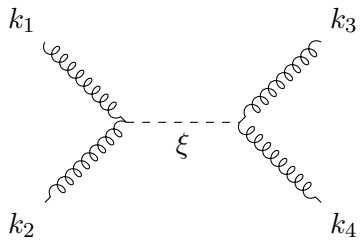


Figure 3.4: Feynman diagram of an interaction between two gluons. Time reads left to right. Two incoming gluons annihilate and some particle with momentum ξ if produced, then two gluons exit.

this process we can only measure the particles that are at the vertices of the diagram, internal lines represent particles that are not observed. These are called virtual particles, while external lines represent real particles. The external lines determine what process is occurring, while the internal lines describe the mechanism by which it occurs. Energy and momentum along with quantum numbers such as spin and charge must be conserved at each vertex. As a consequence virtual particles can have any mass but real particles must have their proper mass.

To calculate the amplitude associated with the Feynman diagram in Figure 3.4 we will follow the rules of Feynman calculus. For simplicity we will only include here terms depending on the energy and momenta of each of the particles involved. Other terms to account for the possible spins, colours, and flavours of particles will be discussed later on. First, we require a factor of

g for each vertex, called the strong coupling. This is a number that determines the strength of the strong interaction and depends on the energies of particles. Secondly, we include a factor called a propagator for each internal line

$$\frac{1}{\xi^2 - m^2}, \tag{3.14}$$

where m is the mass of the virtual particle. m is constant for any interaction and depends on the energies and momenta of the incoming and outgoing particles. To conserve four-momentum we add a delta function for each vertex

$$\delta^4(k_1 + k_2 - \xi), \tag{3.15}$$
$$\delta^4(\xi - k_3 - k_4), \tag{3.16}$$

and integrate over internal momenta. This leaves us with the equation

$$\mathcal{M} \sim g^2 \int \frac{1}{\xi^2 - m^2} \delta^4(k_1 + k_2 - \xi)\delta^4(\xi - k_3 - k_4) d^4\xi. \tag{3.17}$$

The second delta function in Equation 3.17 specifies that $\xi = k_3 + k_4$, which gives

$$\mathcal{M} \sim g^2 \frac{1}{(k_3 + k_4)^2 - m^2} \delta^4(k_1 + k_2 - k_3 - k_4). \tag{3.18}$$

The remaining delta function signifies that momentum is conserved since $k_1 + k_2 = k_3 + k_4$. Using this relation the kinematic portion of the amplitude of Figure 3.4 has

$$\mathcal{M} \sim \frac{g^2}{(k_1 + k_2)^2 - m^2}. \tag{3.19}$$

Using the Mandelstam variables defined in Equation 3.5 we can rewrite this as

$$\mathcal{M} \sim \frac{g^2}{\hat{s}^2 - m^2}. \tag{3.20}$$

The process associated with this Feynman diagram is said to occur through the s-channel. This is not the only Feynman diagram that can describe two-gluon scattering, in order to calculate the amplitude of a particular process we must add up the contributions of all possible diagrams. If we rotate Figure 3.4 by 90 degrees we obtain Figure 3.5 that describes the same process via a different mechanism. Two gluons enter and exchange some particle with momentum ξ and mass m, then two gluons exit. The virtual particle in this interaction is not necessarily the same as in the s-channel process. If we follow the Feynman rules as we did above for this diagram we

Figure 3.5: Feynman diagram of a t-channel interaction between two gluons.

obtain the amplitude

$$\mathcal{M} \sim \frac{g^2}{\hat{t}^2 - m^2}. \tag{3.21}$$

This Feynman diagram represents a t-channel process. Similarly, Figure 3.6 shows a u-channel

process. We can imagine many more Feynman diagrams with the same external lines but many

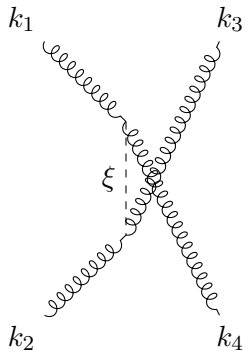


Figure 3.6: Feynman diagram of a u-channel interaction between two gluons.

more vertices and complicated loops such as that shown in Figure 3.7. It turns out that the

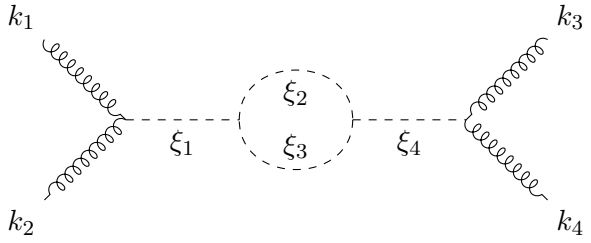


Figure 3.7: Feynman diagram of an interaction between two gluons with one loop.

coupling of the strong force becomes weak at short distances, leading to a phenomenon called asymptotic freedom. This means that interactions between particles that are close together including diagrams with many loops will have very small amplitudes. We will ignore those contributions here and consider only diagrams without loops, called tree-level diagrams.

Typically the cross section is described in terms of the number of particles that scatter into some range of angles in space. In this case the differential cross section

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{|\sin\theta d\theta d\phi|}, \tag{3.22}$$

is the rate at which particle are scattered in a given direction, per unit of solid angle. If we have an incoming beam of particles then the luminosity L is the number of particles travelling through the beam per unit time per unit area. The number of particles per unit time passing through area $d\sigma$ is $dN = Ld\sigma$. This is also the number of particles per unit time scattering into solid angle $d\Omega$. In other words the cross section multiplied by the luminosity gives the event rate [18]. The scattering amplitude represents the probability density of an interaction. For two-parton scattering it is related to the differential cross section by

$$\frac{d\sigma}{d\Omega} = \frac{|\mathcal{M}(12 \to 34)|^2}{64\pi^2 \hat{s}}, \tag{3.23}$$

where $\mathcal{M}(12 \to 34)$ is the amplitude of the particular process [19]. As described above, it must be computed by summing up the contributions of all possible Feynman diagrams. We are

interested in QCD processes where the spins, colours, and flavours of particles can vary. The scattering amplitudes must include factors to account for these variations. Each particle has a spinor that represents its spin. This is a vector with four indices that can vary depending on the spin of a particle, and whether it is a particle or antiparticle. Spinors for initial and final states of each particle must be combined by taking their inner product. Since we cannot differentiate between particles of the same flavour there is a factor within each amplitude to account for the possible permutations of spinor inner products. We must also include a factor associated with colour in QCD amplitudes. Quarks can only carry one colour so their colour factor is a vector like the spinor. Gluons can have more than one colour simultaneously so their colour representation takes the form of a matrix rather than vector. We multiply the different matrices together and take their trace to obtain the gluon colour factor. Each of the six quark flavours has its own associated scattering amplitude and the overall amplitude of a process involving quarks is the sum of these. In most experiments the states of initial particles are random so we can assume that there are equal numbers of each type of parton with each possible spin, colour, and flavour. We want to measure the number of particles scattered in a certain direction or with a certain energy, regardless of its state. Each spin, colour, and flavour combination will have a different amplitude so in order to calculate the cross section we must average all of the amplitudes over initial spin, colour, and flavour configurations and sum over final spin, colour, and flavour configurations [18].

We must take into account that the partons involved in two-parton scatterings in colliders originate from protons. Scattering amplitudes assume equal opportunity for any type of parton to interact. This is not the case since all types of partons are not equally abundant within protons. The likelihood of certain interactions is then not just dependent on the cross section but also depends on the likelihood of any given flavour of parton being present in a proton. This is achieved by convoluting the partonic cross section with the PDFs of the two colliding protons:

$$d\sigma = \sum_{1234} \int\int dx_a\, dx_b\, f_1(x_a, M) f_2(x_b, M) d\tilde{\sigma}(12 \to 34), \tag{3.24}$$

where $d\tilde{\sigma}$ is the differential cross section for each partonic process and $f(x, M)$ are parton density functions. This is summed over the possible interactions between partons. A given parton is equally likely to originate from either end of the beam so if multiple flavours of partons are included in an interaction we must sum up the contributions from either case. For example, for four identical partons there is only one term but for differentiable incoming or outgoing partons there are more terms. Summing over x_a and x_b is not necessary since this is already included by allowing incoming partons to originate from either proton.

In the parton centre-of-mass frame both partons rebound at the same angle θ so their rapidities have the same value but opposite sign, $y_3 = -y_4$. If we wish to boost from the parton centre-of-mass frame to the proton centre-of-mass frame the rapidities of outgoing particles can be written as

$$y_3 = y + Y, \qquad y_4 = -y + Y. \tag{3.25}$$

Rearranging for y and Y gives

$$y = \frac{1}{2}(y_3 - y_4), \qquad Y = \frac{1}{2}(y_3 + y_4). \tag{3.26}$$

We also define

$$\tau = \frac{M}{s}. \tag{3.27}$$

In terms of these new variables the momentum fractions of incoming partons 1 and 2 are

$$x_a = \sqrt{\tau}e^{Y}, \qquad x_b = \sqrt{\tau}e^{-Y}, \tag{3.28}$$

and the Mandelstam variables are

$$\hat{s} = M^2, \qquad \hat{t} = -\frac{M^2}{2}\frac{e^{-y}}{\cosh y}, \qquad \hat{u} = -\frac{M^2}{2}\frac{e^{y}}{\cosh y}. \tag{3.29}$$

We will study the energy of partonic scattering so the differential cross section with respect to the mass M of the collision is the parameter or interest. This is given by [19]

$$\frac{d\sigma}{dM} = M\sum_{1234}\int\int dY\, dy\, x_a\, x_b\, f_1(x_a, M) f_2(x_b, M)\frac{|\mathcal{M}(12 \to 34)|^2}{16\pi^2 \hat{s}}\frac{1}{\cosh^2 y}, \tag{3.30}$$

$$= M\tau\sum_{1234}\int\int dY\, dy\, f_1(x_a, M) f_2(x_b, M)\frac{|\mathcal{M}(12 \to 34)|^2}{16\pi^2 \hat{s}}\frac{1}{\cosh^2 y}. \tag{3.31}$$

If we set an upper bound on the absolute value of the outgoing rapidities $|y_3|, |y_4| < y_{max}$ and use the requirement $x_1, x_2 < 1$ then the upper bound on Y is

$$Y_{max} = \min\left\{\ln\left(\frac{1}{\sqrt{\tau}}\right), y_{max}\right\}. \tag{3.32}$$

The differential cross section becomes [20]

$$\frac{d\sigma}{dM} = M\tau\sum_{1234}\left(\int_{-Y_{max}}^{Y_{max}} dY\, f_1(x_a, M) f_2(x_b, M)\int_{-(y_{max}-|Y|)}^{y_{max}-|Y|} dy\, \frac{|\mathcal{M}(12 \to 34)|^2}{16\pi\hat{s}^2}\frac{1}{\cosh^2 y}\right). \tag{3.33}$$

3.4 String Phenomenology

The amplitude in Equation 3.20 is not a smooth function, it is undefined at the point $\hat{s} = m^2$, called a pole. As we approach this point the amplitude of this process becomes very large. This is associated with a resonance in Regge theory [21]. A Regge state is an intermediate state that is formed in the s-channel during two-parton scattering. These states occur at energies that correspond to a pole in the scattering amplitude of the interaction. In QCD this leads to resonance production when the collision energy approaches the mass of the intermediate state. In string theory string resonances in the form of Regge excitations can occur in pp collisions if the intermediate state in Figure 3.4 is exchanged as an excited string. In this theory SM particles make up the ground state of strings and higher energy states of strings can be observed as excited particles [20].

In various string theories the scattering amplitudes of partonic processes can exhibit resonant behaviour through Regge, Kaluza-Klein, or Winding states. Kaluza-Klein and Winding states appear in amplitudes with four or more quarks. Kaluza-Klein states arise from the compactification of extra dimensions and Winding states are created when objects such as strings or Dp-branes wrap around the extra dimensions. These states carry internal charge so by charge conservation they require at least two pairs of quarks. Regge states are purely string states and do not depend on the compactification of the extra dimensions [22]. Gluon scattering amplitudes involve a single stack of Dp-branes so momentum components along the compactified Dp-brane directions are conserved. As a consequence the poles of these amplitudes arise from the exchange of massless gluons and string Regge excitations only. Conversely, amplitudes involving four quarks depend on the Dp-brane geometry, which can vary between string theories [3]. Scattering between four gluons or two gluons and two quarks is model-independent and consistent for all types of compactification. For this reason we will ignore interactions involving four quarks and consider only those that occur in the LHC that are model-independent with two or more gluons. There are five such interactions:

$$gg \to gg,$$
$$gg \to q\bar{q},$$
$$gq \to gq,$$
$$g\bar{q} \to g\bar{q},$$
$$q\bar{q} \to gg.$$

By measuring the substructure of interactions in the collider we are able to trace with some probability whether each collision includes quarks or gluons [23]. The data that we use does not have any exclusionary criteria based on the number of quarks or gluons but we are able to analyse it for quark/gluon content. Four out of five of the interactions that we are interested in have at least one gluon in the final state and all of the ones that we would like to ignore have zero gluons in the final state. The portion of events with at least one gluon in the final state is then a good estimator of the fraction of events we can consider. We find that approximately 68% of our data is made up of events with at least one gluon. If we include collisions of the form $gg \to q\bar{q}$ then the percentage of events that we can consider becomes even larger. Fortunately, model-independent events make up a large fraction of our data so the analysis will not be hindered by this choice of criteria.

In QCD complete scattering amplitudes must take into account the many different configurations of particles that can occur within a single interaction. Quarks and gluons have quantum numbers that vary between individual particles. Amplitudes require factors to account for the different colour, spin, and flavour combinations of quarks and gluons. String theory amplitudes are different from QCD but they must obey the same rules such as conservation of four-momentum, spin, colour, and quark flavour. We will consider two-parton scattering as

described by Figure 3.2. The complete amplitude of four gluon interactions in this theory is

$$\begin{aligned}\mathcal{M}(g_1,g_2,g_3,g_4) = 4g_s^2\langle 12\rangle^4\Bigg[&\frac{V_t}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle}\mathrm{Tr}\left(T^{a_1}T^{a_2}T^{a_3}T^{a_4}+T^{a_2}T^{a_1}T^{a_4}T^{a_3}\right)\\ &+\frac{V_u}{\langle 13\rangle\langle 34\rangle\langle 42\rangle\langle 21\rangle}\mathrm{Tr}\left(T^{a_2}T^{a_1}T^{a_3}T^{a_4}+T^{a_1}T^{a_2}T^{a_4}T^{a_3}\right)\\ &+\frac{V_s}{\langle 14\rangle\langle 42\rangle\langle 23\rangle\langle 31\rangle}\mathrm{Tr}\left(T^{a_1}T^{a_3}T^{a_2}T^{a_4}+T^{a_3}T^{a_1}T^{a_4}T^{a_2}\right)\Bigg].\end{aligned} \tag{3.34}$$

Here g_s is the string coupling constant, angular brackets represent spinor inner products, and T_i^a are generators of the $SU(3)$ colour group. We define Veneziano factors

$$V_t = V(\hat{s},\hat{t},\hat{u}), \qquad V_u = V(\hat{t},\hat{u},\hat{s}), \qquad V_s = V(\hat{u},\hat{s},\hat{t}), \tag{3.35}$$

as functions of the Mandelstam variables

$$V(\hat{s},\hat{t},\hat{u}) = \frac{\hat{s}\hat{u}}{(\hat{s}+\hat{u})}B(\hat{s},\hat{u}). \tag{3.36}$$

The Euler beta function is defined as

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt. \tag{3.37}$$

In the low energy limit $\hat{s},\hat{t},\hat{u} << 1$

$$V_t = 1+\frac{\pi^2}{6}\hat{s}\hat{u}+\ldots \tag{3.38}$$

the Veneziano factors reduce to unity so the first order terms have no effect on SM amplitudes.

To obtain the cross section of an interaction from its amplitude we must take the moduli squared of each amplitude, sum over final polarisations and colours, and average over initial polarisations and colours. Taking the modulus square of Equation 3.34 summed over initial and final colours is

$$\begin{aligned}|\mathcal{M}(gg\to gg)|^2 = g_s^4(N^2-1)\hat{s}^4&\\ \times\left[2N^2\left(\frac{V_t^2}{\hat{s}^2\hat{u}^2}+\frac{V_s^2}{\hat{t}^2\hat{u}^2}+\frac{V_u^2}{\hat{s}^2\hat{t}^2}\right)+\frac{4(3-N^2)}{N^2}\left(\frac{V_t}{\hat{s}\hat{u}}+\frac{V_s}{\hat{t}\hat{u}}+\frac{V_u}{\hat{s}\hat{t}}\right)^2\right],\end{aligned} \tag{3.39}$$

where N is the number of colours. The amplitude squared, summed over final polarisations and colours, and averaged over all $4(N^2-1)^2$ initial polarisation/colour configurations is

$$\begin{aligned}|\mathcal{M}(gg\to gg)|^2 = g_s^4\left(\frac{1}{\hat{s}^2}+\frac{1}{\hat{t}^2}+\frac{1}{\hat{u}^2}\right)&\\ \times\left[\frac{2N^2}{N^2-1}(\hat{s}^2V_s^2+\hat{t}^2V_t^2+\hat{u}^2V_u^2)+\frac{4(3-N^2)}{N^2(N^2-1)}(\hat{s}V_s+\hat{t}V_t+\hat{u}V_u)^2\right].\end{aligned} \tag{3.40}$$

Following the same procedure for the other model-independent scattering amplitudes and in-

serting the number of colours $N = 3$ and quark flavours $N_f = 6$ yields [3]

$$|\mathcal{M}(gg \to gg)|^2 = g_s^4 \left(\frac{1}{\hat{s}^2} + \frac{1}{\hat{t}^2} + \frac{1}{\hat{u}^2}\right)\left[\frac{9}{4}(\hat{s}^2 V_s^2 + \hat{t}^2 V_t^2 + \hat{u}^2 V_u^2) - \frac{1}{3}(\hat{s}V_s + \hat{t}V_t + \hat{u}V_u)^2\right], \tag{3.41}$$

$$|\mathcal{M}(gg \to q\bar{q})|^2 = g_s^4 \frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\left[\frac{1}{6}\frac{1}{\hat{u}\hat{t}}(\hat{t}V_t + \hat{u}V_u)^2 - \frac{3}{8}V_t V_u\right], \tag{3.42}$$

$$|\mathcal{M}(q\bar{q} \to gg)|^2 = g_s^4 \frac{8}{3}\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\left[\frac{4}{9}\frac{1}{\hat{u}\hat{t}}(\hat{t}V_t + \hat{u}V_u)^2 - V_t V_u\right], \tag{3.43}$$

$$|\mathcal{M}(gq \to gq)|^2 = |\mathcal{M}(g\bar{q} \to g\bar{q})|^2 = g_s^4 \frac{\hat{s}^2 + \hat{u}^2}{t^2}\left[V_s V_u - \frac{4}{9}\frac{1}{\hat{s}\hat{u}}(\hat{s}V_s + \hat{u}V_u)^2\right]. \tag{3.44}$$

Particles created by the vibrations of relativistic strings create Regge states when the energy of partonic processes is greater than the string scale, satisfying $\hat{s} > M_s^2$. Below this threshold virtual Regge excitations can contribute to scattering processes. Interaction cross sections must then include massive string excitations [24]. Infinite poles occur in the s-channel at masses of the string Regge excitations

$$M_n^2 = nM_s^2. \tag{3.45}$$

The total string scattering amplitude is an infinite sum over these s-channel poles, each with an intermediate string state $|k; n\rangle$, as shown in Figure 3.8. The massless $n = 0$ state gives the SM

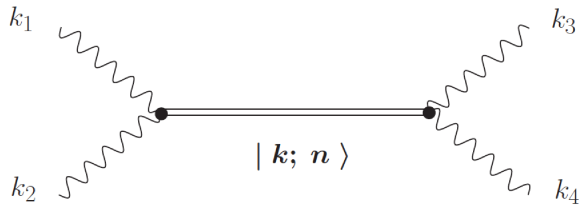


Figure 3.8: Feynman diagram of a tree-level scattering interaction involving a Regge state. Incoming particles have momenta k_1 and k_2 and exchange a Regge excitation with outgoing particles of momenta k_3 and k_4 [3].

contribution to this Feynman diagram, the exchange of a massless gluon. The other massive $n \neq 0$ terms sum to a finite term that describes the interactions resulting from heavy string states [3].

The resonant string theory scattering amplitudes can be obtained by factorising the polarised four-parton scattering amplitudes on kinematic poles at $\hat{s} = nM_s^2$, where Regge excitations occur [24]. n is an integer that represents the order of the resonance. To consider the poles of the amplitudes above we can take the limit $\hat{s} \to nM_s^2$ where

$$V_t \approx \frac{1}{\hat{s} - nM_s^2}\frac{M_s^{2-2n}}{(n-1)!}\prod_{J=0}^{n-1}(\hat{u} + M_s^2 J). \tag{3.46}$$

J is the spin of particles exchanged in the s-channel with $0 \leq J \leq n+1$. For simplicity we will consider here only $n = 1$, corresponding to the first and lowest-mass string Regge resonance. In

this case

$$V_t \approx \frac{\hat{u}}{\hat{s} - M_s^2}, \qquad V_u \approx \frac{\hat{t}}{\hat{s} - M_s^2}. \tag{3.47}$$

V_s is finite in this limit

$$V_s = \frac{\Gamma(1-\hat{t})\Gamma(1-\hat{u})}{\Gamma(1+n)}. \tag{3.48}$$

String theory amplitudes result in a singularity at the resonant energy since the terms in Equation 3.47 are undefined when $\hat{s} = M_s^2$. s-channel resonances must have a finite width so we can model the scattering amplitudes as a Breit-Wigner distribution. Since Equations 3.41-3.44 result from averaging over initial polarisations they contain contributions from both spin $J = 0$ and $J = 2$ Regge excitations due to incident gluons in $(\pm\pm)$ and $(\pm, \mp)$ helicity configurations, respectively. Resonance widths depend on both spin and identity of the intermediate states: excited gluon g^*, excited quark q^*, and excited colour singlet C^*. After softening the amplitudes we have

$$\begin{aligned}|\mathcal{M}_{1st}(gg \to gg)|^2 = \frac{19}{12}\frac{g^4}{M_s^4}\Bigg\{&\frac{25}{57}\left[\frac{M_s^8}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{g^*,1st}^{J=0})^2} + \frac{\hat{u}^4+\hat{t}^4}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{g^*,1st}^{J=2})^2}\right] \\ &+ \frac{32}{57}\left[\frac{M_s^8}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{C^*,1st}^{J=0})^2} + \frac{\hat{u}^4+\hat{t}^4}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{C^*,1st}^{J=2})^2}\right]\Bigg\},\end{aligned} \tag{3.49}$$

$$|\mathcal{M}_{1st}(gg \to q\bar{q})|^2 = \frac{7}{24}\frac{g^4}{M_s^4}\left[\frac{5}{7}\frac{\hat{u}\hat{t}(\hat{u}^2+\hat{t}^2)}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{g^*,1st}^{J=2})^2} + \frac{2}{7}\frac{\hat{u}\hat{t}(\hat{u}^2+\hat{t}^2)}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{C^*,1st}^{J=2})^2}\right], \tag{3.50}$$

$$|\mathcal{M}_{1st}(q\bar{q} \to gg)|^2 = \frac{56}{27}\frac{g^4}{M_s^4}\left[\frac{5}{7}\frac{\hat{u}\hat{t}(\hat{u}^2+\hat{t}^2)}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{g^*,1st}^{J=2})^2} + \frac{2}{7}\frac{\hat{u}\hat{t}(\hat{u}^2+\hat{t}^2)}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{C^*,1st}^{J=2})^2}\right], \tag{3.51}$$

$$\begin{aligned}|\mathcal{M}_{1st}(gq \to gq)|^2 &= |\mathcal{M}_{1st}(g\bar{q} \to g\bar{q})|^2 \\ &= \frac{4}{9}\frac{g^4}{M_s^2}\left[\frac{M_s^4(-\hat{u})}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{q^*,1st}^{J=1/2})^2} + \frac{(-\hat{u})}{(\hat{s}-M_s^2)^2 + (M_s\Gamma_{q^*,1st}^{J=3/2})^2}\right],\end{aligned} \tag{3.52}$$

where Γ_{g^*,q^*,C^*}^{J} are total decay widths of the first excited states of gluons, quarks, and the colour singlet, respectively. These are calculated to be [20]

$$\Gamma_{g^*,1st}^{J=0} = \frac{g^2}{4\pi}M_s\frac{3}{4}, \qquad \Gamma_{g^*,1st}^{J=2} = \frac{g^2}{4\pi}M_s\frac{9}{20}, \tag{3.53}$$

$$\Gamma_{q^*,1st}^{J=1/2} = \frac{g^2}{4\pi}M_s\frac{3}{8}, \qquad \Gamma_{q^*,1st}^{J=3/2} = \frac{g^2}{4\pi}M_s\frac{3}{16}, \tag{3.54}$$

$$\Gamma_{C^*,1st}^{J=0} = \frac{g^2}{4\pi}M_s\frac{3}{2}, \qquad \Gamma_{C^*,1st}^{J=2} = \frac{g^2}{4\pi}M_s\frac{3}{4}. \tag{3.55}$$

By inputting the PDF sets discussed in Section 3.2 and scattering amplitudes that we obtained above to Equation 3.33 we can calculate differential cross sections for model-independent pp scattering with massive string Regge excitations. These can then be integrated over M to obtain cross sections, which can be used to search for evidence of this theoretical model in data.

3.5 Jets

Jets are described as a highly collimated grouping of particles. They occur in many high energy collisions of elementary particles and are made up of various hadrons. In high energy pp collisions such as those at hadron colliders quarks and gluons are produced. Due to confinement individual quarks and gluons cannot exist at energies below approximately 150 MeV. When protons collide at high energies they are fragmented into their constituent quarks and gluons that collide with each other. Quarks at high energies can emit gluons, similar to electrons radiating photons. These gluons are concentrated in the directions of the partons in the primary collision, giving a highly collimated group of partons. In order to obey confinement it is energetically favourable to produce a quark-antiquark pair from the vacuum to restore a colourless state than for the singular gluon to exist on its own. These quarks may then radiate more gluons, resulting in a branching process. The partons lose energy as they radiate outwards, increasing the strength of interactions between them, until they cluster together to form hadrons detected as final state particles. The combination of partons into hadrons is called hadronisation. This entire process from collision to hadronisation occurs very quickly and over a short distance so we cannot measure any of the partons involved. We can only observe these final state hadrons so we rely on them for information on the interactions that created them. We do not have a good model for how hadronisation takes place but experimental evidence suggests that the process does not significantly impact the momenta of the partons involved. A highly energetic quark will result in a similarly energetic jet of hadrons [12]. We are interested in scattering events between two partons so we will focus on events with two jets, called dijet events. One such event recorded at the LHC is depicted in Figure 3.9.

Many properties of jets can be measured and each provides us with information on the interactions that produced them. The number of jets, their energies, and the type of particles that make them up can tell us different properties of the partonic interaction. If a particle decays to many partons that create jets then the sum of invariant masses of the jets should approximately match that of the original particle. This can be used to measure the masses of the partons in the primary collision. Measuring the energy and momentum of the hadrons that make up jets can provide information on the kinematic properties of the partons that created them. The shape of jets and multipicity of particular particles within them can be used to identify the flavour of partons involved in the primary collision [26]. Measuring the angle between emitted particles in a jet also gives us information on the type of parton that created it [12].

In most circumstances making testable predictions with QCD is extremely difficult, due to the infinite number of possible interactions that can occur during a single collision as described in Section 3.3. Over short distances this infinite number of terms can be approximated accurately by a finite number of terms, called perturbative QCD. The number of Feynman diagrams required to compute an amplitude grows factorially with the number of final state particles. For example, an amplitude for producing eight gluons in the final state results from the sum of more than one million Feynman diagrams. This makes calculations involving QCD processes very resource and time intensive so we usually limit terms to the lowest or second-lowest order, called loading order (LO) or next-to-leading order (NLO), respectively. LO calculations involve

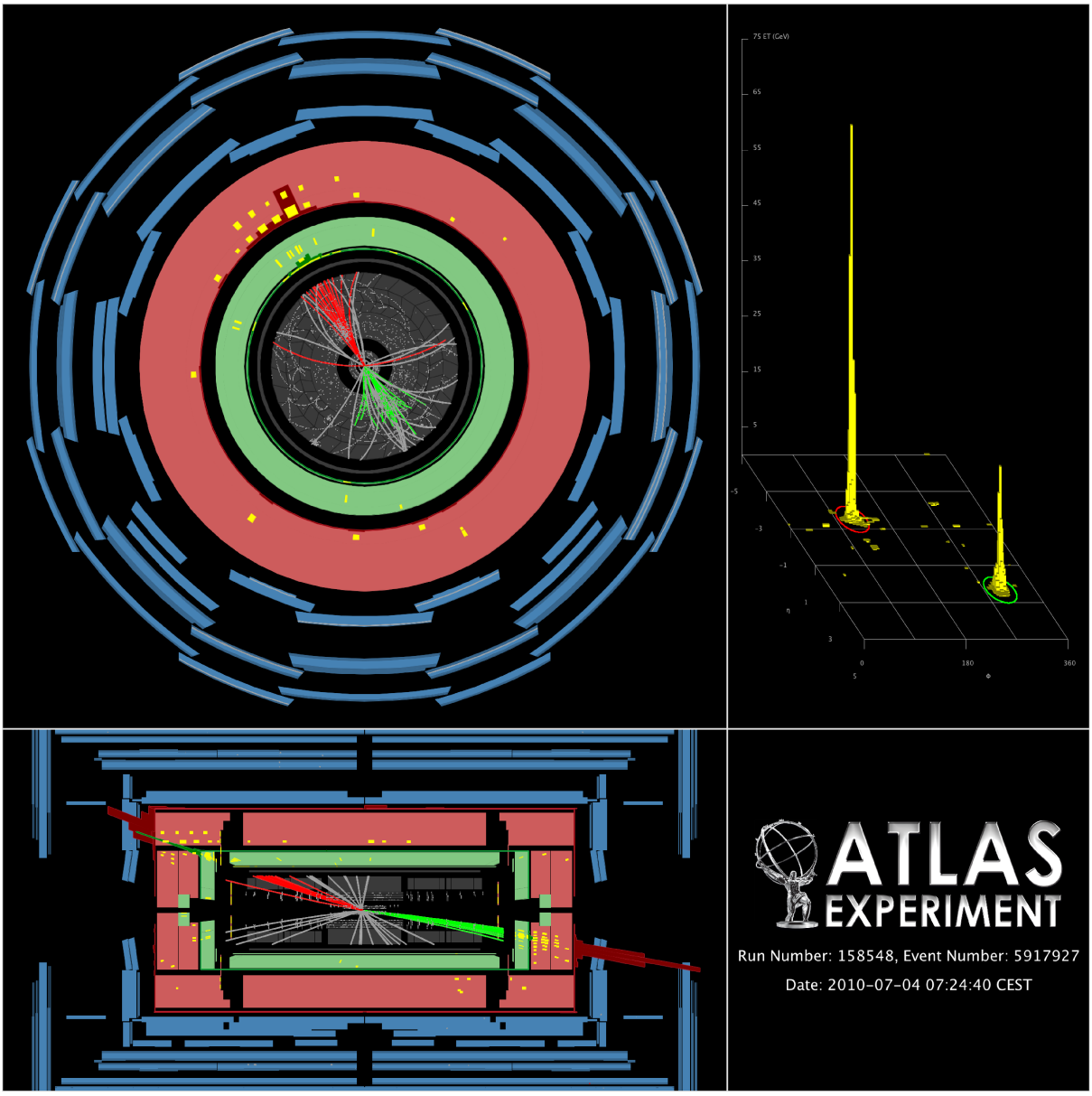


Figure 3.9: Dijet event recorded at the LHC. Top and side view of the detector are shown, and energy deposit of the jets. Individual tracks are grouped into jets and colour coded [25].

only tree-level Feynaman diagrams while NLO calculations include second order diagrams such as those with one loop. This has given successful approximations to many QCD processes and is experimentally verified. In order to use jets to study partonic collisions we must employ perturbative QCD. Comparisons of tree-level calculations with experimental data have given us information on the nature of quarks and gluons. For instance the existence of gluon self-interactions was confirmed from angular correlations between pairs of jets in four-jet events in electron-positron annihilation. The transverse momentum of the leading jet in events where a Z boson and one additional jet are created is shown in Figure 3.10. The LO QCD prediction

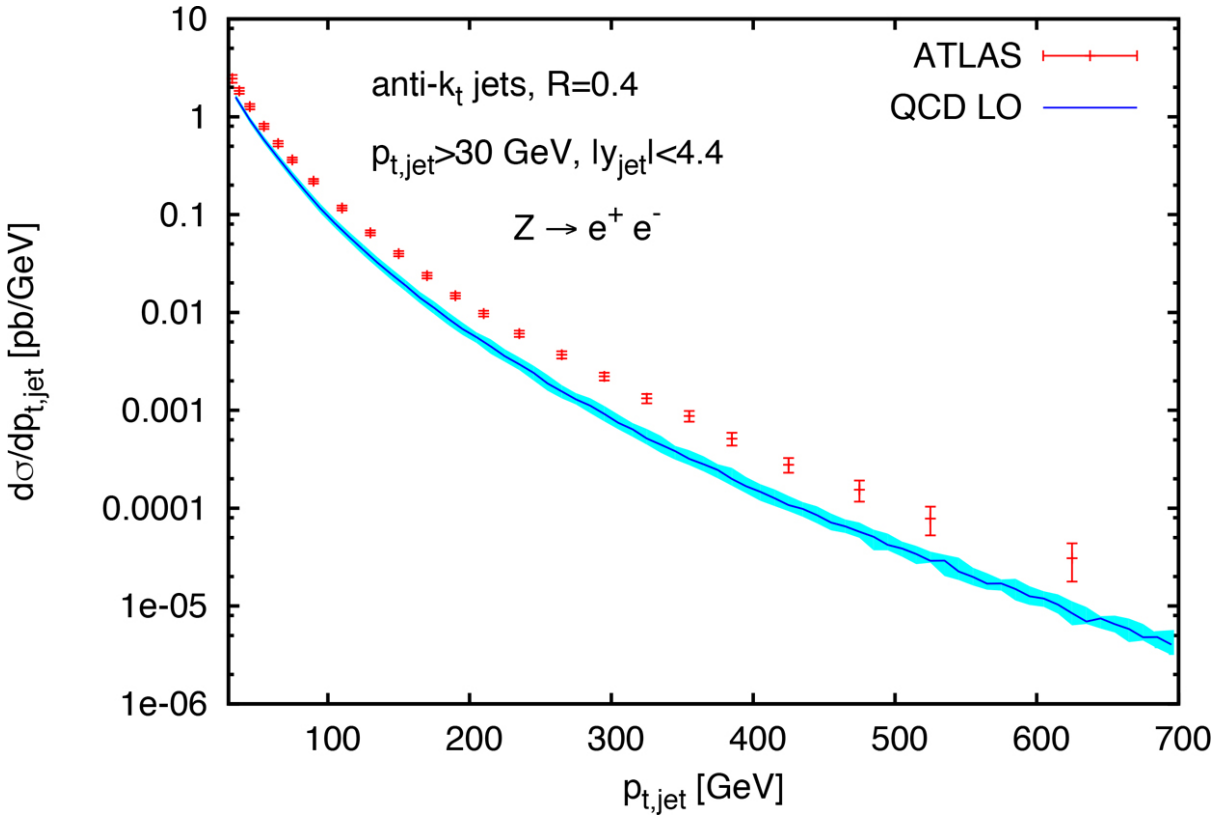


Figure 3.10: The distribution in the transverse momentum of the leading jet in Z plus one jet events measured by the ATLAS collaboration compared to a LO QCD prediction [12].

is close in shape to data but underestimates jet p_T over the entire range. This indicates that we need higher order terms to accurately describe the interactions. NLO computations of QCD interactions are much more accurate but the uncertainty of theoretical predictions remains approximately 10-20%. For many processes the NLO terms are similar in value to LO ones, suggesting a slow convergence of the QCD perturbative series. Precise predictions require even higher order corrections but presently no algorithm exists to consistently compute two-loop amplitudes [12].

4 Large Hadron Collider and ATLAS Detector

4.1 The Large Hadron Collider

The LHC is a two-ring-superconducting-hadron accelerator and collider at CERN near Geneva, Switzerland. The collider is built underground in a tunnel with a circumference of 36.7 km and a varying depth ranging between 45 and 170 m. It is built in eight arcs of length 2987 m connected by eight straight sections of length 528 m. The LHC is a particle-particle collider so there are two rings with counter-rotating beams that collide at each of the detectors. Out of the eight sections there are four that house intersection points where detectors are placed. The collider is capable of producing pp, p-Pb, and Pb-Pb beams and contains six different detectors [27]. The A Toroidal LHC ApparatuS (ATLAS) [28] and Compact Muon Solenoid (CMS) [29] detectors are large general-purpose detectors that are capable of capturing the highest luminosity collisions. The other experiments at the LHC are much smaller and study more specific phenomena such as the LHCbeauty (LHCb) [30] detector that studies decays of B hadrons and CP violation, and the LHCforward (LHCf) [31] detector that measures neutral particles emitted to the very forward region of collisions to study cosmic rays.

The LHC was built between the years 1998-2008. The first period of data acquisition (Run 1) from 2009-2013 reached total energies of 7 TeV with an approximate integrated luminosity 50 fb^{-1}, and 8 TeV with an approximate integrated luminosity 23 fb^{-1}. The collider was then shut down for upgrades from 2013-2015, after which Run 2 began. From 2015-2018 the LHC recorded data with centre-of-mass energy 13 TeV and reached an integrated luminosity of 139 fb^{-1}. Another shutdown took place from 2019-2020 and Run 3 will begin in 2021. The plan is to reach an integrated luminosity of 300 fb^{-1} at 14 TeV by the end of 2022. Another shutdown is planned between 2023-2025, which will upgrade the collider to the High-Luminosity LHC (HL-LHC). This aims to achieve luminosities as high as 4000 fb^{-1} by the late 2030s [32].

4.2 The ATLAS Detector

The ATLAS detector is 44 m long and 25 m tall, weighing approximately 7000 tonnes. An illustration of the entire ATLAS detector is shown in Figure 4.1. The detector is symmetric with respect to the beam axis and is comprised of many different components that detect various types of particles produced in collisions. There is an inner tracking cavity surrounded by a thin superconducting solenoid, a series of calorimeters, and three large superconducting toroids (one barrel and two end caps) with eight-directional azimuthal symmetry.

Approximately 1000 particles are created from a collision point every 25 ns with $|\eta| < 2.5$ within the detector. This means that very high resolution is required to record accurate momentum and vertex measurements. The inner detector is shown in Figure 4.2. It measures 6.2 m long and has a diameter of 2.1 m, and is made up of a combination of trackers. The precision tracking detectors are a series of discrete, high-resolution semiconductor trackers that cover the region $|\eta| < 2.5$. These are eight layers of silicon microstrip detectors in the centre, as well as three cylindrical layers of pixel detectors along the barrel and three pixel tracker disks on each end. The transition radiation tracker, which uses many layers of gaseous straw tubes covers $|\eta| < 2.0$. The inner detector is placed in a 2 T solenoidal field generated by the

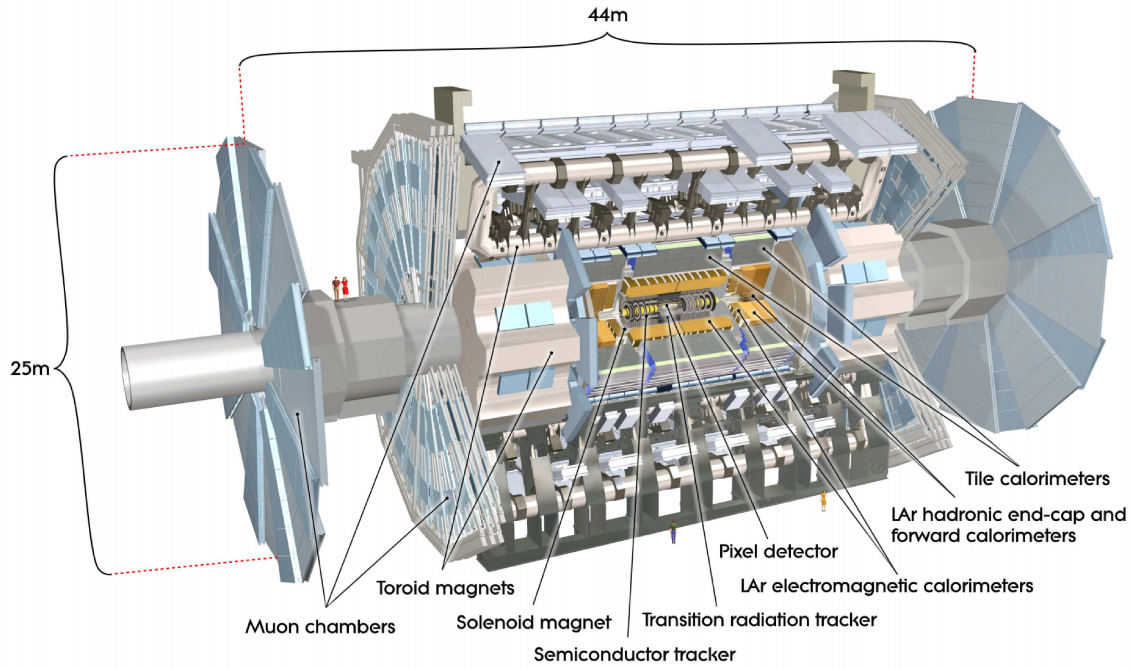


Figure 4.1: The ATLAS detector with major components indicated. The detector is 25 m tall and 44 m long and weighs approximately 7000 tonnes. From [28].

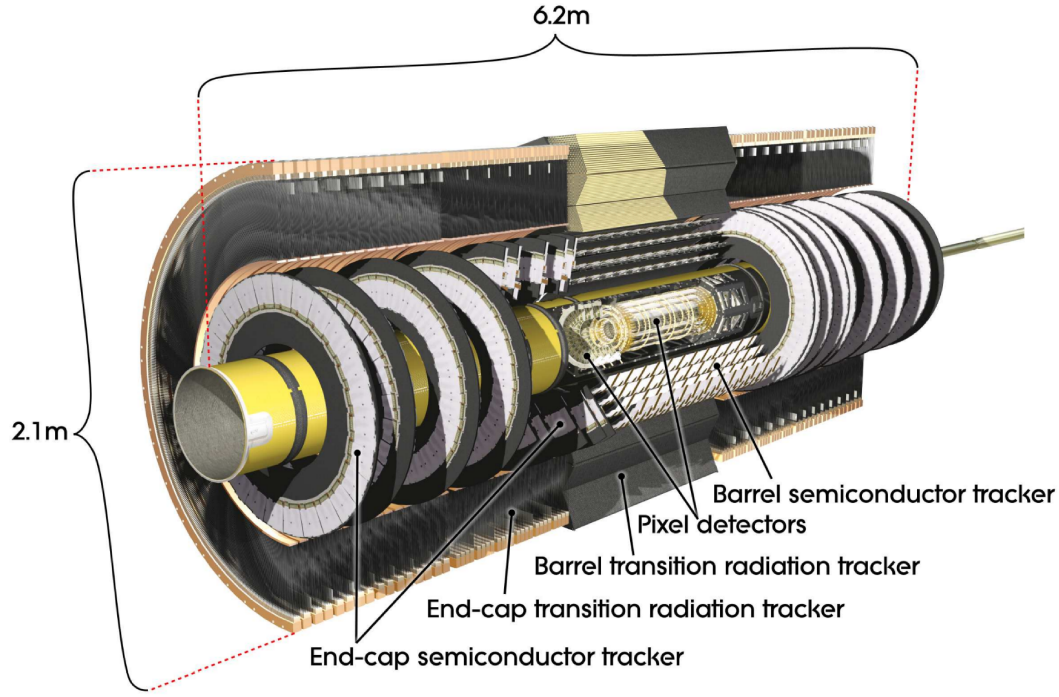


Figure 4.2: The ATLAS inner detector with important components indicated. This detector contains high resolution pixel and silicon strip detectors along the barrel and both ends. From [28].

central solenoid, which bends charged particles, allowing their momentum to be recorded by the detector.

Outside of the inner detectors lie a series of calorimeters that cover the range $|\eta| < 4.9$. The calorimeter system is shown in Figure 4.3. Their purpose is to record both electromagnetic

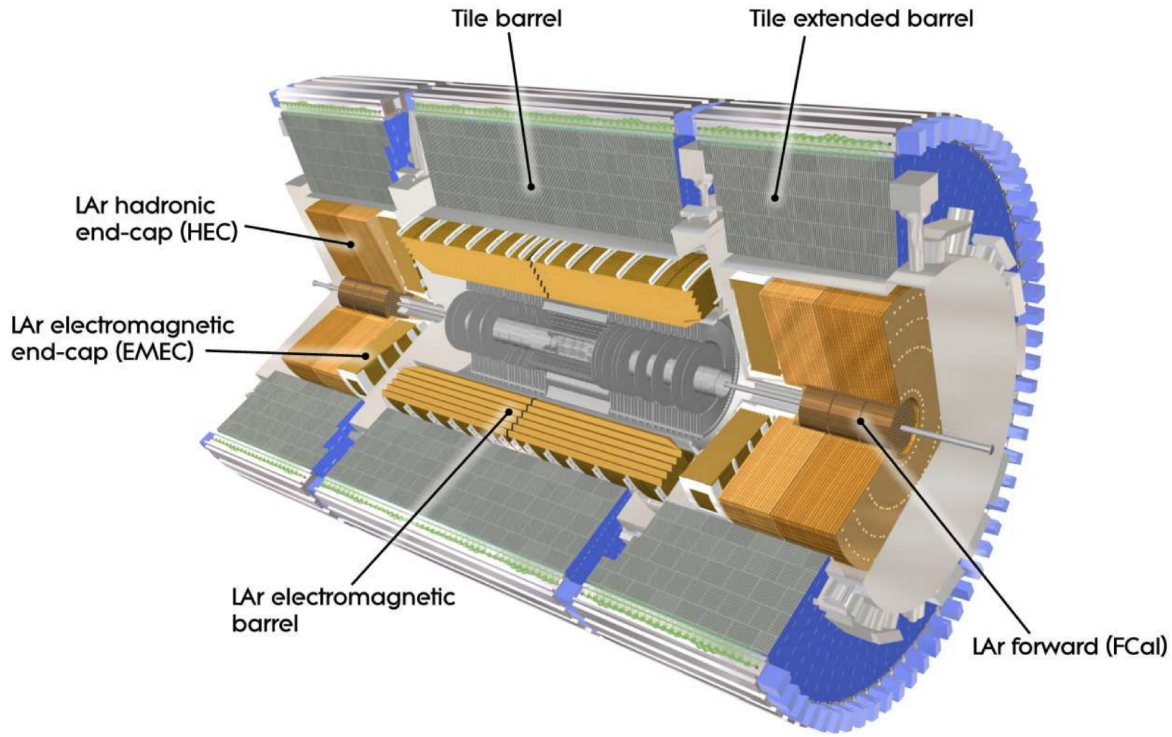


Figure 4.3: The ATLAS calorimeter system with important components indicated. This detector contains electromagnetic and hadronic calorimeters [28].

and hadronic processes, and to contain them as much as possible to reduce their impact on the outer muon system. Over the η region covered by the inner detector electromagnetic sampling calorimeters containing liquid-argon record precision measurements of the energy and position of incident electrons and photons. This is divided into a barrel part with $|\eta| < 1.475$ and end caps with $1.375 < |\eta| < 3.2$. Directly outside of this is a scintillator-tile hadronic calorimeter, separated into one large central barrel covering $|\eta| < 1.0$ and two smaller extended barrel cylinders on either side with range $0.8 < |\eta| < 1.7$. The two other components of the hadronic calorimetery system are the liquid-argon end-cap calorimeter and forward calorimeter. The hadronic end-cap calorimeter covers $1.5 < |\eta| < 3.2$. It is directly behind the electromagnetic calorimeter and is made up of two independent wheels per end, each divided in two for a total of four layers per end-cap. The forward calorimeter measures both electromagnetic and hadronic interactions and covers the region $3.1 < |\eta| < 3.9$.

The outermost layer around the calorimeters is the muon spectrometer, shown in Figure 4.4. It consists of large superconducting magnets to deflect muon tracks and high-precision tracking chambers to record measurements. A large barrel toroid acts over the range $|\eta| < 1.4$ while muon tracks for $1.6 < |\eta| < 2.7$ are bent by two smaller end-cap magnets. Over the region $1.4 < |\eta| < 1.6$ bending is provided by a combination of both barrel and end-cap magnets.

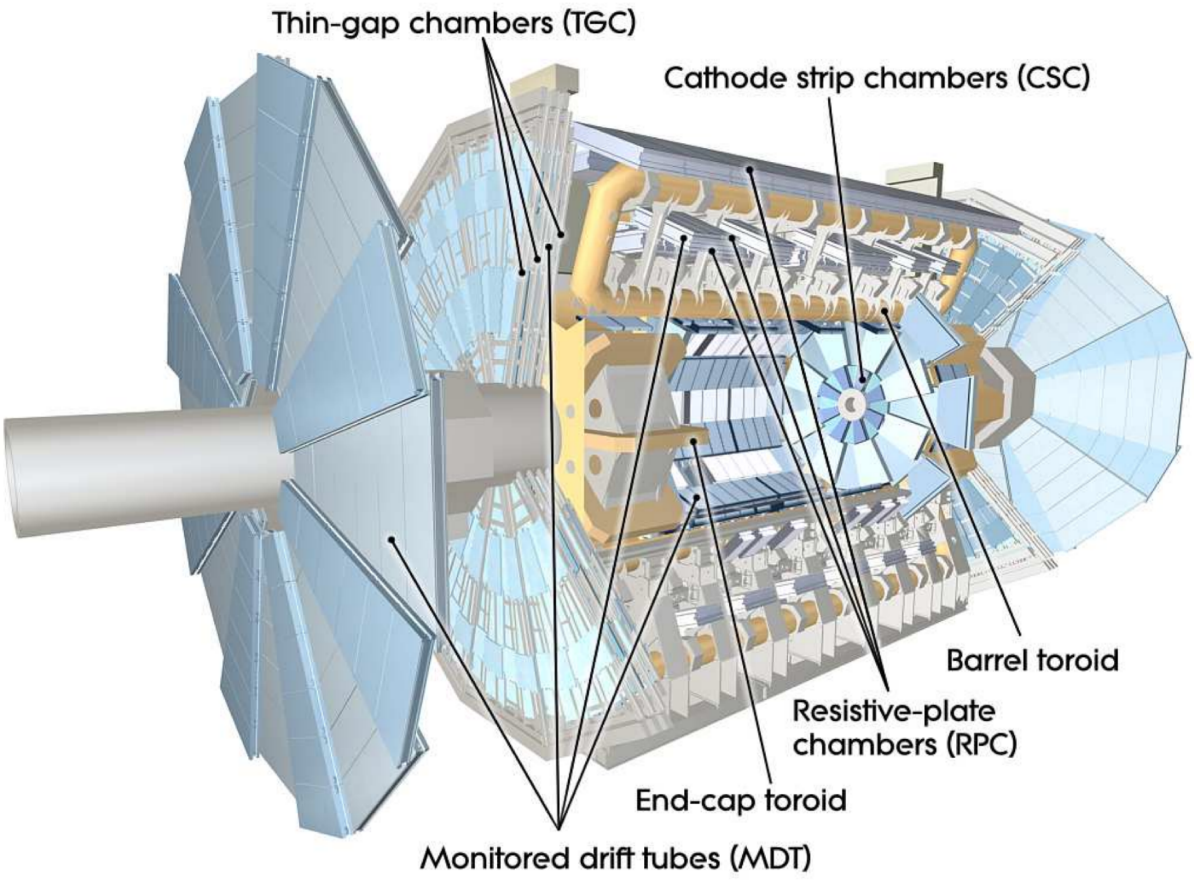


Figure 4.4: The ATLAS muon system with important components indicated. This detector consists of magnets that bend the muon tracks, and detectors to measure muon positions and momenta [28].

Precision measurements of the track coordinates are provided by monitor drift tubes in the range $|\eta| < 2.7$ while cathode strip chambers are utilised in the range $2.0 < |\eta| < 2.7$. The identification of muon tracks is done by the resistive plate chambers in the barrel and thin gap chambers in the end-cap regions [28].

ATLAS uses a coordinate system shown in Figure 4.5. The origin lies at the centre of the detector where collisions take place. The beam axis is defined as the z-axis and the xy plane lies perpendicular, known as the transverse plane. The positive x-axis points towards the centre of the LHC ring and the positive y-axis points upwards towards the surface of the earth. Cylindrical coordinates are often used where θ is the polar angle measured from the z-axis. The azimuthal angle ϕ is measured from the x-axis around the beam from 0 to $+\pi$ in the top half of the detector and 0 to $-\pi$ in the bottom half.

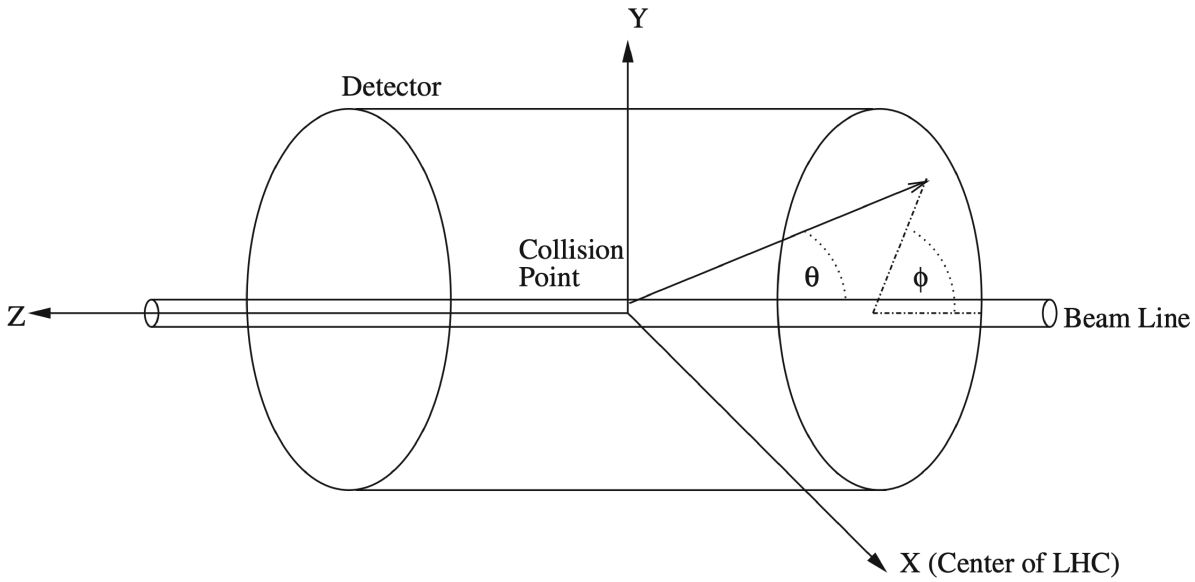


Figure 4.5: Coordinate system used by the ATLAS detector [33].

4.3 Data Acquisition

During data taking runs there are multiple pp collisions inside the ATLAS detector every 25 ns. These are collisions between bunches of approximately 50 protons, called a bunch-crossing. In each bunch-crossing there may be many pp collisions, called vertices. Jets are associated to the vertices that created them and vertices with jet energies much higher than others in the same bunch-crossing are called primary vertices. Each bunch-crossing may have one or more primary vertices. As described in Section 3.5 these protons fragment into partons that collide and create jets. We cannot measure the partons since they quickly combine into hadrons, which are deposited into the calorimeters. To reconstruct events we must use the information recorded by the calorimeters. The event rate is on the order of GHz so it is not possible to store all of this data due to limitations in sampling frequency and storage space. Furthermore, many collisions are at low energy and are not of interest. The Trigger and Data Acquisition (TDAQ) system decides in real time whether or not to record data from each event within the

detector [34]. Whenever a hard scattering collision happens there are many additional particles produced from soft scatters that make it difficult to reconstruct the hard scatter. This effect is called pileup, and can introduce significant noise into jet reconstruction [35]. Signals received from the calorimeters span multiple bunch-crossings since the sampling frequency of the trigger system is slower than the time between collisions. With so much information being measured by the detector it is important to record only events that are of particular interest for physics analysis. The trigger system aims to reduce the amount of noise in data by storing only data from high energy scattering events.

The trigger system contains three levels of filtering. The Level-1 (L1) trigger system uses information from certain components of the detector to decide whether or not to continue processing each event. The two other levels are the Level-2 (L2) trigger and event filter, combined they are called the High-Level Trigger (HLT). Each trigger level refines decisions made at the previous level and applies additional selection criteria, if required. The L1 trigger searches for decays of high transverse momentum muons, electrons, photons, jets, and τ-leptons into hadrons. It also searched for objects with large missing and total transverse energy. The L1 trigger identifies Regions-of-Interest (ROIs) in η and ϕ to be looked into by the second stage of the trigger process. The HLT uses ROI information on coordinates, energy, and types of signatures and applies further kinematic selections to reduce the amount of data recorded. The event filter processes events and categorises them according to ATLAS-defined data streams, for example: electrons, photons, or jets [34]. The naming convention of single-jet triggers follows either 'Jnnn' for L1 triggers or 'jnnn' for HLT trigger, where 'nnn' is a number specifying the nominal jet p_T threshold for the trigger in GeV.

4.4 Jet Reconstruction

After any event within ATLAS particles are deposited into the detector components, which measure the energy and location of incident particles. We are interested in events containing quarks and gluons so we will focus here on the reconstruction of jets from hadrons. Considering the event in Figure 4.6 it is not immediately evident how many jets are present, particularly looking along the beam line. This is classified as a four-jet event, with each jet energy grouped by colour in the bottom right panel. Each of the lines is called a track, and represents a particle or group of particles that creates an energy deposit in one of the calorimeters. These are generated by the process of quarks radiating gluons that subsequently fragment into more partons and branch into many jets described in Section 3.5. Jets are usually concentrated along the direction of the particle that created them but it is possible that two tracks separated by some angle were created by the same parent parton. We must determine how many jets are present and assign all tracks to one of the jets for each event within ATLAS. It is thus very important to establish a set of rules for identifying and grouping particles into jets, called a jet reconstruction algorithm.

In QCD the probability of a parton emitting a gluon with zero energy is infinite, referred to as soft divergence. In this case the term soft refers to a particle whose energy is much smaller than the typical energy of the other particles. This would seem problematic at first glance but it is fortunately counter-balanced by quantum fluctuations. A parton splitting into a pair of

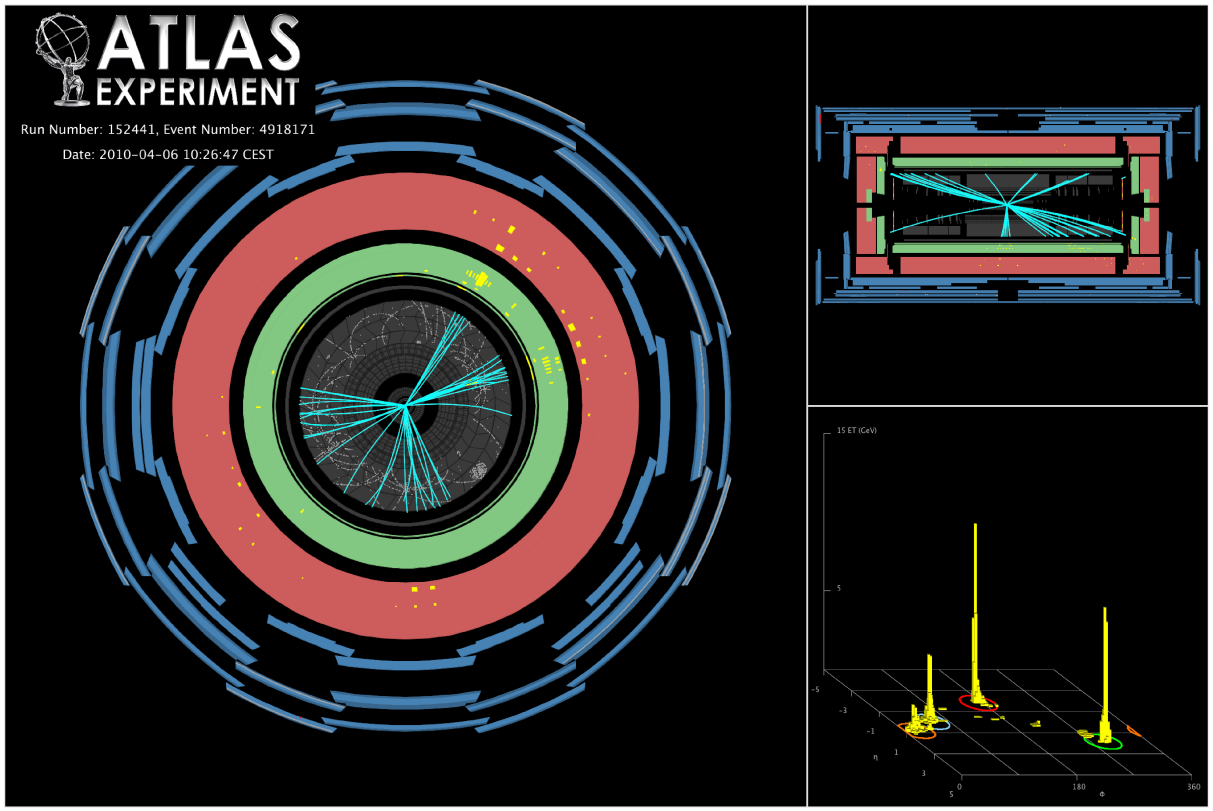


Figure 4.6: Event with four reconstructed hadronic jets recorded by the ATLAS detector at the LHC. Top and side view of the detector are shown, and energy deposits of the jets [25].

collinear partons gives rise to collinear divergence. The probability that a gluons is emitted and reabsorbed before it is observed is similarly infinite but with opposite sign. Correspondingly, we want to create our observable such that it is affected in the same way by a soft gluon and the corresponding quantum fluctuation to cancel the associated infinities and give a finite result. The most important property that a jet reconstruction algorithm must have is being infrared and collinear safe. Infrared safety means that adding an arbitrary number of soft partons will not change the number of jets or their momenta. Collinear safety ensures that if an arbitrary number of partons in the jet split into two collinear partons the number of jets and their momenta is unchanged. This ensures that jet reconstruction is not affected by either soft or collinear divergence. These types of algorithms aim to define jets the same way at the detector, hadron, and parton levels [12].

Calorimetric cells cannot resolve the energy deposit of single particles when many particles are incident. Therefore, jets must be defined using the transverse energy, pseudorapidity, and azimuthal angle of individual calorimetric cells as inputs, rather than the momenta of individual hadrons. Topological clusters [36] are formed from energy deposits in the hadronic calorimeters, which are then used as jet reconstruction inputs. Topological clusters are groups of neighbouring calorimeters cells that register significantly higher energy deposits than the expected noise. There are many jet reconstruction algorithms used at different colliders, ATLAS uses one called anti-k_t. We define the distance between an object i and the beam by

$$d_{iB} = p_{Ti}^{-2}, \tag{4.1}$$

where p_T is the transverse momentum of the object, which can be an individual particle, calorimetric cell, or jet. We can then define the distance between objects i and j as

$$d_{ij} = min\{d_{iB}, d_{jB}\}\frac{\Delta_{ij}^2}{R^2} = min\{d_{iB}, d_{jB}\}\frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2}, \tag{4.2}$$

where y is the rapidity and ϕ is the azimuthal angle of the calorimeter cluster, and R is a free parameter. To start the clustering process the smallest overall distance is calculated and if it is a d_{ij} then objects i and j are recombined, if it is a d_{iB} then i is labelled a jet and removed from the list of objects. This process is repeated until no more entries remain and all have been assigned to a jet [37].

Reference jets are obtained by applying the jet algorithm to stable particles that emerge from the fragmentation process and define the reference jet scale. The calorimeter responses depend on the type of particle incident, and is on average higher for electrons and photons than for hadrons. Particles lose energy as they travel through non-instrumented parts of the detector before they reach the calorimeters, and some low momentum particles are also bent in or out of the reconstructed jet by the magnetic field inside the tracker. For these reasons reconstructed jet energy is on average 30% lower than the reference jet energy and includes many non-linearities. Jets in ATLAS must be calibrated by simulating the fundamental processes that create them. The calibration aims to make the reconstructed jet energy as close as possible to the reference jet energy.

There are two types of calibration used in ATLAS, global and local. The global method looks

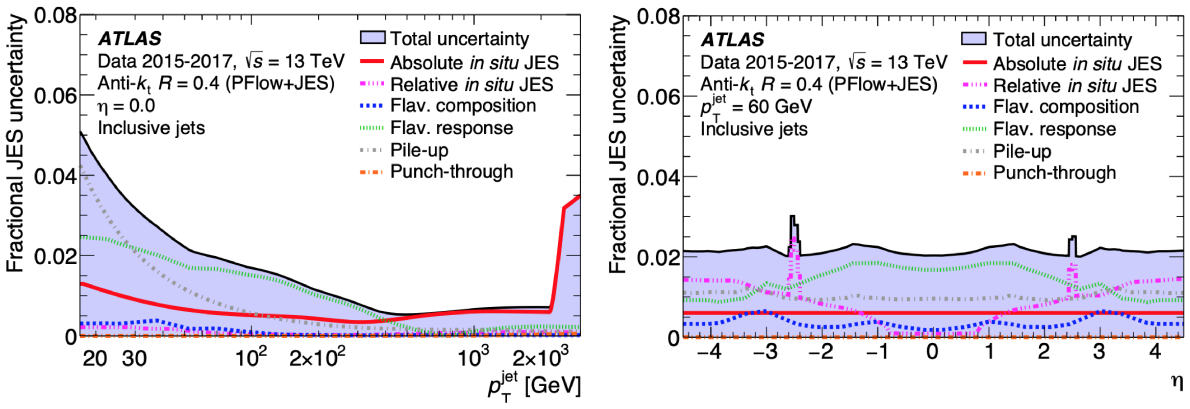
at energy density of the calorimeter cells that are included in each jet. The shower produced by a jet in the calorimeter contains a hadronic and electromagnetic component. The electromagnetic component contains electrons and photons so electromagnetic calorimeter deposits in calorimeters are compact and thus have a high energy density. The hadronic component has a broader energy deposit so its energy density is lower. The hadronic component's reconstructed energy is lower than the energy of the incident hadrons. Global calibration assigns each calorimeter cell a correction weight depending on its energy density that minimise the difference between the total reconstructed energy and the reference jet energy. This calibration method is able to recover the linearity of the reconstructed jet energy at around 2% level and improve resolution. For local calibration the topological clusters are calibrated before they are reconstructed into jets. Clusters can be classified as electromagnetic, hadronic, or unknown, according to their shape. Hadronic clusters are weighted as in the global calibration method. These calibrated clusters are then built into jets [38].

The dominant experimental uncertainty for many events at the LHC is the jet energy scale (JES). When jets are reconstructed they begin at either the EM energy scale for electrons or local cluster weighting (LCW) energy scale for hadrons. The calibration process brings jets from their measured energy scale to the JES. This corrects each cluster for energy that cannot be measured in the calorimeter, such as nuclear reactions, and for energy losses in inactive areas of the detector and due to noise thresholds. The calibration process scales measured jet energies to their proper values, and the uncertainty on this scaling is a systematic uncertainty [39]. The uncertainty accounts for differences between MC generators, the effects of noise thresholds on the jet energy, the effects of modelling, and any additional material in the detector [40]. The uncertainty is determined by comparing data to simulated jets and certain well-measured benchmark objects such as photons. The value of the uncertainty depends on jet p_T and varies between regions of the detector. The fractional JES uncertainty for jets reconstructed with the anti-k_t algorithm with $R = 0.4$ as a function of p_T and η is shown in Figure 4.7. The uncertainty is largest in the low and high p_T regions. Several factors are associated with this increase in uncertainty including noise from multijet background events in data. There are asymmetric peaks in the uncertainty around $\eta = \pm 2.5$ due to uncertainty between data and simulation. This η region represents the upper limit accessible by the inner detector. The transition between different calorimeter technologies causes a large difference between data and simulation.

The precision of measurements of the energy of jets is also an important value, this is known as the jet energy resolution (JER). This is the uncertainty in the final measurement of the jet energy after all calibration is completed. To measure detector resolution we can look at the width of the distribution of balance between jets and well-measured photons. The JER can be parameterised by the following equation:

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C, \tag{4.3}$$

where N represents noise, S represents the stochastic effect from the sampling nature of the calorimeters, and C is a p_T independent constant term. Noise from the calorimeter electronics

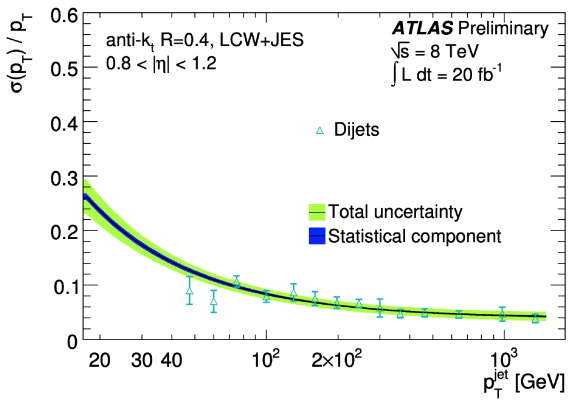


(a) Uncertainty vs p_T .

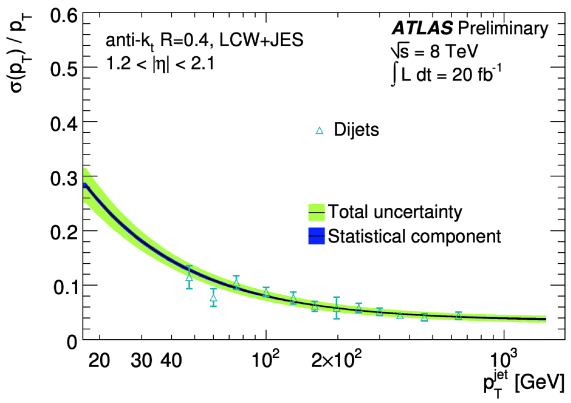
(b) Uncertainty vs η.

Figure 4.7: The total jet energy scale uncertainty as a function of p_T and η for events with two jets. Relative in situ JES refers to the uncertainty from the di-jet η-intercalibration, punch-through refers to the uncertainty in the muon-based stage of the global sequential correction, and Absolute in situ JES refers to other measurements such as single-particle uncertainty [41].

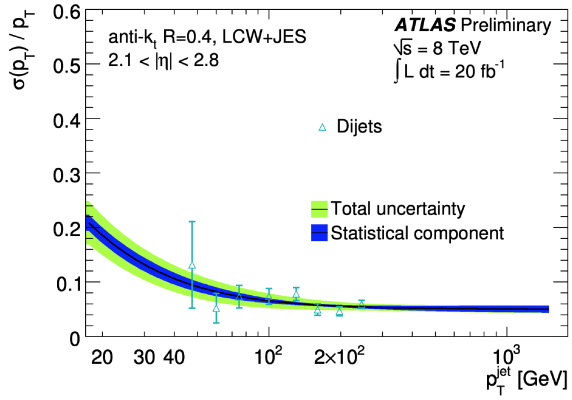
and pileup forms a significant component of the JER at low p_T [41]. The JER uncertainty for jets reconstructed with the anti-k_t algorithm with for $R = 0.4$ as a function of p_T in three regions of the detector is shown in Figure 4.8, it is greatest at low p_T and peaks around 3%.



(a) $0.8 < |\eta| < 1.2$.



(b) $1.2 < |\eta| < 2.1$.



(c) $2.1 < |\eta| < 2.8$.

Figure 4.8: The jet p_T resolution as a function of p_T in three forward regions of the detector for events with two jets. The final fit using the function in Equation 4.3 is included with its associated statistical and total uncertainty [39].

5 Analysis

5.1 STRINGS Event Generator

STRINGS-1.00 is a MC event generator for the production and decay of string resonances in pp collisions [42]. The STRINGS-1.00 script was written before the start of this research but during the analysis several problems were identified within the code so it has been modified. Changes made in the updated version STRINGS-2.00 will be described below. The generator is capable of producing first and second string resonances, along with QCD tree-level events for $2 \to 2$ partonic scattering based on the amplitudes described in Section 3. Interference terms between resonances are not considered. Events are randomly generated according to their differential cross sections with respect to M, which are calculated using the equations in Section 3.3. Differential cross sections of first string resonances generated by STRINGS-2.00 as a function of M with string scale $M_s = 7.0$ TeV are shown in Figure 5.1. The expected resonant

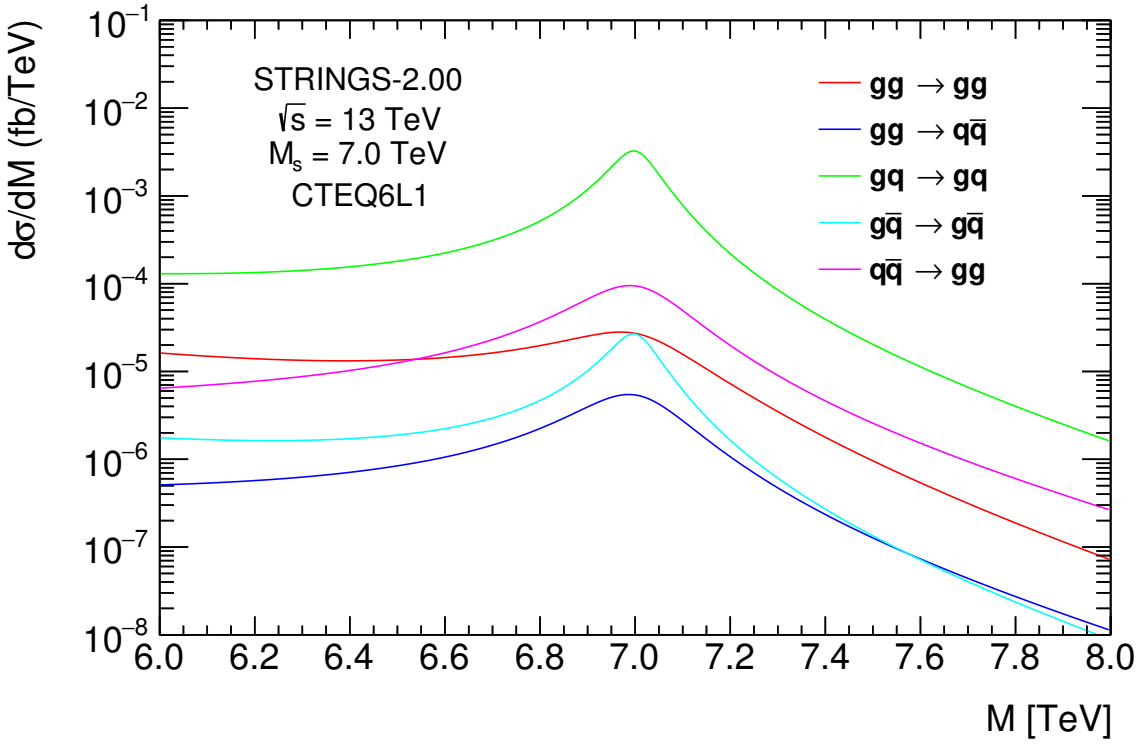


Figure 5.1: Differential cross sections as a function of M of first string resonances generated by STRINGS-2.00, with string scale $M_s = 7.0$ TeV using PDF set CTEQ6L1 [14].

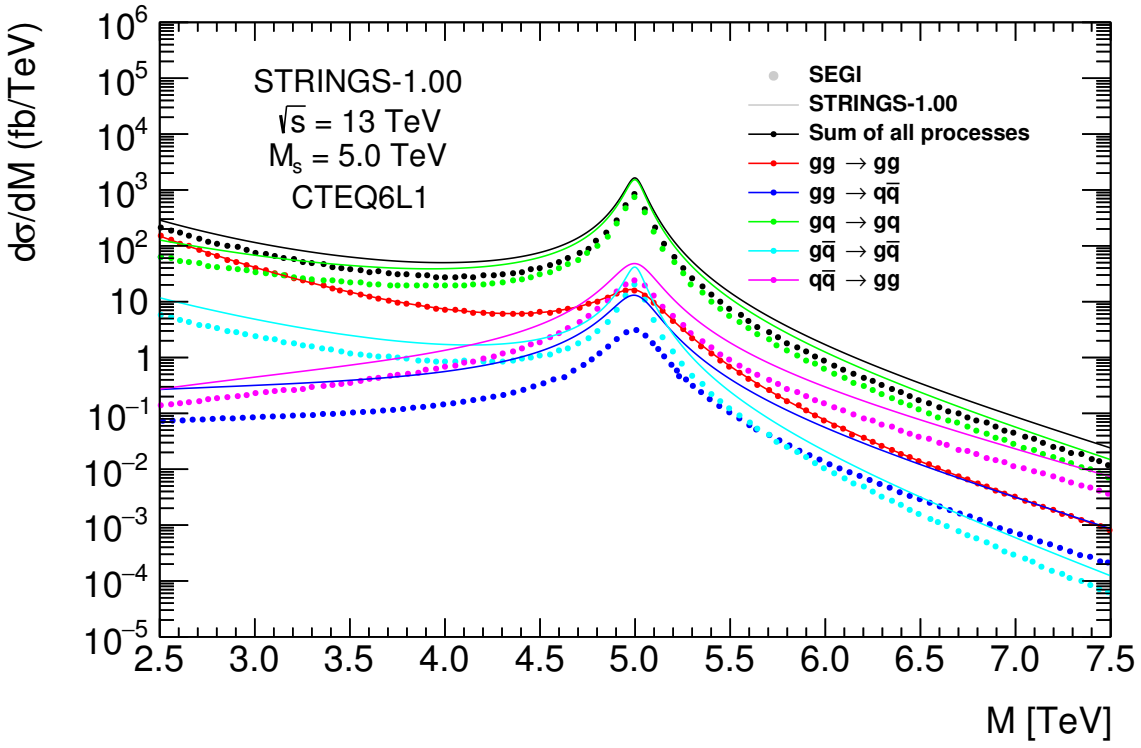
behaviour when the energy of the collision becomes close to the string scale $M \sim M_s$ is clearly visible in each of the quark/gluon processes we consider.

The generator first randomly generates M for each event. The scattering amplitudes for all processes are integrated using Equation 3.33 to give differential cross sections with respect to M. The differential cross sections are updated in STRINGS-2.00. We compare the differential cross sections generated by STRINGS-1.00 and STRINGS-2.00 to those generated by SEGI from Ref. [20] in Figure 5.2. The shapes of the differential cross sections generated by STRINGS-

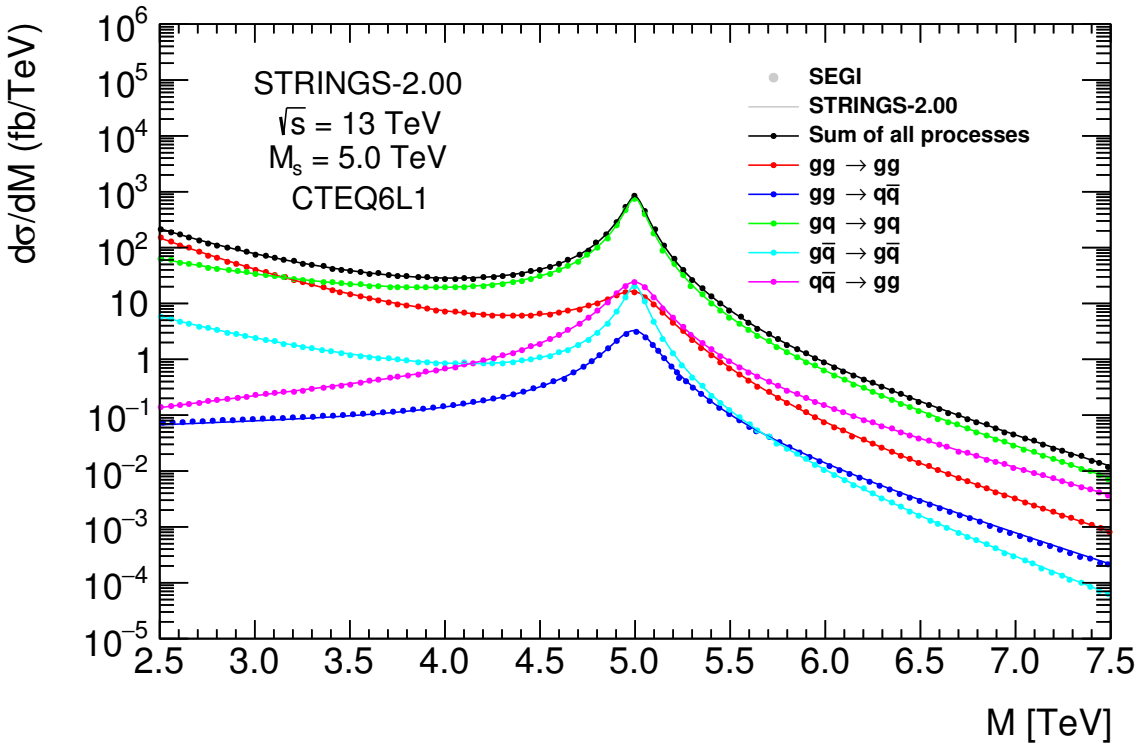
1.00 are correct for all processes but normalisations are incorrect for four out of five. The scattering amplitudes in STRINGS-1.00 multiply the formulae given in Section 3 by erroneous constants. These factors have been removed and scattering amplitudes in STRINGS-2.00 now equal Equations 3.49-3.52. This is confirmed by comparing to results generated by SEGI in Ref. [20]. Once all differential cross sections have been calculated a uniform random number between 0 and the largest differential cross section is used to determine which process each event will be. After a process type has been selected another uniform random number is generated between the minimum and maximum possible differential cross sections of that process. If the second random number is larger than the differential cross section evaluated at the given M then the random process is restarted from the beginning, otherwise the event is kept and the algorithm continues.

Y and y are also randomly assigned a value within their bounds. A third uniform random number is generated between the maximum and minimum scattering amplitudes for the chosen process. The miminum value is zero while the maximum value occurs at $Y = y = 0$. If the random number is larger than the calculated scattering amplitude then the random process is restarted from the beginning, otherwise the event is kept and the algorithm continues. With the process type determined, the code then calculates the kinematics of the event. All interactions in STRINGS-1.00 assume massless incoming and outgoing particles as described in Section 3.1. STRINGS-2.00 is updated to allow massive outgoing quarks as described in Appendix A. Colour, quark flavour, and electric charge are conserved in all collisions so the flavours of outgoing partons are determined when the process type is selected. The exception is the $gg \to q\bar{q}$ process, wherein the outgoing partons can have any of the six flavours with equal probability [42]. The kinematic variables of partons in each event are written out to a Les Houches Event file that can be input into other programs to simulate jet production and hadronisation in the detector [43].

Once all events have been generated the total cross section is calculated, as well as the cross section of each process. The algorithm used in STRINGS-1.00 first sums all differential cross sections for events of each process separately. This is then multiplied by the maximum M minus the minimum M and divided by the total number of events generated to give a total cross section for each process. These are summed to give the total cumulative cross section for all events generated. This leads to the fraction of each process's total cross section being different from its fraction of total events. In order to remedy this problem a new algorithm is implemented to calculate cross sections in STRINGS-2.00. The differential cross sections for all events are summed and multiplied by the maximum M minus the minimum M and divided by the total number of events generated to give a total cumulative cross section. To obtain the cross section of each process we multiply the total cross section by the number of events of each process and divide by the total number of events generated. The fraction of total events and fractional cross section for a sample of 10000 events generated by STRINGS-1.00 and are shown for the $gg \to gg$ process in Figure 5.3. With STRINGS-1.00 the process makes up a much larger fraction of all events than of the cumulative cross section. These numbers should be equal since the cross section is a representation of the probability of an event of each type occurring. STRINGS-1.00 uses the differential cross section of each event to calculate the total cross section of each process. The dominant process produces differential cross sections multiple

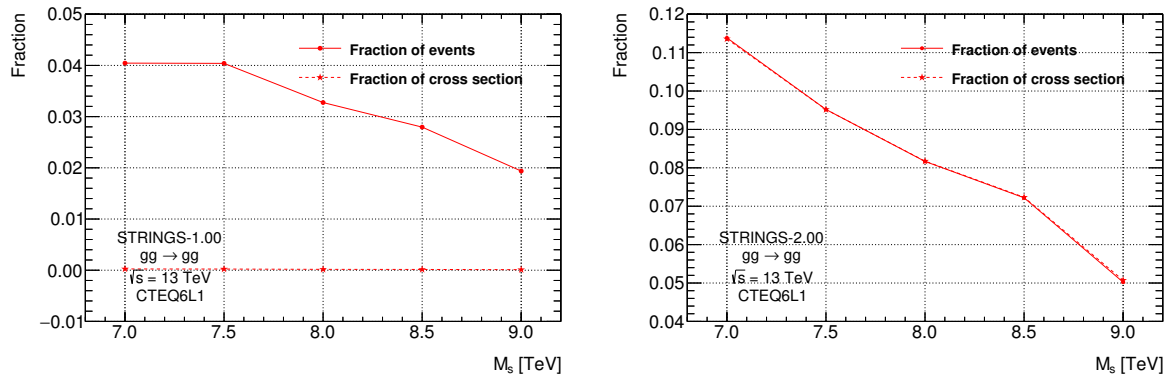


(a) STRINGS-1.00.



(b) STRINGS-2.00.

Figure 5.2: Differential cross sections as a function of M generated by SEGI (dots) obtained from Ref. [20] and STRINGS (lines) for all processes. $M_s = 5.0$ TeV and PDF set CETQ6L1 are used to generate both sets of distributions.



(a) STRINGS-1.00.

(b) STRINGS-2.00.

Figure 5.3: Fraction of total number of events (solid line) and total cross section (dashed line) for the $gg \to gg$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.

orders of magnitude larger than the other processes for certain invariant masses. When these differential cross sections are summed the most common process then makes up almost 99% of the total cross section while only contributing around 90% of the events. The other processes make up a non-negligible fraction of the events but contribute nothing to the total cross section. With this algorithm the total cross section is dominated by only a few events with very high differential cross sections, instead of accurately describing the probability distribution of all processes.

With STRINGS-2.00 event and cross section fractions are the same for all processes using the updated algorithm. The values for each fraction are not exactly equal due to rounding error. The number of events is known to infinite precision but the cross sections must be truncated to a certain number of decimal places. When the cross section fraction of each process is calculated there is some precision lost. This error is reduced for more events as the error on cross section scales with the inverse square root of the number of events generated. The fraction of events and cross section of other processes are shown in Appendix B.

5.2 MC Simulated and Data Samples

There are three types of samples used in this analysis: data, background, and string signals. Data is collected from collisions at the LHC, background is estimated from a model fit to the data, and signal is simulated. QCD processes are also simulated to validate the SM background estimate.

The data used in this analysis is collected by the ATLAS detector during the years 2015-2018. They are pp collisions at the LHC with a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb^{-1}. The uncertainty in integrated luminosity of the full data is 1.7% [44]. Data campaigns are separated by the year in which they were recorded so we have four datasets named data15, data16, data17, and data18, where the last two digits correspond to the year. These samples have an integrated luminosity of 3.2 fb^{-1} for data15, 33.0 fb^{-1} for data16, 44.3 fb^{-1} for data17, and 58.5 fb^{-1} for data18. The full data contains all events that occur within ATLAS that pass through the TDAQ system described in Section 4.3.

We use only data that is on the Good Run List (GRL), which means that at the time it was measured all detector systems were functioning and able to record high quality data. This contains events that we are interested in and many more, so we will apply selection criteria in the following section to isolate events of interest for our analysis.

MC generated events are used to model the SM background distribution. The MC16 simulation campaign consists of three samples that are meant to mimic the conditions of the data. The data15 and data16 samples contain similar beam conditions so they are combined for the purposes of MC generation. The MC samples used here are named 16a, 16d, and 16e, and are produced for data15 and data16, data17, and data18, respectively. These are simulated QCD events inside the detector and each sample contains different pileup simulations to mimic beam conditions in its associated data taking period. QCD events are simulated using Pythia v8.186 [45] with A14 tune parameters [46]. QCD processes from the SM are simulated at LO and NLO in perturbative QCD. Hard scattering events have two incoming and two outgoing partons. Pythia simulates the primary collisions between partons, showering, fragmentation, hadronising, and decaying to final state particles [47]. The entire process that occurs inside the collider is simulated with Pythia. The initial protons are simulated, along with the collisions between partons within them. The branching process described in Section 3.5 of partons creating jets is also modelled, followed by hadronisation where these partons combine into hadrons. The calorimeter deposits that these hadrons form in the detector are also simulated so that the MC samples can be analysed in the same way as data.

QCD MC samples cover a large range in cross section so the production must be split in order to obtain comparable statistical precision across the kinematic range of interest. A large number of events is generated for each MC sample, spanning a large range in energy and momentum for both incoming and outgoing partons. Each QCD MC sample is sliced into 13 regions according to leading jet p_T. These are named $JZ\{0-12\}W$, where W indicates that the slices are weighted [48]. There are two variables used to separate the regions, $\hat{p}_T$ and p_T. $\hat{p}_T$ is the transverse momentum of outgoing partons in the centre-of-mass frame and p_T is of the leading jet in the lab frame. $\hat{p}_T$ can only be determined at the generator level and must be measured within the simulation. p_T is an observable quantity after jet reconstruction, measured using the anti-k_t jet algorithm with $R = 0.6$ for final state particle four-vectors. The p_T requirements and weights of each slice are shown in Table 5.1. Samples at high p_T have a low cross section so many random events must be generated in order to produce enough events in the high p_T region. This requires a lot of time and computational power. The looser $\hat{p}_T$ requirement allows for increased efficiency in producing these events since it can be applied at the generator level. Events that do not fulfil this requirement need not undergo the entire process of showering, fragmentation, hadronising, and decay to final state particles. Events that satisfy the $\hat{p}_T$ requirement are allowed in the simulation and those events that also satisfy the p_T requirement make up each slice. A weighting factor is associated with each slice to simulate the correct shape of the p_T distribution after all slices are combined. The slices are weighted by their cross section times filter efficiency times luminosity, divided by the sum of weights for each slice. The luminosity is different for each sample, MC 16a uses the summed luminosities of data15 and data16, MC16d uses the luminosity of data17, and MC16e uses the

Slice	Minimum $\hat{p}_T$ (GeV)	Minimum p_T (GeV)	Maximum p_T (GeV)	Weight
JZ0	0	0	20	6.1954E+07
JZ1	0	20	60	6.1925E+07
JZ2	15	60	160	1.8089E+04
JZ3	50	160	400	4.5050E+02
JZ4	150	400	800	1.3797E+01
JZ5	350	800	1300	5.3340E-01
JZ6	600	1300	1800	7.7851E-02
JZ7	950	1800	2500	1.2240E-02
JZ8	1500	2500	3200	3.5431E-03
JZ9	2200	3200	3900	6.6163E-04
JZ10	2800	3900	4600	7.8627E-05
JZ11	3500	4600	5300	4.1545E-05
JZ12	4200	5300	∞	2.4181E-05

Table 5.1: p_T requirements and weights of JZW slices for Pythia QCD dijet samples. $\hat{p}_T$ is of outgoing partons measured in the centre-of-mass frame and p_T is measured in the lab frame for jets. Weight is defined as cross section $\times$ filter efficiency $\times$ luminosity.

luminosity of data18. The filter efficiency is the ratio of the number of events in each slice that satisfies both $\hat{p}_T$ and p_T requirements to the number of events in the slice that passes only the $\hat{p}_T$ requirement. It takes a lot of computational power and time to generate events with high p_T so this partitioning of event generation allows for more efficient use of resources. The procedure used to produce these MC samples is described in detail in Ref. [48].

STRINGS-1.00 [42] described in Section 5.1 is used to generate the string resonance signal samples. This script simulates Regge excitations in collisions between quarks and gluons. The PDF set CTEQ6L1 [14] is used in the generation of these events. The samples were generated using STRINGS-1.00 before updates were made to the generator. It was found that numerical factors can be applied to the existing samples to account for the corrections made in STRINGS-2.00. This eliminates a need for the generation of new samples, saving computational resources. Samples are generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV. Several searches have been performed for string resonances with other experiments such as CMS [49, 50] and the maximum lower limit on the string scale to date is $M_s > 7.7$ TeV [51]. Only one of the five processes that we consider here was used in that search so we will study values of M_s in this range and above. The higher values of M_s that we consider here are expected to be at the upper limit of energies accessible in the data.

The string resonance differential cross sections can be significantly enhanced by the PDFs at low x, corresponding to low M. We are interested in a narrow resonance structure in the M spectrum so the low mass region is not included in sample generation. The maximum value of M included in sample generation is 13 TeV for all string samples since this is the largest possible energy of collisions in the detector. There are two possibilities for the lower bound on mass generation $M_{Low\,cut}$ of each string sample. The first option is the M value corresponding to the minimum sum of differential cross sections for all processes to the left of the resonance peak. The second option is the M value such that the interval $[M_{Low\,cut}, 13]$ TeV covers 95% of the area

under the Breit-Wigner curve $[M_s - 1.0, 13]$ TeV. Whichever interval gives the larger range in M is selected for each string signal. The $M_s = 7.0$, 7.5, and 8.0 samples are generated with masses using the first option described and the $M_s = 8.5$ and 9.0 TeV samples are generated with masses according to the second option. Figure 5.4 shows the differential cross sections generated by STRINGS-2.00 over the mass range $M = [0.1, 13]$ TeV for $M_s = 7.0$ TeV with $M_{Low\,cut}$ indicated. The entire range of possible energies is shown but only energies in the range $[M_{Low\,cut}, 13]$ TeV are allowed when generating the string samples. The differential cross section distributions for the remaining samples can be found in Appendix C. String samples are generated in the MC16

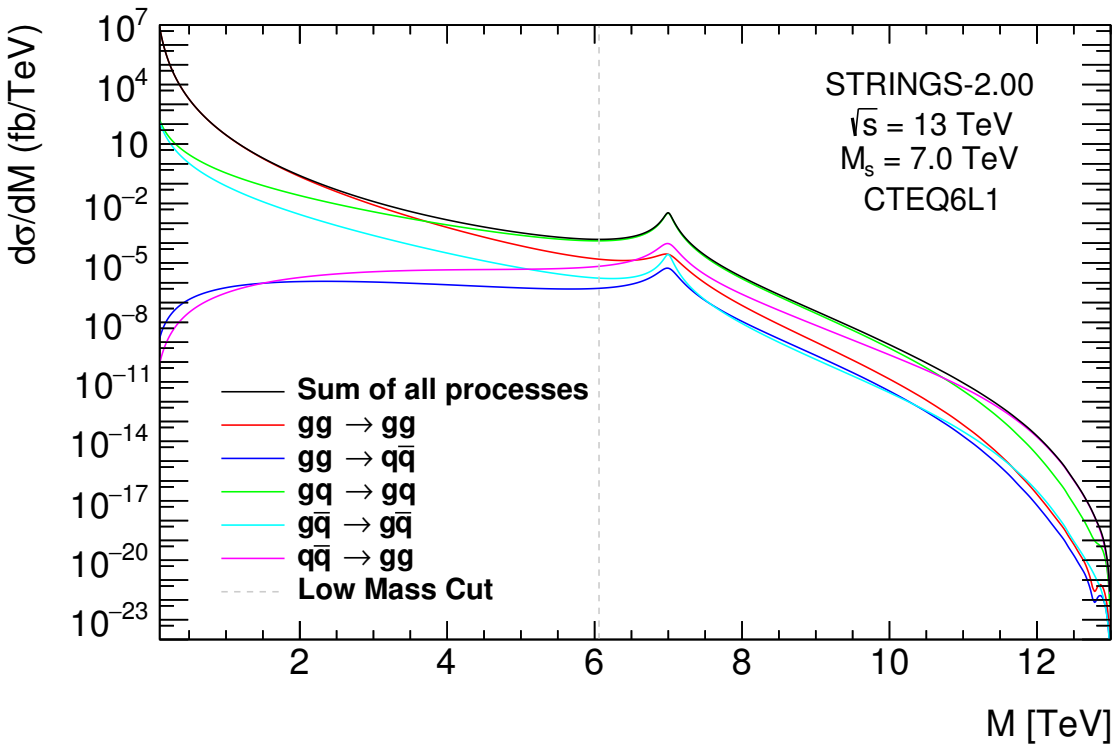


Figure 5.4: Differential cross sections as a function of M for all five processes generated by STRINGS-2.00 with $M_s = 7.0$ TeV using PDF set CETQ6L1 [14]. The black line is $M_{Low\,cut} = 6.06$ TeV.

campaign. 20000 events are generated within subcampaigns a, d, and e, for 60000 total. The string scales used here along with $M_{Low\,cut}$ and their corresponding total cross sections for all processes are shown in Table 5.2. The relative contribution of each process to the total cross

M_s (TeV)	$M_{Low\,cut}$ (TeV)	σ (fb)
7.0	6.06	7.09
7.5	6.60	1.86
8.0	7.14	0.456
8.5	7.60	0.100
9.0	8.05	0.0199

Table 5.2: Total cross sections for 60000 events generated using STRINGS-1.00 [42].

section varies with M_s , as shown in Table 5.3. For all values of M_s considered here the dominant

process is $gq \to gq$, which contributes between 81-88% of all events. The fraction of events of this type increases with M_s. The samples were generated using STRINGS-1.00, which does not

Process	M_s (TeV)				
	7.0	7.5	8.0	8.5	9.0
$gg \to gg$	11.36%	9.52%	8.17%	7.23%	5.07%
$gg \to q\bar{q}$	0.29%	0.28%	0.22%	0.19%	0.17%
$gq \to gq$	81.36%	83.12%	84.58%	85.54%	87.59%
$g\bar{q} \to g\bar{q}$	0.68%	0.58%	0.52%	0.51%	0.49%
$q\bar{q} \to gg$	6.31%	6.50%	6.51%	6.53%	6.67%

Table 5.3: Relative contribution of string resonance processes to the total cross section for different string scales M_s.

calculate the correct differential cross sections as outlined in Section 5.1. To account for the fixes implemented in STRINGS-2.00 we must apply correction factors. We apply the weights in Table 5.4 to the events to recover the correct signal strengths in our generated samples. These weights are determined by dividing the cross section percentages given by STRINGS-2.00 by the cross section percentages given by STRINGS-1.00 for each process and string scale. The MC samples simulate events inside the detector so we cannot be completely certain of which process occurs within each event. The incoming partons can be differentiated but not the outgoing ones so the $gg \to gg$ and $gg \to q\bar{q}$ processes must be combined in these event weights. All events are

Process	M_s (TeV)				
	7.0	7.5	8.0	8.5	9.0
$gg \to q\bar{q} + gg$	2.836	2.161	2.457	2.422	1.569
$gq \to gq$	0.901	0.924	0.930	0.939	0.951
$g\bar{q} \to g\bar{q}$	4.857	5.780	3.587	4.591	6.213
$q\bar{q} \to gg$	1.158	1.164	1.175	1.136	1.153

Table 5.4: MC string resonance process event weights. These must be applied to samples generated with STRINGS-1.00 to account for incorrectly calculated differential cross sections.

weighted with the following formula

$$w = \frac{\sigma L_i c_{ij}}{\sum_j c_{ij}}, \tag{5.1}$$

where i runs over the MC subcampaigns and j runs over events in i. σ is the cross section for events at a given string scale M_s from Table 5.2, L is the luminosity for the data taking period associated with the MC subcampaign, and c is the event correction from Table 5.4, required to get the correct q/g ratios. These weights ensure that the signal samples and processes are the proper strength independent of the number of events generated. The signal events are weighted to the luminosity of the data so the event rate depends only on the cross section of string resonances. We ensure that the number of signal events corresponds to the theoretical prediction for the number of data events occurring within the detector. In this way the number of data events does not need to be known before the creation of signal samples and the same

number of signal events can be generated for each string scale, even though their cross sections are different.

STRINGS-1.00 is interfaced to Pythia v8.240 [52] for modelling of the parton shower, hadronisation, and underlying event, with A14 tune parameters [46]. As with the QCD MC samples Pythia is used to simulate the processes that occur within the detector after the primary collision. The EvtGen 1.6.0 program [53] is used to decay bottom and charm hadrons. Pileup is modelled by overlaying simulated inelastic pp events generated with Pythia v8.186 [45] using the NNPDF2.3LO set of PDFs [54] and the A3 tune [55] over the original hard scattering event.

The output for all three types of samples is in the form of calorimeter deposits that are grouped into topological clusters, whether they be real or simulated. This information is then put through the jet reconstruction procedure described in Section 4.4. The specific parameters used during jet reconstruction are listed in Appendix D.

5.3 Event Selection

5.3.1 Pre-Selection

For consistency data and MC samples are analysed in the same manner. We apply pre-selection criteria after reconstruction to ensure that we use only high quality information. We then make further event selections to keep only events that are relevant for our resonance search. The invariant mass m_{jj} of dijet events is calculated from the two jets with the highest p_T in each event. The highest p_T jet is called the leading jet, while the second highest p_T is called the subleading jet. The calibrated four-momenta of the two jets are summed to give

$$m_{jj} = \sqrt{(E_3 + E_4)^2 - (\vec{p}_3 + \vec{p}_4)^2}, \tag{5.2}$$

where E_3 (E_4) and $\vec{p}_3$ ($\vec{p}_4$) are the energy and three-momentum of the leading (subleading) jet, respectively [56]. m_{jj} is approximately equal to the invariant mass M of the outgoing partons in the initial collision but due to uncertainties in jet reconstruction there may be some discrepancy between M and m_{jj} . In the SM the dominant source of dijet events is $2 \to 2$ partonic scattering from QCD processes. The m_{jj} spectrum of these events is smooth and continuously falling so is a good place to search for resonant behaviour since a resonance should be clearly visible an an excess above background.

When data events are reconstructed calorimeter cells that register energy deposits significantly higher than noise levels are grouped together into jets as described in Section 4.4. We use the anti-k_t algorithm with $R = 0.4$. As mentioned in Section 4.4 reconstructed jets from calorimeter deposits must be corrected since particles can lose energy as they move through the detector. The data, QCD MC, and signal samples are treated slightly differently since they are created in different ways. The first selection made for data samples is to reject any events that are not on the GRL, meaning one or more of the detector instruments was not functioning at full capacity when the event was recorded. Next, events registering errors in detector instrumentation are rejected. This includes the liquid-argon calorimeters, tile calorimeters, and semiconductor detectors. If there are any errors detected with the event reconstruction then the data is also rejected. This is called the core selection and identifies events that indicate an

incomplete event build from lack of calorimeter information or possible errors in calibration. These filters are only applied to data since MC samples model the detector in good working order.

The next step is to apply cleaning cuts to all reconstructed jets following the loose criteria. The criteria required for this selection is outlined in Ref. [58] and is designed to isolate "good" jets that are created by high energy objects in a collision, and remove "fake" jets that arise from background processes. There is also a requirement that the vertex with the highest sum of p_T^2 for all associated jets must have at least two associated tracks. This ensures that the primary vertex in all selected events is associated with at least two outgoing particles since we are interested in studying dijet events. These cuts are applied to data and MC events and are used as a baseline for events that can safely be used in most analyses. The QCD MC samples also undergo MC cleaning to remove any events that produce errors in jet reconstruction. After reconstruction and pre-selection events are stored in the form of ntuples, which hold kinematic information, weights, and other details of each event. The Data, QCD MC, and string ntuples used in this analysis are listed in Appendix E.

5.3.2 Object Selection

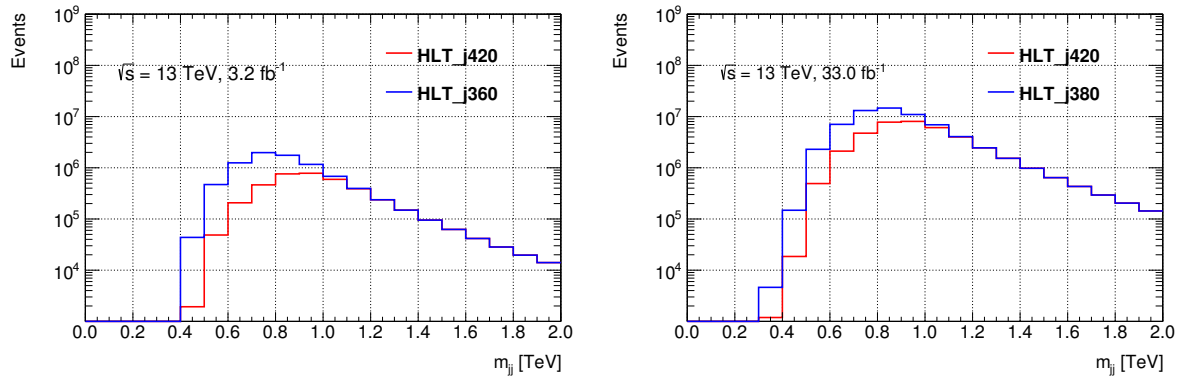
After all samples have been cleaned and stored in ntuples we can specify further requirements. Selections are typically placed on the objects being studied, which in this case are jets. We specify a requirement called jet selection that aims to remove low energy jets attributed to background processes. Previous ATLAS studies have shown that pileup effects can be ignored for jets with $p_T > 50$ GeV [59]. We require that the leading jet p_T must be greater than 380 GeV and the subleading jet p_T must be greater than 150 GeV. These selections remove a large number of low energy events to reduce data storage requirements for our samples. This selection has no effect on the signal samples but many background events are rejected.

The lowest unscaled single-jet trigger in 2015 was HLT_j360, in 2016 was HLT_j380, and in 2017 and 2018 was HLT_j420. We require that events pass at least one of the triggers in the list {HLT_j360, HLT_j380, HLT_j400, HLT_j420, HLT_j440}. We use the loosest requirement on trigger since this selection typically removes a large number of events. Using a higher trigger requirement could potentially lead to a loss of signal events. In order to maximise the amount of signal after selection we use HLT_j420 since it is the lowest unscaled single-jet trigger used in all data taking periods. This trigger requires at least one jet with p_T above 420 GeV in each event.

5.3.3 Event Selection

Further selections are made to reduce the number of events, specifically to remove events that are not useful in our analysis. Since we are looking for high energy dijet events we will require at least two reconstructed jets in every event. The signal processes that we consider contain two outgoing particles so are very unlikely to correspond to events that create a single jet in the detector. We are interested in $2 \rightarrow 2$ scattering events which should result in two jets but during the process of parton showering and hadronisation it is possible that additional jets could be created in the detector. For this reason we include events that have two or more associated jets.

We also set a lower limit on the value of m_{jj} since the p_T selections remove many low m_{jj} events and we wish to have a smooth spectrum. The lower limit that we set on m_{jj} depends on the choice of trigger. A mass selection will be applied to remove events where the trigger efficiency is less than 99.5%. To measure the trigger m_{jj} turn-on in data an unbiased sample was obtained using the HLT_j360 trigger for data15 and HLT_j380 trigger for data16. These are the lowest unscaled single-jet triggers used in their respective years. Data17 and data18 use HLT_j420 as the lowest unscaled single-jet trigger so we are not able to compare trigger efficiencies within these samples as there is no reference trigger available. Figure 5.5 shows the number of events for both triggers as a function of m_{jj}. The efficiency is the ratio of number of events passing the lower trigger to number of events passing HLT_j420. The mass at which the efficiency is 99.5% is approximately 1.2 TeV for both samples so this is chosen as the lowest value of m_{jj} allowed by our selection criteria.



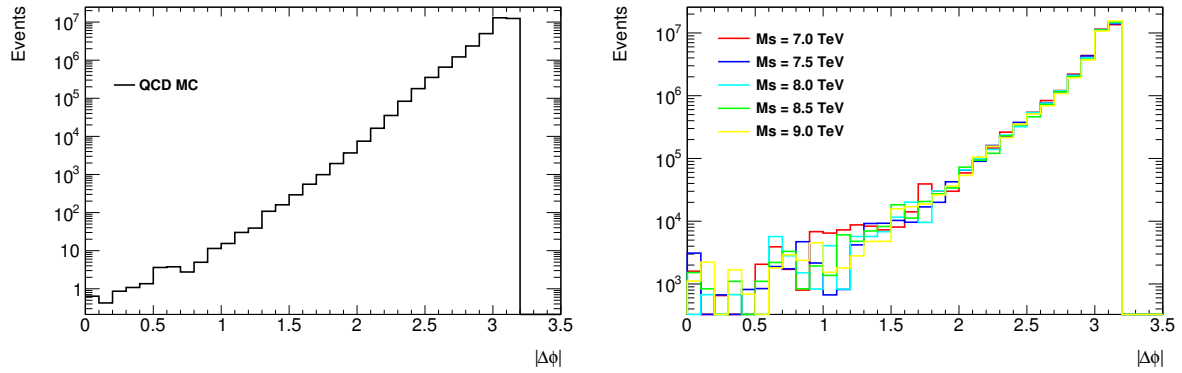
(a) Data15. (b) Data16.

Figure 5.5: Number of events in m_{jj} data spectrum from data15 and data16 with trigger HLT_j360 and HLT_j380, respectively, compared to HLT_j420.

5.3.4 Analysis Selection

After ensuring that our samples contain high quality events and that many low m_{jj} background events have been removed we specify selection criteria unique to our resonance search. These selections are specific to this analysis and aim to increase our signal-to-background ratio by removing many background events and few signal events.

We are searching for high energy s-channel processes that contain back-to-back jets in the transverse plane. We anticipate that the two leading jets in signal events will have a large difference in azimuthal angle ϕ. Background processes have a smoother distribution of azimuthal angle separation so in order to remove background events while retaining signal events we will place a bound on $|\Delta\phi| = |\phi_3 - \phi_4|$ of the leading and subleading jets in each event. The $|\Delta\phi|$ distributions of QCD MC and string signal samples before imposing a $|\Delta\phi|$ selection are shown in Figure 5.6. Signal samples are scaled to the same number of events as QCD MC for a simple visual comparison. The $|\Delta\phi|$ distribution of background is smooth throughout, while the distributions for each of the signal samples are discontinuous at low values. Each signal sample has few events with $|\Delta\phi| \lesssim 1.0$ and an abrupt increase in events near $|\Delta\phi| \approx 1.0$. The signal distributions then all become smoothly increasing for $|\Delta\phi| \gtrsim 1.0$. We aim to remove the



(a) QCD MC.

(b) String signals.

Figure 5.6: $|\Delta\phi|$ distributions for QCD MC and string signal samples. Signal samples are scaled to the same number of events as QCD MC. The selection criteria described above are applied to these samples but no selection is placed on $|\Delta\phi|$.

largest amount of background events while retaining the largest number of signal events. The same selection criteria must be used for all signal samples so we choose a value of $|\Delta\phi|$ selection that suits all signals. There is a sharp increase in signal events in the vicinity of $|\Delta\phi| = 1.0$ for each signal sample but the exact location varies between signals. This is chosen as an adequate selection for each signal so we require that the azimuthal angle separation $|\Delta\phi|$ between the two leading jets must be larger than 1.0. This selection is in accordance with the most recent ATLAS dijet resonance search [57].

As described in Section 3.3 the QCD scattering amplitudes of s-channel processes scale with $\hat{s}^{-2}$. Similarly t and u-channel processes scale with $\hat{t}^{-2}$ and $\hat{u}^{-2}$, respectively. In the high energy interactions that we are studying the momentum of both incoming particles is quite high, leading to large values of $\hat{s}$. In interactions where the energies of incoming and outgoing particles are very similar we will have small values of $\hat{t}$ and $\hat{u}$. This means that with QCD processes t and u-channel events dominate over the s-channel. The Mandelstam variables can be expressed in terms of the scattering angle in the centre-of-mass frame θ^* as

$$\hat{t} = -\frac{1}{2}\hat{s}(1 - \cos\theta^*), \qquad \hat{u} = -\frac{1}{2}\hat{s}(1 + \cos\theta^*). \tag{5.3}$$

This is related to the rapidity separation of the two leading jets. We define

$$y^* = \frac{y_3 - y_4}{2} \tag{5.4}$$

where y_3 and y_4 are the rapidities of the leading and subleading jets, respectively, in the lab frame. y^* is related to the scattering angle by

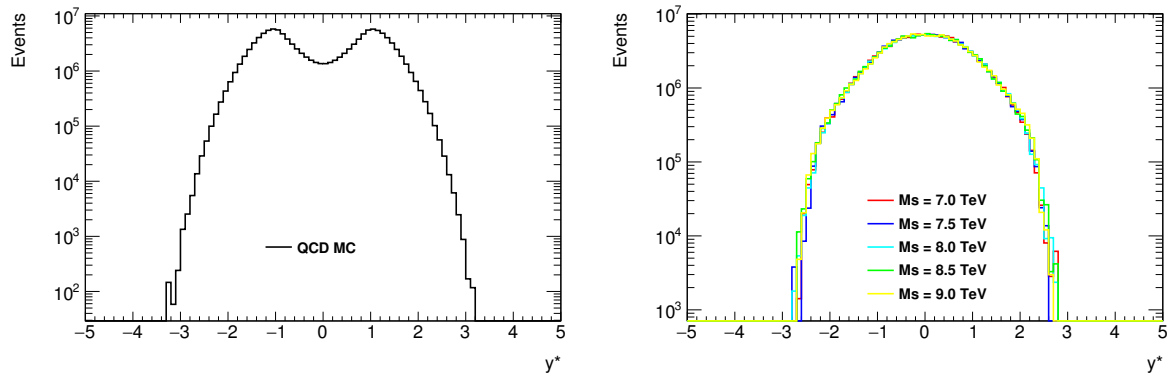
$$\cos\theta^* = \tanh y^*. \tag{5.5}$$

In QCD high energy dijet events are predominantly generated through the t-channel with an

angular distribution similar to that of Rutherford scattering

$$\frac{dN}{d\cos\theta^*} \sim \frac{1}{\hat{t}^2} \sim \frac{1}{(1-\cos\theta^*)^2}. \tag{5.6}$$

The signal events that we anticipate do not have the same angular distribution as QCD processes. String resonances are produced by the exchange of massive string excitations in the s-channel so the cross sections are dominated by this contribution rather than t and u-channel interactions. We can see that Equation 5.6 depends on $\hat{t}$, peaking when it is small, while Equations 3.49-3.52 depend on $\hat{s}$, peaking when it is close to M_s. Equation 5.6 peaks strongly near $\cos\theta^* = 1$, this means that the $|y^*|$ distribution of QCD background processes will be minimal at 0. Signal events are expected to have a much flatter $\cos\theta^*$ distribution than QCD, peaking at 0. This corresponds to a peak in y^* of 0 for signal events [60]. The y^* distributions of QCD MC and string signal samples before imposing a $|y^*|$ selection are shown in Figure 5.7. Signal samples are scaled to the same number of events as QCD MC for visual purposes.



(a) QCD MC.

(b) String signals.

Figure 5.7: y^* distributions for QCD MC and string signal samples. Signal samples are scaled to the same number of events as QCD MC. The selection criteria described above are applied to these samples but no selection is placed on y^*.

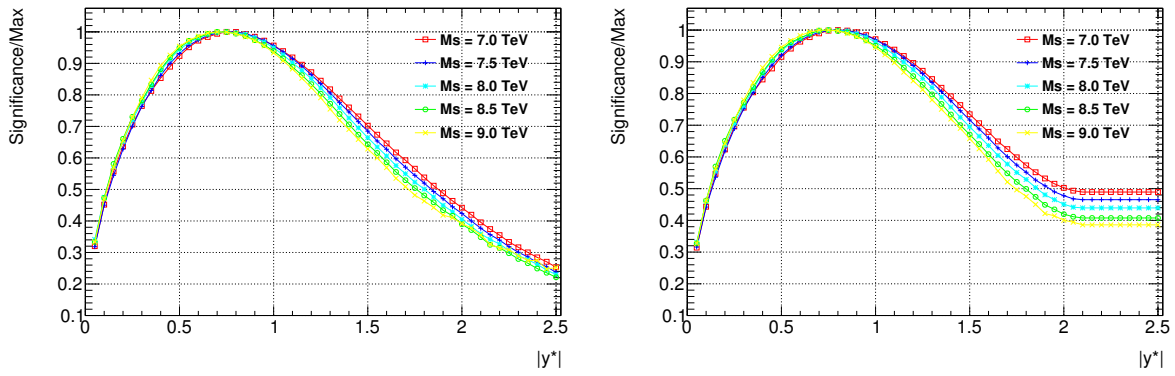
To reduce contribution from background QCD processes and optimise the signal included a requirement is placed $|y^*|$. By setting an upper bound on $|y^*|$ we reject t-channel QCD processes while admitting anticipated signal events. To optimise the $|y^*|$ selection we produce signal and background m_{jj} distributions with various $|y^*|$ thresholds. We then calculate the significance of each selection [61]

$$\text{Significance} = \sqrt{\sum_i 2\left[(S_i + B_i)\cdot \ln\left(1+\frac{S_i}{B_i}\right) - S_i\right]}, \tag{5.7}$$

where S_i (B_i) is the number of signal (background) events in bin i of the m_{jj} distribution. QCD MC events are used as the background sample for this calculation. Not all bins in the distribution are used to calculate significance, we use only bins that represent 95% of the total number of signal events. This ensures the significance calculation includes only bins that have a large ratio of signal over background, which the selection attempts to enhance. For ATLAS purposes this analysis is being performed in conjunction with a study of the same signals using

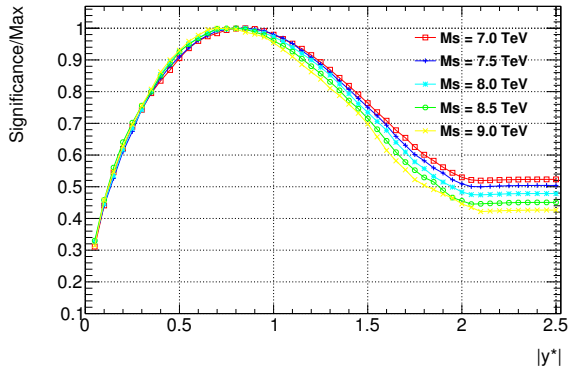
the application of quark/gluon tagging. The analysis categorises jets to determine if studying only events with a certain quark or gluon content can lead a more significant signal. Depending on the number of charged particles with p_T above a threshold of 500 MeV in a certain jet it can be determined whether the jet originated from a quark or gluon, called tagging. This process is described in detail in Ref. [62]. Here a jet is tagged as originating from a gluon if $n_{trk} > -7.541 + 3.233 \ln(p_T)$[1], where n_{trk} is the number of associated tracks that have $p_T > 500$ MeV associated with the jet. We produce two distributions in addition to the inclusive sample discussed in detail here. The one gluon tagged sample requires that at least one of the two leading jets in each event is tagged as originating from a gluon, and the two gluon tagged sample requires that both leading jets in each event are tagged as originating from a gluon.

The same selection criteria for all variables are applied to the inclusive (performed here), one gluon tagged, and two gluon tagged cases. It was found that all other selection criteria have approximately the same impact on the tagged and untagged cases but the choice of $|y^*|$ selection depends on quark/gluon tagging. To determine the optimal $|y^*|$ selection criteria we calculate significances using Equation 5.7 for the inclusive and tagged samples. The relative significances for various values of $|y^*|$ selection are shown in Figure 5.8. The $|y^*|$ selection leading to the



(a) Inclusive.

(b) ≥ 1 gluon tag.



(c) 2 gluon tag

Figure 5.8: Significance, calculated using Equation 5.7, divided by the maximum value for various $|y^*|$ selection criteria.

maximum significance for each string scale is listed in Table 5.5. The significance peaks at

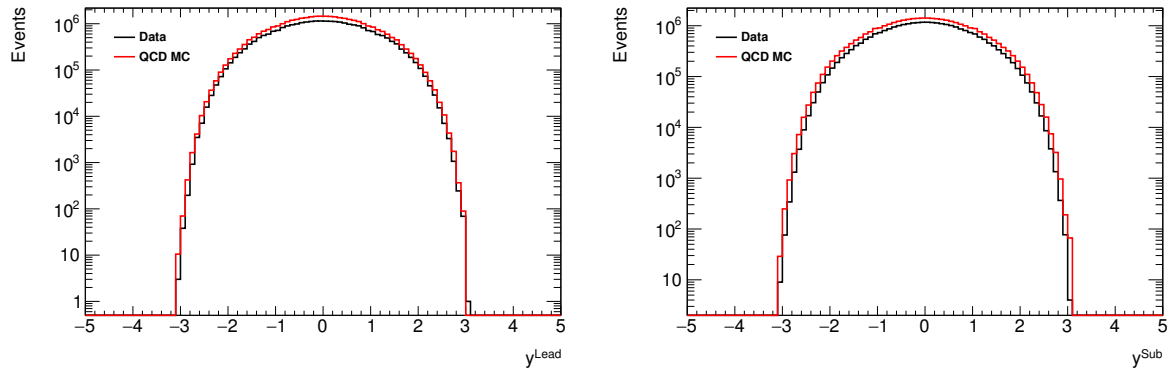
[1]These values are obtained from a study performed for the ATLAS exotics q/g tagged dijet analysis. This work is not available publicly.

M_s (TeV)	Optimal Selection		
	Inclusive	≥ 1 g tag	2 g tag
7.0	0.75	0.80	0.85
7.5	0.75	0.80	0.85
8.0	0.75	0.75	0.85
9.5	0.70	0.75	0.80
9.0	0.70	0.75	0.70

Table 5.5: $|y^*|$ selection leading to the maximum significance value. Significance is calculated using Equation 5.7.

$|y^*| < 0.75$ for the inclusive and one gluon tagged samples, and $|y^*| < 0.85$ for the two gluon tagged samples. The selection of $|y^*| < 0.8$ is chosen to approximate the optimal value for the tagged and untagged analyses.

No selection criteria is placed on the ϕ or y values of the leading or subleading jets in this analysis. The separation between these kinematic variables is bounded but individual jets are allowed to have any value. The string signal samples are generated with $|y_3|, |y_4| < 2.5$ but there are very few events with $|y_3|, |y_4| > 2.0$. Bounds on the values of ϕ and y for each of the leading jets may be required in some analyses to ensure that the jets being studied are deposited in the calorimeters. It is more difficult to accurately reconstruct jets that have large values of y but in this case after all other selections have been made no bound is required. The distributions of leading and subleading jet y for data and QCD MC are shown in Figure 5.9. After all other



(a) Leading jet.

(b) Subleading jet.

Figure 5.9: y distribution of leading and subleading jets for data and QCD MC, before nomarlization to data. All selection criteria have been applied.

selections have been made the largest $|y|$ value for both data and QCD MC is approximately 3.1. These events all fall within well-instrumented areas of the detector so further selection criteria on $|y_3|$ and $|y_4|$ would be redundant.

The event selection criteria are summarised in Table 5.6. The distributions of kinematic variables that are used in event selection are shown in Appendix F for all samples. Tables with the number of events remaining after each selection criteria for data, MC QCD, and string samples can be found in Appendix G. The selection criteria are applied to each sample and then the samples are binned into a histogram according to m_{jj} . The m_{jj} bin size is optimised

Selection
Leading jet $p_T > 380$ GeV
Subleading jet $p_T > 150$ GeV
Trigger HLT_j420
Number of jets ≥ 2
$m_{jj} > 1.2$ TeV
$\lvert\Delta\phi(jj)\rvert > 1.0$
$\lvert y^*\rvert < 0.8$

Table 5.6: Event selection criteria applied to all samples.

to minimise JER effects following the studies performed for the 2015 dijet analysis [63] and the Run 2 dijet analysis [64]. The bins have variable widths since the number of events decreases with increasing m_{jj} . Low mass regions have narrow bins while higher mass regions have wider bins. The full m_{jj} binning can be found in Appendix H.

5.4 Background Estimate

After all event selection criteria have been applied we compare the QCD MC dijet sample to data. Typically QCD multijet processes do not agree with data since perturbative QCD predictions have a large uncertainty. The result is that our simulated dijet distribution is similar in shape to data in the range we consider but its total cross section is larger. To have an accurate background estimate we normalise our QCD MC distribution to the number of data events. The raw and normalised distributions are compared with data in Figure 5.10. The raw distribution is scaled by 0.802.

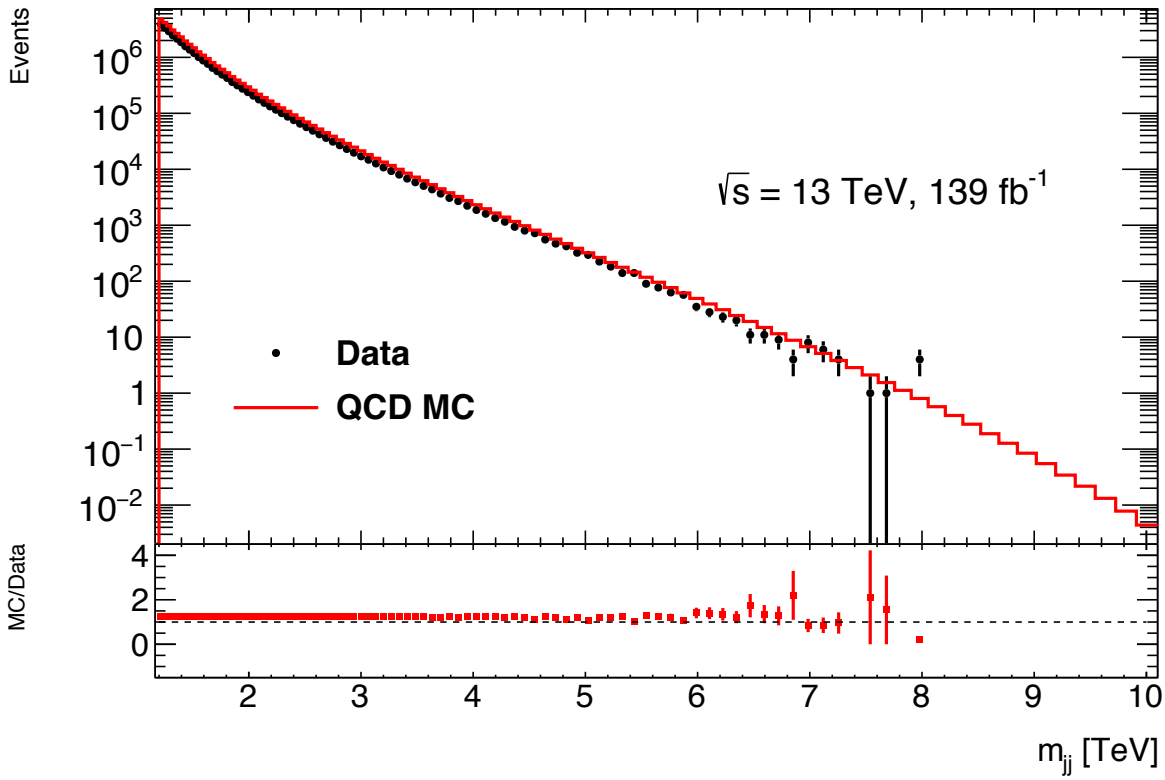
The majority of SM dijet events are produced through QCD multijet processes, which have a smooth m_{jj} spectrum. There are large theoretical uncertainties on QCD processes as discussed in Section 3.5 and a limited simulation size so the QCD MC samples are not an accurate enough background estimate. To determine the SM contribution to the data distribution we do not use the QCD MC samples but rather fit the data m_{jj} spectrum. The data is fit with a parametric function

$$f(x) = p_1(1-x)^{p_2}x^{p_3+p_4\ln x+p_5(\ln x)^2}, \tag{5.8}$$

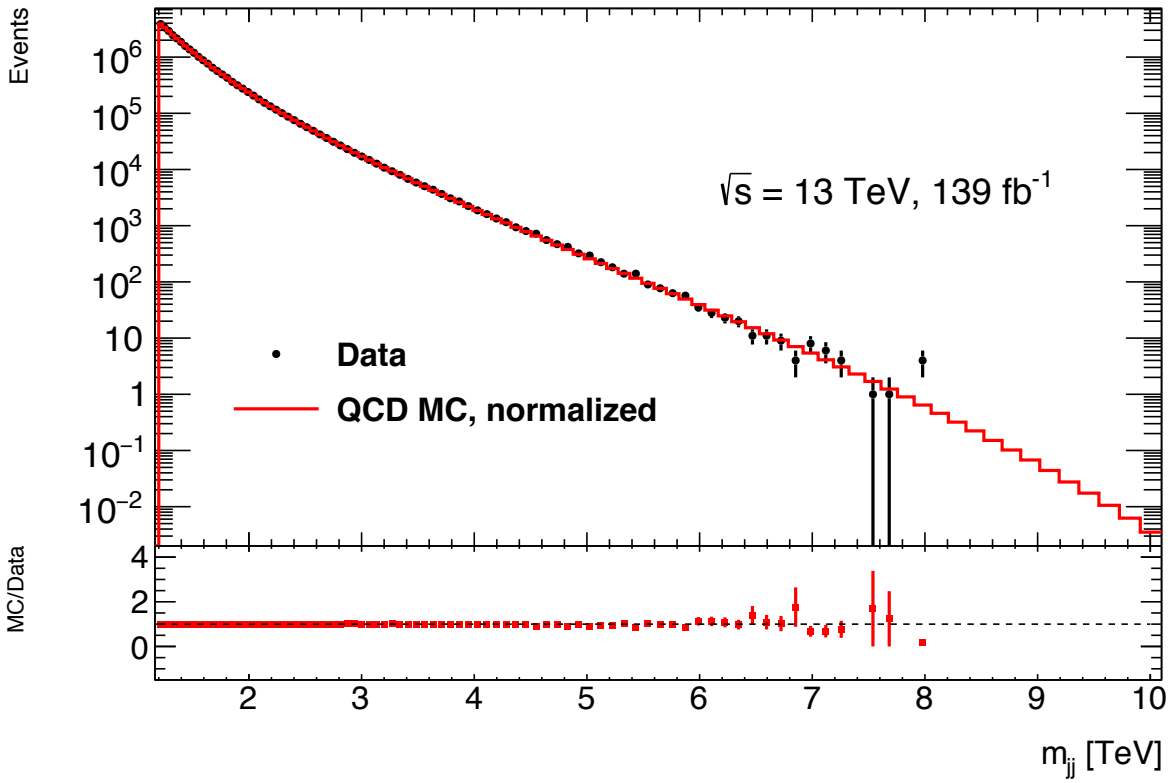
where

$$x = \frac{m_{jj}\ (\text{TeV})}{\sqrt{s}} = \frac{m_{jj}\ (\text{TeV})}{13\ \text{TeV}}, \tag{5.9}$$

and $p_{1,2,3,4,5}$ are fit parameters. This function is motivated by the (x^p) form of LO QCD matrix elements and the $((1-x)^p)$ form of the PDFs [65]. Previous ATLAS searches with 3.57 fb^{-1} [66] and 37.0 fb^{-1} [67] of data found that Equation 5.8 with three ($p_4 = p_5 = 0$) and four ($p_5 = 0$) parameters, respectively, were able to sufficiently describe the data. With increased statistics we find that we require more parameters to describe the full m_{jj} distribution. The most recent search from ATLAS found that the five parameter function accurately describes the m_{jj} distribution predicted by LO and NLO QCD MC [57]. With increasing data that spans more than 6 orders of magnitude it becomes challenging to fit the entire m_{jj} spectrum with a single global fit function so we fit the spectrum using the sliding window method [65].



(a) Raw distribution.



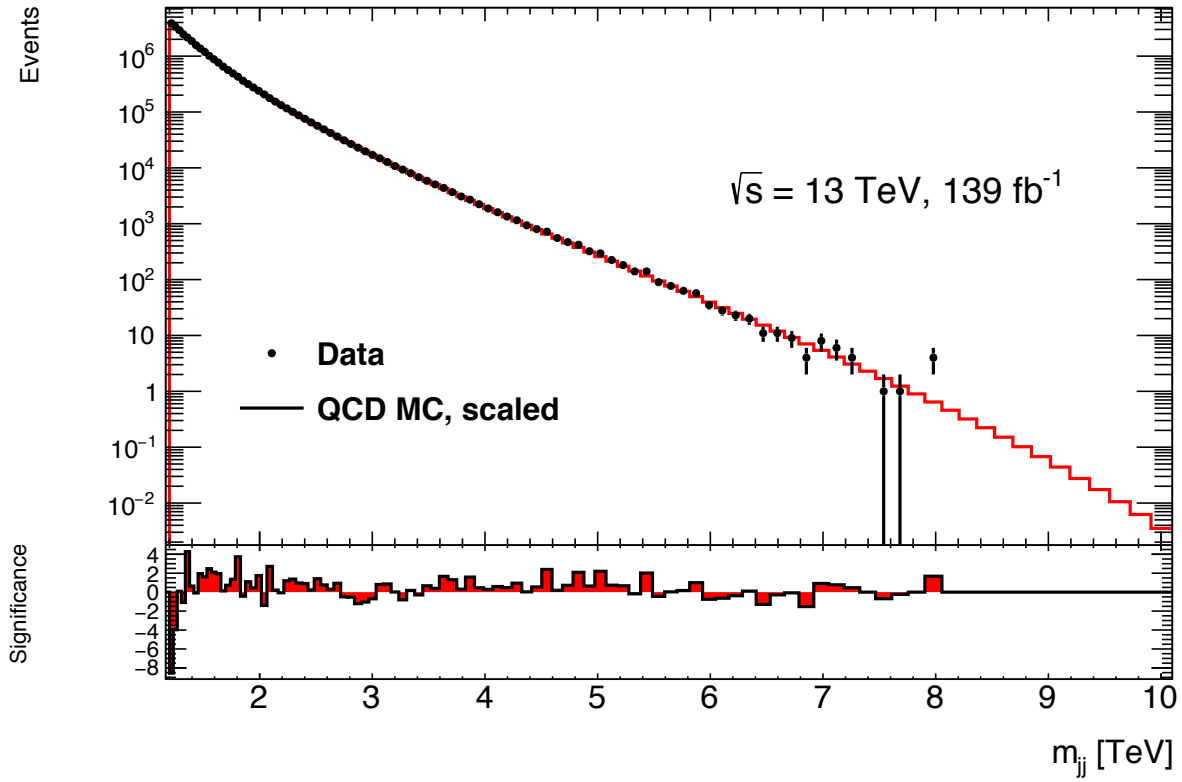
(b) Normalised distribution.

Figure 5.10: QCD MC distribution before and after being normalised to the number of data events.

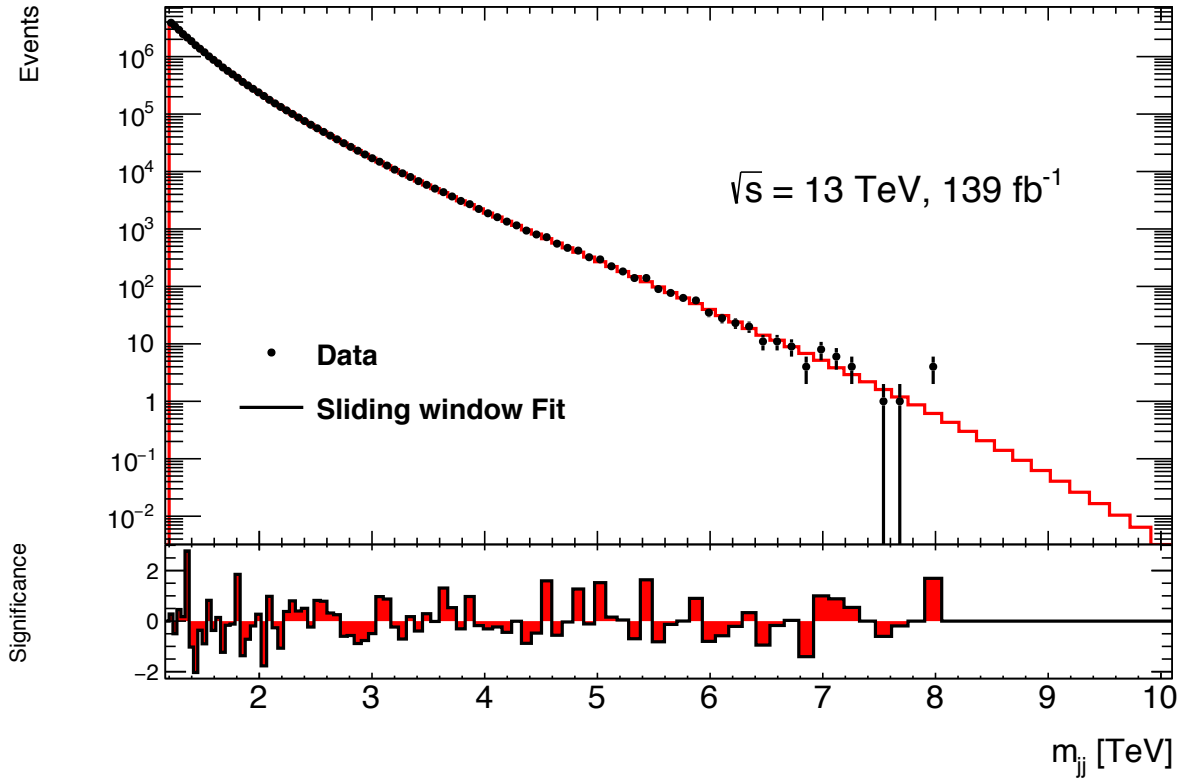
In contrast to a global fit function with the sliding window method the data is fit in overlapping windows centred around each bin, giving the bin-wise background estimate. A global fit function determines the optimal parameter values to fit the entire spectrum. The sliding window fit allows the parameter values to vary in separate regions of the distribution. The algorithm performs both a background fit and a resonance search to ensure that the background-only spectrum is not influenced by any signal that may be present in the data. It takes as input two background-only functions, called the nominal and alternative. These are the four and five-parameter versions of Equation 5.8, respectively. In each window background-only and signal-plus-background fits are performed with both the nominal and alternate fit functions. These are evaluated using a χ^2 value to determine if the nominal or alternate function should be chosen in each window. The window size, defined in terms of the window half-width, or the number of mass bins to the left and right of a window center, is 24 bins. The background spectrum is constructed bin-by-bin by performing a likelihood fit to data in each window. The fit in the central bin of each window is used as the background for the given window. The values from all windows are then combined to produce the full background spectrum. When all windows have been scanned a p-value is computed for the likelihood that the data contains a signal. If an excess is identified in the data, corresponding to a p-value less than 0.01, then the signal is subtracted from the data and a new background fit is produced from the data-minus-signal distribution [65]. Our fit corresponds to a p-value of 0.81, indicating that the background-only fit is successful and no resonance has been found in the data so the data-minus-signal distribution is not computed. This background prediction does not involve simulation so is not affected by uncertainties due to MC modeling or MC statistics.

The data and background distributions for the scaled QCD MC and sliding window fit are shown in Figure 5.11. The second panel of each plot shows the significance, defined as the difference between the data and the fit divided by the square root of the data. Significance is set to zero for all bins in which there is no data. The large spikes in the significance at low mass in the MC distribution are due to limited MC statistics. The significances for the MC distribution are consistently larger than the sliding window fit. Furthermore, significances with MC are one-sided, they are positive for the majority of the distribution. This indicates that the MC QCD consistently under-represents the data for large regions of the m_{jj} spectrum, making it a poor fit. This is also seen in the fractional difference between the data and background in Figure 5.12. The error bars in both plots show the error on the background distributions. The fractional differences with the sliding window fit are more evenly distributed around zero, indicating less bias in the fit than QCD MC. The sliding window background fit is used as the background estimate since it is a better fit to data than the QCD MC. The MC distribution will be used to test and validate the sliding window background, since it provides a signal-free mass distribution that approximately follows the shape of the data.

We perform tests to validate the background spectrum and determine whether or not it includes any bias. This can be done using the scaled QCD MC distribution as the basis for pseudoexperiments, which are Poisson fluctuations of the nominal distribution. First a signal injection test is performed, which requires adding signal to the expected background to test if the fit to the combined spectrum is able to calculate the correct signal yield, or how many

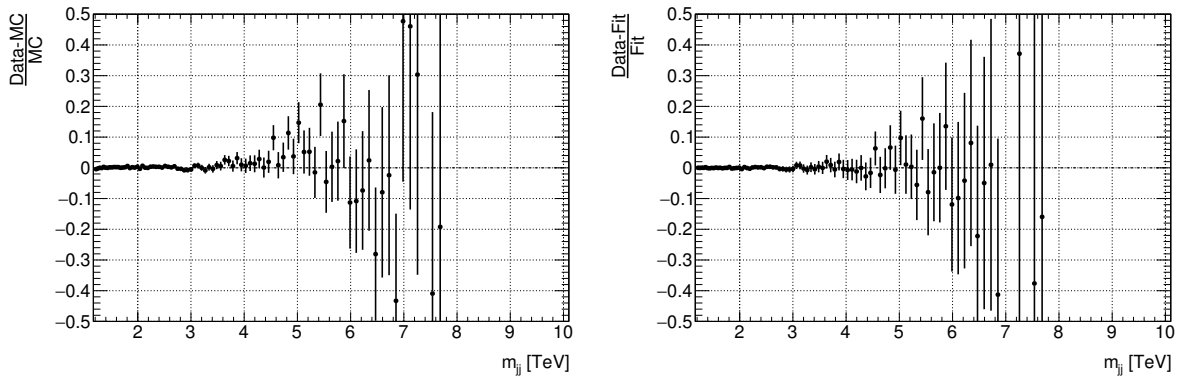


(a) Scaled QCD MC.



(b) Sliding window fit.

Figure 5.11: Data with normalised QCD MC distribution and sliding window background fit. The bottom panel shows significance, defined as the difference between the data and the fit divided by the square root of the data. If there is no data then the significance is set to zero.



(a) Scaled QCD MC.

(b) Sliding window fit.

Figure 5.12: Fractional difference between data and QCD MC distribution, and sliding window background fit.

signal events are present. In this test we take the sliding window background distribution and inject a signal with various strengths to obtain a signal-plus-background fit. We then generate 500 pseudoexperiments as signal-plus-background distributions and fit these to obtain a new background-only distribution. To calculate the signal yield we subtract the number of events in each bin of the background fit from the signal-plus-background distribution and compare this to how many signal events were injected. We anticipate that the signal-plus-background fit will be able to identify signals with a high statistical significance over background. To test the sensitivity of the background it is more interesting to inject a lower amount of signal. The string samples with $M_s = 7.0$, 7.5, and 8.0 TeV represent a signal with over 3σ statistical significance using the method described in Section 6.1, while samples with $M_s = 8.5$ and 9.0 TeV represent a signal with 0σ and 1σ statistical significance, respectively. We will inject a signal of approximately the same strength for each of the signal samples. A signal injection test is run for each string signal sample with an injected signal cross section corresponding to 1σ statistical significance. We are interested in studying the background distribution in the vicinity of each of the string signal masses so we do not consider all m_{jj} bins when calculating the number of signal events. The range of bins included in the calculation for each signal sample is symmetric about the mean such that approximately 50% of the area under the resonance curve is considered.

The injected cross section and number of events for each signal sample are shown in Table 5.7, along with the results of the tests. The results of the signal injection tests are visualised in

M_s (TeV)	Cross section (fb)	Injected Events	injected/extracted ± stat. unc.
7.0	0.50	23.0	1.7±1.5
7.5	0.31	15.6	1.25±0.89
8.0	0.20	9.84	0.91±0.46
8.5	0.13	6.35	0.85±0.43
9.0	0.0083	0.413	0.6±1.7

Table 5.7: Signal cross section and number of events used to perform signal injection tests for each string sample, along with results.

Figure 5.13. The number of extracted signal events and its statistical uncertainty are obtained

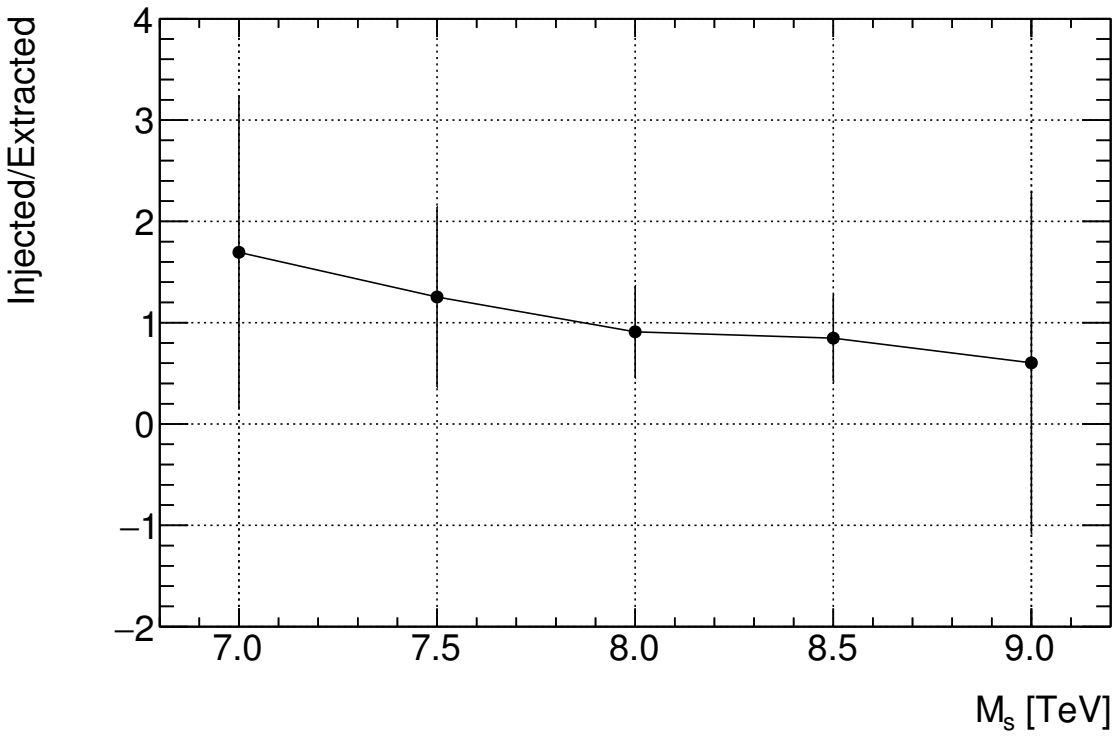


Figure 5.13: Signal injection test results with 500 pseudoexperiments for each of the signal samples.

by taking the mean and standard deviation of 500 pseudoexperiments. There is no uncertainty on the number of injected signal events. We find that the extracted signal yield is consistent with the injected signal within the statistical uncertainty for all string samples. There is a downward trend in the ratio of injected to extracted signal events as M_s increases. This indicates that there may be some bias in the background distribution leading to fewer signal events being identified in the lower M region. The uncertainty on the number of extracted signal events is largest for the minimum and maximum M_s values. The M_s value with the ratio of injected to extracted signal events closest to one corresponds to the ratio with the lowest amount of uncertainty. Since the largest discrepancies between injected and extracted signal events also have the largest amount of uncertainty this trend will not be treated as a systematic uncertainty. Given the small number of pseudo experiments and the large amount of uncertainty on the number of extracted signal events we cannot conclude that there is a significant bias in the background distribution.

Spurious signal is defined as the amount of signal extracted from a distribution with no real signal present, and it quantifies the amount of bias in a distribution. For a spurious signal test we generate 500 pseudoexperiments as Poisson fluctuations of the scaled QCD MC distribution and fit them with the sliding window method. To extract a signal yield we subtract these fits from the nominal background spectrum to check for excess or deficit in distributions with no injected signal. The signal yield is then used as an estimate for the amount of spurious signal, which indicates the efficiency of the fit function in modelling the background. To determine the level of bias in the background fit we calculate the mean number of extracted signal events from 500 pseudoexperiments. This is shown for each of the signal masses in Figure 5.14 where

the error bars represent the standard deviation. The same m_{jj} range as above is used where

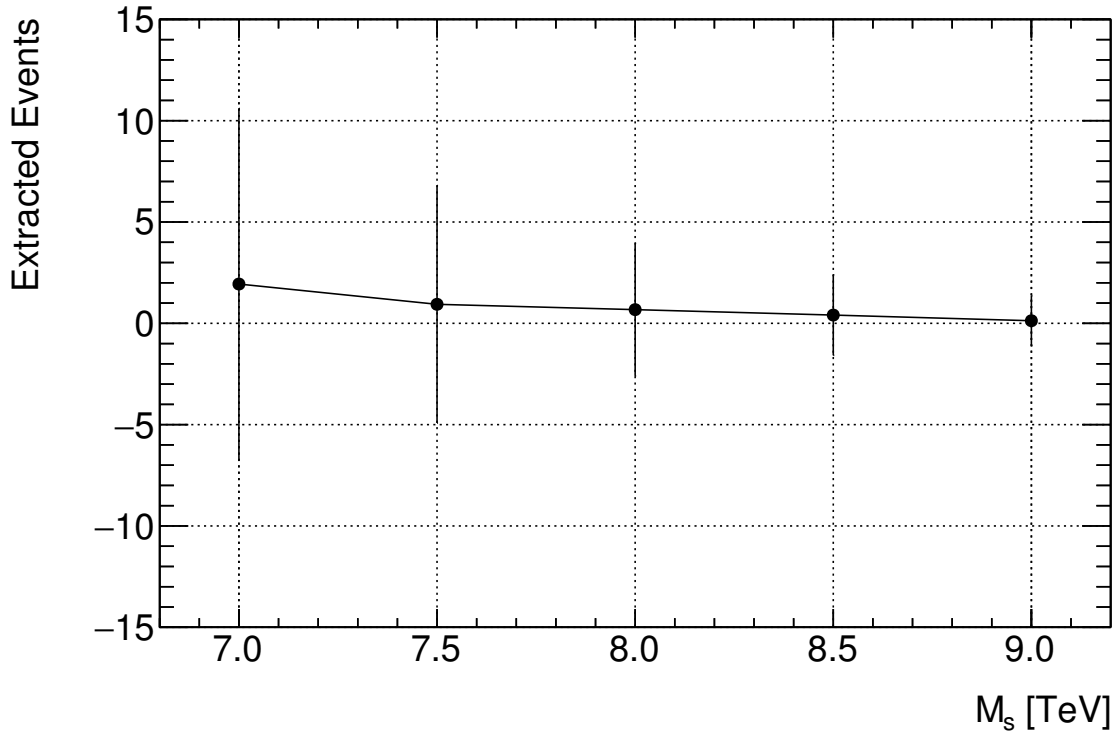


Figure 5.14: Spurious signal test results for each of the signal samples. The mean and standard deviation of extracted signal events are calculated from 500 pseudo-experiments.

only bins symmetric about the mean that make up approximately 50% of the total signal events are considered. The largest number of extracted signal events is approximately 2 for the $M_s = 7.0$ TeV signal sample. There is a high degree of uncertainty on this value so it is within 1σ from its expected value. The uncertainty could be reduced by producing a larger number of pseudoexperiments but the fitting process requires a large amount of time and computational resources so this was not possible for the purposes of our analysis. No significant bias is found in the background spectrum for all string signal masses.

5.5 Uncertainty

All of the samples used in this analysis have some level of uncertainty due to various factors that are out of our control. When we fit the data against our model this will become important to determine the quality of the fit. The data are integer event counts so there is no associated uncertainty but there is some statistical uncertainty in the form of Poisson fluctuations.

The most significant systematic uncertainty in MC signal samples are those associated with modelling the JES and JER. These are estimated using jets in 13 TeV data and simulation [68] and are modelled according to the uncertainties described in Section 4.4. The JES and JER uncertainties are represented by nuisance parameters (NPs). There are three NPs associated with JES and seven associated with JER. This is the smallest number of NPs that can be used to represent the uncertainties without significant loss of information [69]. Each NP corresponds

to two distributions, one for a shift up in mass and one for a shift down. The JES uncertainty changes the m_{jj} of the nominal signal histogram, shifting it either higher or lower. The JER uncertainty changes the width of the nominal signal histogram, shifting it either larger or smaller. It is found that the NP with the most significant impact is the JES 3 NP for all string signal samples. The distribution of both up and down shifts of this NP along with the nominal distribution for $M_s = 7.0$ TeV are shown in Figure 5.15. The nominal distributions

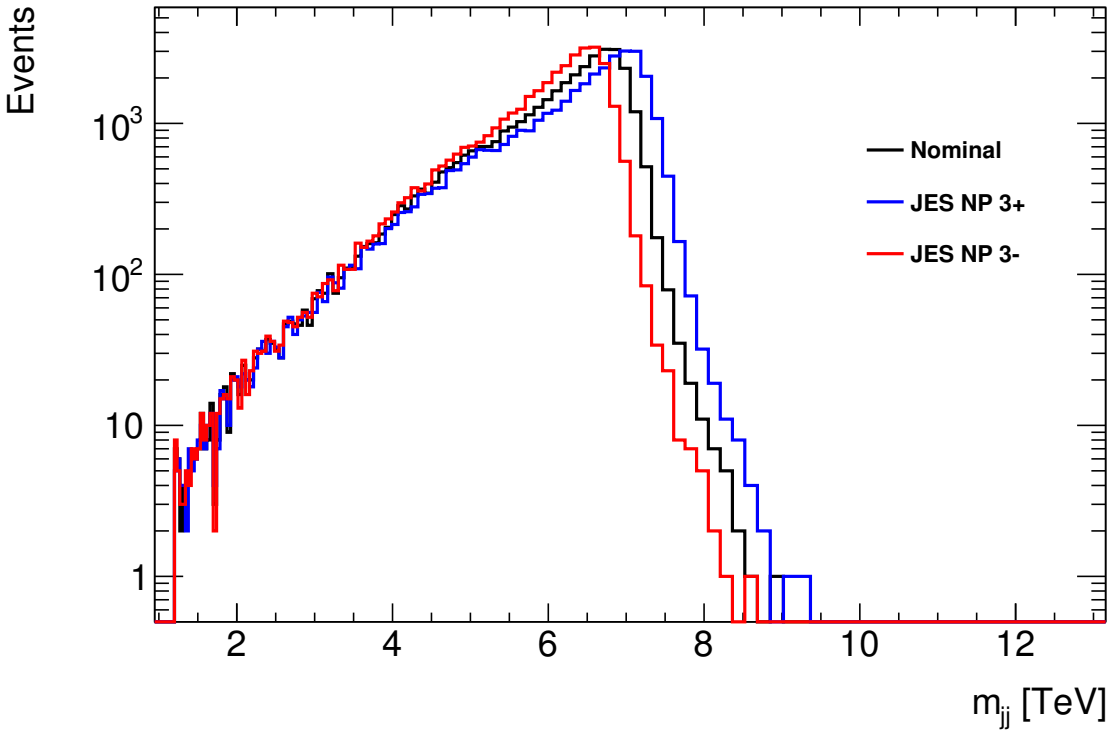


Figure 5.15: JES NP 3 up and down shift with $M_s = 7.0$ TeV nominal distribution.

compared to all three JES NPs and all seven JER NPs for each string sample can be found in Appendix I. The effect of JES on the signal samples is studied by considering the shift in the mean of the distribution, while the effect of JER is studied by considering the change in the standard deviation of the signal distribution. Both fluctuations are applied to the signal spectra with NPs for each uncertainty added in quadrature in Figure 5.16. JES uncertainty is around 1% and JER uncertainty is around 3-4% depending on M_s. This procedure is used as a way to visualise the effect that JES and JER have on our signal samples. However, in our analysis each of the NPs will be treated as a fluctuation of the nominal distribution separately.

Ideally the uncertainty on the background distribution due to the choice of fit parameters can be derived as a confidence band on the function determined by the covariance matrix of the fitted parameters. In this case the parameters of the function are strongly correlated so it is difficult to accurately compute the covariance matrix. Additionally, the sliding window fit contains different parameter values in each window so determining the uncertainty on each would be computationally intensive. The confidence interval of a function represents the 1σ region in which the fit would fall in a large number of repeated trials so it can be obtained from fitting many pseudoexperiments. This method does not require precise estimation of the uncertainty

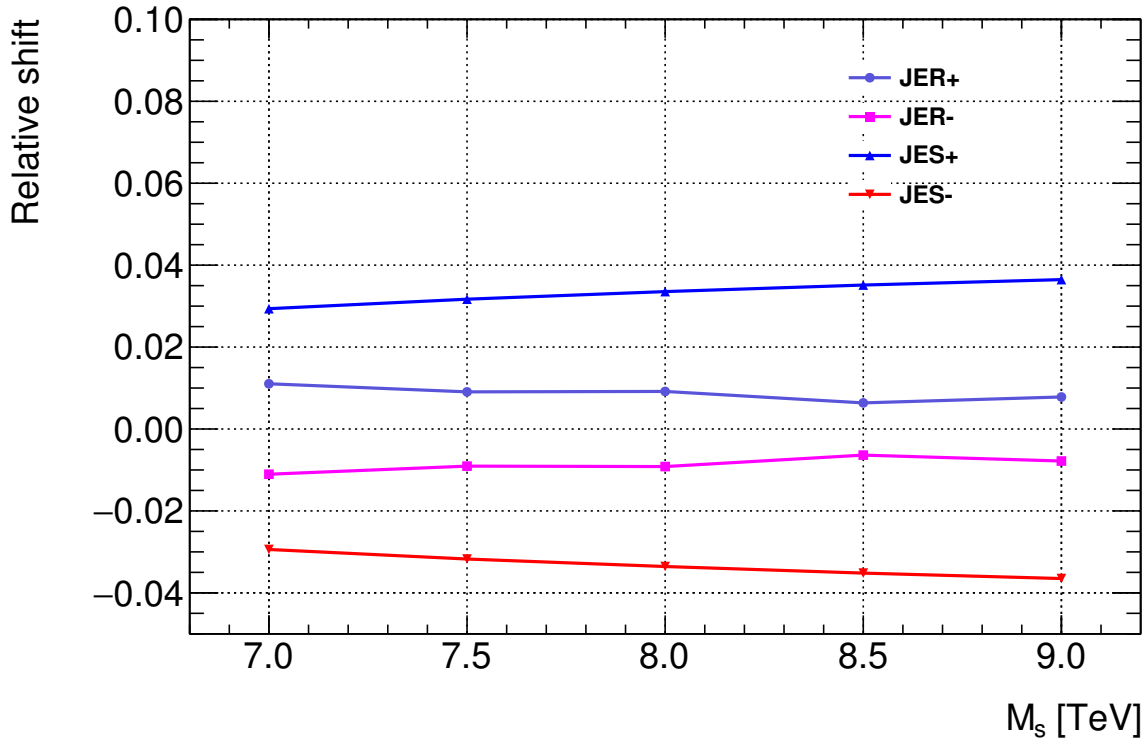


Figure 5.16: Relative shift in the mean of the m_{jj} distributions for each string signal sample due to JES uncertainty, and relative change in the standard deviation of the m_{jj} distributions for each sample due to the JER uncertainty.

on each of the parameters in the fit. To determine the statistical uncertainty in the background fit we follow the procedure used in the previous ATLAS dijet analysis [57]. We produce 10000 pseudoexperiments from the nominal background spectrum and fit them using the same settings and initial conditions as the fit to data. The error on the fit in each bin is defined as the root mean square of the function value in that bin for all pseudoexperiments. The background distribution and relative uncertainty are shown in Figure 5.17. Statistical uncertainties should be random in nature while the error on the background is not. The background distribution is obtained from fitting the data so the uncertainty on background is treated in a systematic way rather than statistical. Any statistical fluctuations in data will influence the computation of the uncertainty on the background distribution so to avoid including these fluctuations more than once we will treat the error on background as a systematic uncertainty. The error on luminosity is also included in the analysis as a systematic uncertainty, this applies to both signal and background.

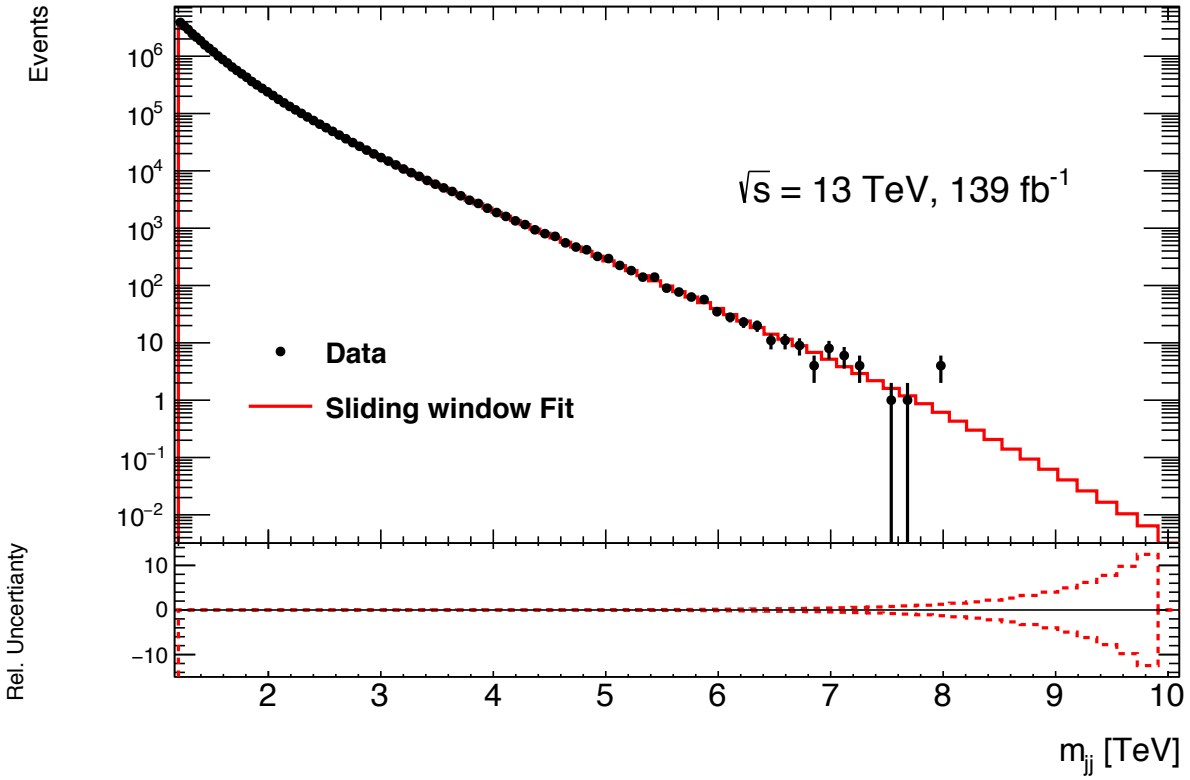


Figure 5.17: Sliding window background fit and data m_{jj} distributions with the relative uncertainty on the background fit shown in the lower panel.

6 Results

6.1 Resonance Search

We search the data for resonant behaviour, visible as an excess over the expected background. We compare the data to our signal model in Figure 6.1. The signal model is the simulated string resonance samples with varying string scales plus the expected background distribution. The signal model is a guideline for the distribution that we expect the data to follow if our signal hypothesis is correct. The amount of signal present in the data spectrum is quantified using BumpHunter [70], which locates the largest excess in the data distribution and calculates its statistical significance.

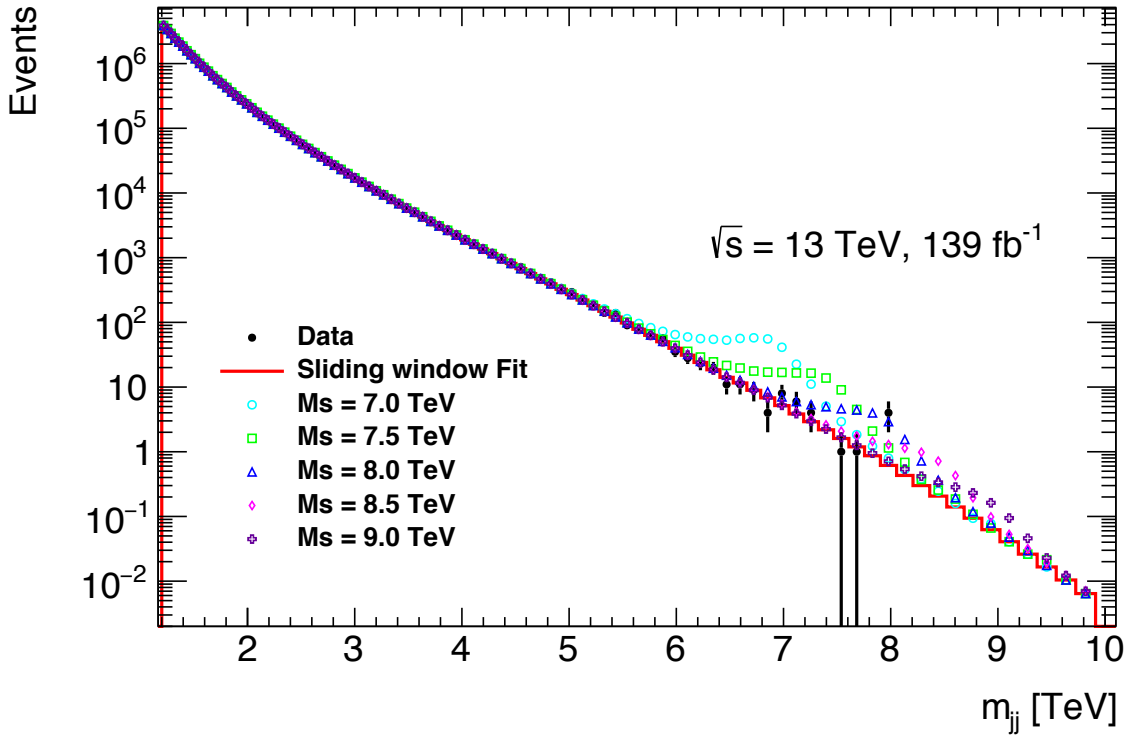


Figure 6.1: Data, sliding window background fit, and signal model with M_s = 7.0, 7.5, 8.0, 8.5, and 9.0 TeV.

When searching to discover a new phenomenon we define the null hypothesis H_0, which describes only known processes. Here our H_0 is the background-only hypothesis. The BumpHunter algorithm compares the data to H_0 and a discovery can be claimed by establishing inconsistency of the data with known processes, in this case the SM. This is convenient because it is model-independent and does not require a specific signal hypothesis for the existence of a discovery. Once a discovery is established tests can be performed to categorise it.

The BumpHunter algorithm analyses the data for an excess over background in each window in the search, these are separate from the windows used in the sliding window fit. Window widths here are variable, ranging from two m_{jj} bins to half of the m_{jj} distribution. In each window the number of data (d) and background (b) events are counted and the BumpHunter test statistic

t_0 and p-value are calculated. The test statistic is defined as

$$t = \begin{cases} 0 & d \leq b, \\ f(d-b) & \text{otherwise}, \end{cases} \tag{6.1}$$

where f can be any positive monotonically increasing function such as $(d-b)^2$. The p-value is defined as the probability that, assuming the background-only hypothesis, the test statistic will be greater than or equal to its calculated value obtained by comparing the data to the background:

$$p\text{-value} = P(t \geq t_0 | H_0). \tag{6.2}$$

Pseudoexperiments are generated from H_0 and t is computed in each interval comparing the data to the pseudoexperiments. The p-value is estimated as a binomial success probability using Bayes' theorem and depends on how the pseudoexperiments fluctuate around H_0. The p-value can be interpreted as a false-discovery probability since if we use it to declare discovery when the p-value $\leq \alpha$ then the probability to wrongly rule out the background-only hypothesis is α. The p-value is defines as

$$p\text{-value} = \begin{cases} 1 & d \leq b, \\ \mathscr{P}(d,b)(1-10^3)^2 & \text{otherwise}, \end{cases} \tag{6.3}$$

where

$$\mathscr{P}(d,b) = \begin{cases} \Gamma(d,b) & d \geq b, \\ 1 - \Gamma(d+1,b) & \text{otherwise}. \end{cases} \tag{6.4}$$

This calculation requires use of the relation

$$\Gamma(d,b) = \sum_{n=d}^{\infty} \frac{b^n}{d!} e^{-b} = \frac{1}{\Gamma(d)} \int_0^b t^{d-1} e^{-t} dt, \tag{6.5}$$

where $\Gamma(d)$ represent the Gamma function.

The p-value describes the probability of the background-only hypothesis creating an excess as significant as the largest one identified in the window. These steps are performed in each window for every possible window width. The interval with the largest excess from the background model is defined as the window that is least likely to arise from Poisson background fluctuation and corresponds to the window with the minimum p-value. The statistical significance $\tilde{S}$ of the largest identified excess can be calculated with [70].

$$\tilde{S} = \ln(p\text{-value}_{min}) \tag{6.6}$$

The background distribution is obtained by fitting the data so there is the possibility that if an excess exists in the data it will influence the background. To ensure that the background is not influenced by any signal that may be present any excess in data with a p-value less than 0.01 can be excluded from the data and the background fit can be rerun.

Results from BumpHunter are shown in Figure 6.2 with the most discrepant region between

the vertical lines. The most significant excess in data is found in the region 1.305-1.378 TeV with a p-value of 0.89. The p-value is greater than 0.01, indicating that no significant excess above the background spectrum has been found in the data. This value corresponds to the background-only hypothesis, not to a discovery.

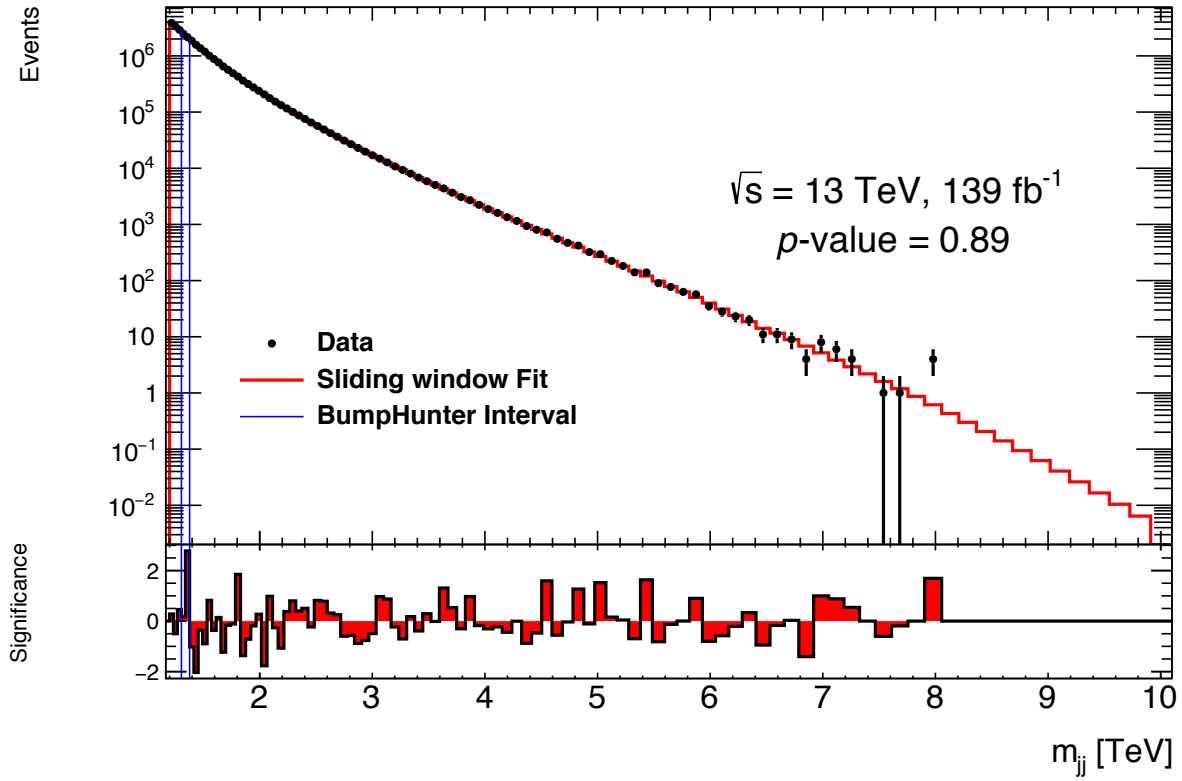


Figure 6.2: BumpHunter results with data and the sliding window background fit. The largest excess in the data spectrum is identified as 1.305-1.378 TeV, with a p-value of 0.89.

6.2 Limit Calculation

In the previous section we searched the data m_{jj} distribution for an excess over the expected background. No significant deviation from the expected background has been observed but constraints on the signal model can be computed. These include both upper limits on the cross section of the signal and lower limits on the string scale of our model. This allows us to exclude regions of the model parameter space where string resonances are not present. Limits are computed using a confidence level (CL) [71], which calculates the level of signal present in the data. We use frequentist statistics based on likelihood functions to compute these CLs. In accordance with other limit studies a 95% CL, corresponding to a p-value of 5%, excludes regions where a signal is not compatible with the data [72]. We test the background-only hypothesis H_0 against H_1, which contains both background and our signal. To determine the outcome of a search for an undiscovered signal we seek to quantify the level of agreement between our data and a given hypothesis by computing a p-value. This p-value represents the probability that our observations will be of equal or greater incompatibility with the predictions of our hypothesis.

The hypothesis can then be excluded if its p-value is below a certain threshold. We use the statistical formulation from Ref. [73] and Ref. [61], outlined below.

To calculate how well our signal model represents the data we can define the likelihood function $L(\mu, \theta)$. This is a function of parameters that can be varied to alter the signal model and measures how well the values of each parameter fit the data. For example if $L(\mu_1, \theta) > L(\mu_2, \theta)$ then the sample we have observed is more likely to have occurred if $\mu = \mu_1$. We may compute the likelihood function for every value that the parameters can take and the parameter values corresponding to the maximum calculated likelihood represent the data best. The likelihood function of our signal model is the product of Poisson distributions of event counts and the systematic uncertainties.

If we have S number of signal events and B number of background events we can denote their probability density functions as $f_S(x, \theta)$ and $f_B(x, \theta)$. These describe the distributions of the events and are normalised such that $\int dx\, f(x, \theta) = 1$. Here θ are NPs that can be tuned when fitting to data. When we calculate the likelihood function in each bin there are several NPs corresponding to systematic uncertainties that we allow to vary. We can then introduce μ as a parameter that determines the strength of the signal process. The minimum value of μ is 0, which corresponds to the background-only hypothesis and $\mu = 1$ corresponds to the signal hypothesis. This is our parameter of interest. We are interested in determining the probability of obtaining n data events where our search variable (m_{jj} in this case) for event i has value x_i. The full data set is denoted as $\{x_1, ..., x_n\}$. We must include the Poisson probability of obtaining n events when $\mu S + B$ events are expected, and the probability density of obtaining x_i for a given value of μ. From this we obtain a "marked Poisson model" [73]:

$$L(\{x_1, .., x_n\}|\mu, \theta) = \text{Pois}(n|\mu S + B) \left[\prod_{i=1}^{n} \frac{\mu S f_S(x_i, \theta) + B f_B(x_i, \theta)}{\mu S + B} \right]. \tag{6.7}$$

Systematic uncertainties correspond to NPs α_l that fluctuate between 0 and 1. 0 represents the nominal distribution while 1 represents one standard deviation away from the nominal. The systematic uncertainties are the product of various distributions depending on the effect of the uncertainty on the nominal distribution. They are collectively represented by C_{syst}. Each NP can be approximated by a Gaussian, Poisson, or log-normal distribution. We use the Gaussian parametrisation

$$C_{syst} = \sum_{l=1}^{N} \ln G(\alpha_l^0 - \alpha_l), \tag{6.8}$$

where N is the number of systematic uncertainties and α_l^0 are the central values around which the α_l can be varied. These functions are multiplied by the Poisson model to account for the systematic uncertainties on signal and background. Each bin corresponds to a different value of α_l^0 and α_l for every systematic uncertainty. These functions multiply Equation 6.7 to give

$$L(\{x_1, .., x_n\}|\mu, \theta) = \text{Pois}(n|\mu S + B) \left[\prod_{i=1}^{n} \frac{\mu S f_S(x_i, \theta) + B f_B(x_i, \theta)}{\mu S + B} \right] \times C_{syst}. \tag{6.9}$$

To test a hypothesised value of μ we consider the profile likelihood ratio defined as

$$t_\mu = -2 \ln \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}, \tag{6.10}$$

where $\hat{\mu}$ and $\hat{\theta}$ correspond to the absolute maximal value of L, while $\hat{\hat{\theta}}$ maximises L for a given value of μ. Higher values of t_μ correspond to greater incompatibility between the data and the signal-plus-background model with a specified value of μ [61]. We can use

$$\text{Pois}(n|\nu) = \nu^n \frac{e^{-\nu}}{n!}, \tag{6.11}$$

to rewrite the log of L as

$$-\ln L(\mu, \theta) = (\mu S + B) + \ln n! - \sum_{i=1}^{n} \ln[\mu S f_S(x_i, \theta) + B f_B(x_i, \theta)] - \sum_{l=1}^{N} \ln G(\alpha_l^0 - \alpha_l). \tag{6.12}$$

To quantify the level of disagreement between the data and our model we can compute a p-value, defined as

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) dt_\mu, \tag{6.13}$$

where $t_{\mu,obs}$ is the value of t_μ observed from the data and $f(t_\mu|\mu)$ is the probability density function of t_μ assuming signal strength μ. If there is a large number of data events then the bias of maximum likelihood estimators tends to zero and t_μ can be approximated by a Gaussian function with other terms neglected. In this approximation the cumulative distribution of t_μ is

$$F(t_\mu|\mu) = 2\Phi(\sqrt{t_\mu}) - 1, \tag{6.14}$$

where Φ is the cumulative distribution of the standard Gaussian. It follows that the p-value of a hypothesised value of μ for an observed value t_μ is [61]

$$p_\mu = 1 - F(t_\mu|\mu) = 2\left(1 - \Phi\left(\sqrt{t_\mu}\right)\right). \tag{6.15}$$

We will take the value of μ for which the median p-value is 0.05 as the expected upper limit on μ at a 95% confidence level.

First we run a signal-plus-background fit where we fit our signal model to the data. The background is not modified but the signal is scaled to best fit the data. This calculates the parameter μ, which describes the signal strength after fitting, using the statistical framework outlined above. To perform the fit all parameters are set to a constant except for one that is allowed to float. This parameter is varied and the value that yields the largest calculated likelihood is determined. This process is repeated for each parameter until the maximum possible L has been calculated. The value of the signal strength for which the p-value falls below 5% is computed for a 95% CL upper limit. Multiple hypothesis tests are executed to evaluate the CL values for a wide range of signal strengths. A second scan follows in a smaller interval using the expected upper limit derived from the first scan [72].

The results of the limit scan for the $M_s = 7.0$ TeV signal model are shown in Figure 6.3. The p-value is computed for several values of μ along with the associated CLs. Black, blue, and red dots represent the CL for the background-only, signal-plus-background, and signal hypothesis, respectively. The black dashed line is the linear interpolation of the expected value of the signal hypothesis CL. These lines are not exactly equal due to statistical fluctuations in the data. The green and yellow bands are 1σ and 2σ intervals on the median expected CL line, implying that the CL should lie within the bands 68% and 95% of the time, respectively. The p-value of 0.05 is denoted by the red horizontal line. The value of μ for which the p-value of the signal CL is 0.05 is the upper bound on μ that is consistent with the data at a 95% CL. Limit scans for other string scales can be found in Appendix J.

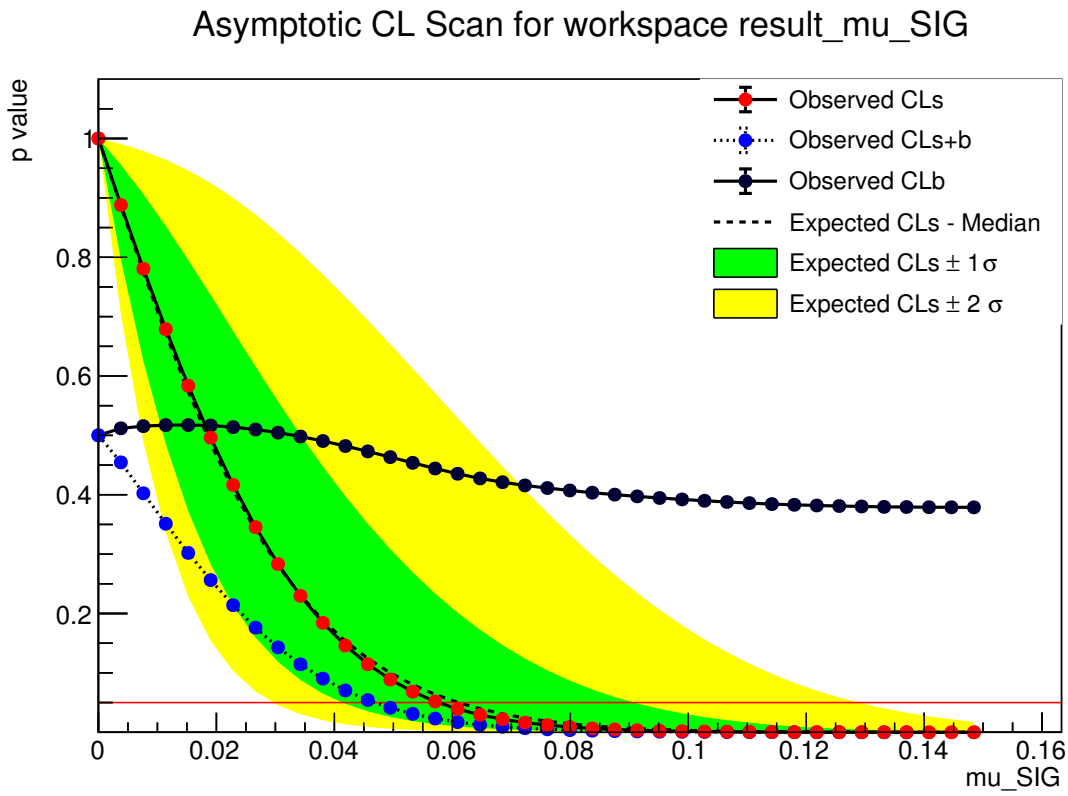


Figure 6.3: Limit scan results for $M_s = 7.0$ TeV string resonance signal model.

The data event with the largest value of m_{jj} is around 8.0 TeV and several of our string signals contain events at higher energies. It is clear from Figure 6.1 that the data does not contain any resonant behaviour but we can still consider signals with resonant peaks at higher energies than the data points. Since there is no data at energies higher than 8.0 TeV we cannot rule out string signals that peak at significantly higher values of m_{jj}. The μ value required to fit the data for lower M_s string signals is very low since the data is very close to background but for higher string masses μ is substantially larger. This is because there are few signal events in the mass region where we have data for samples with M_s greater than the highest m_{jj} bin in data. It is possible for μ to be greater than 1 since there is no data near the resonant peak. The expected and observed upper limit on μ at a 95% CL for each of the string signals is shown in Table 6.1.

M_s [TeV]	Expected	Observed
7.0	0.061	0.058
7.5	0.150	0.171
8.0	0.402	0.459
8.5	1.231	1.649
9.0	4.288	4.313

Table 6.1: Expected and observed upper limits on μ at a 95% CL.

The number of observed signal events N is related to the cross section of the signal σ by the following equation:

$$\sigma = \frac{N}{LA\epsilon B}, \tag{6.16}$$

where L is the luminosity recorded by ATLAS, A is the fraction of accepted signal events after selection criteria, ϵ is the efficiency of the signal reconstruction due to algorithms and detector effects, and B is the resonance branching ratio of the final state of interest. Here we take $L = 139$ fb^{-1}, and $B = 1$ since we consider all processes inclusively. The number of events lost due to detector effects is taken as ϵ. Since MC samples model the detector in good working order no signal events are lost to errors in the calorimeters or other instrumentation. All events are able to be reconstructed in signal samples so we will treat detector inefficiencies as negligible and take $\epsilon = 1$. We can use the upper limit on μ calculated by the limit scan to determine the upper limit on number of events given by the signal model by multiplying μ by the total number of signal events. This can then be used to determine the upper limit on $\sigma \times \epsilon \times A$ with Equation 6.16. The calculated 95% CLs are shown in Figure 6.4. The solid black points give the observed upper limit on $\sigma \times \epsilon \times A$ with 95% CL. The black dashed line is the linear interpolation of the expected value at a 95% CL. The green and yellow bands are 1σ and 2σ intervals on the median expected upper limit line. The blue dashed line represents the theoretical $\sigma \times \epsilon \times A$ for each string scale computed using the STRINGS event generator. If the expected cross section limit in our signal hypothesis is larger than the cross section predicted by the CL method in a certain mass region then we can exclude that mass region with 95% CL. As seen in Figure 6.4 the expected upper limit on string resonances is $M_s = 8.40$ TeV and the observed limit is $M_s = 8.30$ TeV. The blue dashed line is above the observed 95% upper limit in the region $M < 8.30$ TeV, which indicates that we must reject the signal hypothesis in this mass region as we have a 95% CL that string resonances do not occur here.

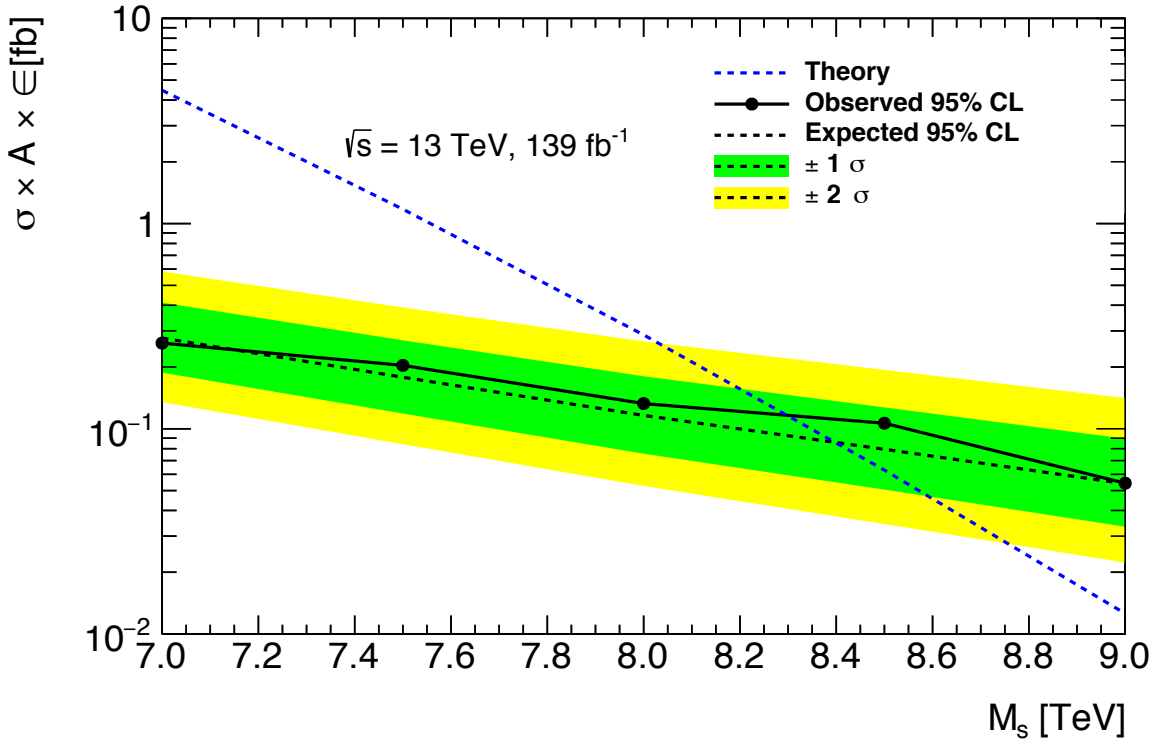


Figure 6.4: 95% CL plot used to set limits on the observable string signal for $M_s = 7.0$, 7.5, 8.0, 8.5, 9.0 TeV. The expected upper limit on string resonances is $M_s = 8.40$ TeV and the observed limit is $M_s = 8.30$ TeV.

7 Summary

The SM is the most complete theory that we have presently to explain the particles that we observe and the ways in which they interact. Since its formulation we have discovered several problems within the theory. There are experimental observations that cannot be explained fully by the SM in its present state. This leaves us with the conclusion that we must alter the SM or replace it entirely if we wish to have a model that describes everything that we have observed. One such alternative is string theory, which is proposed to replace the SM. This theory is self-contained and encompasses all forces that we observe, while the SM does not include gravity. There are several reasons why this model is attractive, mainly owing to the fact that outstanding questions within the SM may be answered by string theory. The Dp-brane formulation of string theory allows for the possibility of large compactified extra spatial dimensions that we have not yet observed. If this formulation is correct then the scale of string theory can be on the order of TeV, accessible with current collider technology. This formulation of string theory provides an elegant solution to the hierarchy problem, which questions why the fundamental scale of gravity is many orders of magnitude larger than that of the other forces. In this model resonances occur in proton-proton collisions when the energy of the quarks or gluons within the colliding protons approaches the energy scale of the string theory M_s. This phenomenon should be measurable at the LHC if M_s is indeed on the order of TeV.

We used theoretical string theory interaction amplitudes to search for evidence of string resonances in proton-proton collisions at the LHC. We simulated quark/gluon interactions with string resonances for five string scales between $7.0 < M_s < 9.0$ TeV. We estimated QCD background processes by fitting a function to data. The background and signal-plus-background models were compared to data taken with the ATLAS detector between the years 2015-2018. We found no resonance in ATLAS data at energy $\sqrt{s} = 13$ TeV with an integrated luminosity of 139 fb^{-1}. The statistical significance of the largest excess in data was found have a p-value of 0.89. This does not satisfy the criteria for discovery, which is a p-value < 0.01. Limit scans were run on each of the string signal samples to determine the upper limit on the amount of signal that is consistent with data. We set a lower limit on the string scale based on our observations. The previous lower limit on the string scale set by the CMS collaboration in 2018 is $M_s = 7.7$ TeV [49]. Our expected lower limit on the string scale at a 95% CL is $M_s = 8.40$ TeV while the observed limit is $M_s = 8.30$ TeV. We found with 95% CL that string resonances are not present in ATLAS data with $M_s < 8.30$ TeV, improving upon the previous lower limit.

References

[1] N. ArkaniHamed, S. Dimopoulos, and G. Dvali. "The hierarchy problem and new dimensions at a millimeter". In: *Physics Letters B* 429.3-4 (1998), pp. 263–272. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(98)00466-3. URL: http://dx.doi.org/10.1016/S0370-2693(98)00466-3.

[2] B. Zwiebach. *A First Course in String Theory*. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511841682.

[3] D. Lüst, S. Stieberger, and T. R. Taylor. "The LHC string Hunters companion". In: *Nuclear Physics B* 808.1-2 (2009), pp. 1–52. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2008.09.012. URL: http://dx.doi.org/10.1016/j.nuclphysb.2008.09.012.

[4] A. B. Arbuzov. "Quantum Field Theory and the Electroweak Standard Model". In: (2018), 1–34. 35 p. DOI: 10.23730/CYRSP-2017-004.1. eprint: 1801.05670. URL: https://cds.cern.ch/record/2315477.

[5] C. P. Burgess and G. D. Moore. *The standard model: A primer*. Cambridge University Press, 2006. ISBN: 978-0-511-25485-7.

[6] S. V. Troitsky. "Unsolved problems in particle physics". In: *Physics-Uspekhi* 55.1 (2012), pp. 72–95. DOI: 10.3367/ufne.0182.201201d.0077. URL: https://doi.org/10.3367/ufne.0182.201201d.0077.

[7] L. Susskind. "The gauge hierarchy problem, technicolor, supersymmetry, and all that". In: *Physics Reports* 104.2 (1984), pp. 181–193. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(84)90208-4. URL: https://www.sciencedirect.com/science/article/pii/0370157384902084.

[8] R. C. Myers and S. E. Vázquez. "Quark soup al dente: applied superstring theory". In: *Classical and Quantum Gravity* 25.11 (2008), p. 114008. DOI: 10.1088/0264-9381/25/11/114008. URL: https://doi.org/10.1088/0264-9381/25/11/114008.

[9] ATLAS Collaboration. "Search for strong gravity in multijet final states produced in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *Journal of High Energy Physics* 2016.3 (2016). ISSN: 1029-8479. DOI: 10.1007/jhep03(2016)026. URL: http://dx.doi.org/10.1007/JHEP03(2016)026.

[10] C. D. Hoyle et al. "Submillimeter tests of the gravitational inverse-square law". In: *Physical Review D* 70.4 (2004). ISSN: 1550-2368. DOI: 10.1103/physrevd.70.042004. URL: http://dx.doi.org/10.1103/PhysRevD.70.042004.

[11] M. S. Gao, S. R. Yuan, and J. Gao. "Top-quark mass determination from t-channel single top production at the LHC". In: *JHEP* 04 (2021), p. 054. DOI: 10.1007/JHEP04(2021)054. arXiv: 2007.15527 [hep-ph].

[12] A. Banfi. *Hadronic Jets*. 2053-2571. Morgan & Claypool Publishers, 2016. ISBN: 978-1-6817-4073-7. DOI: 10.1088/978-1-6817-4073-7. URL: http://dx.doi.org/10.1088/978-1-6817-4073-7.

[13] R. Placakyte. *Parton Distribution Functions*. 2011. arXiv: 1111.5452 [hep-ph].

[14] P. M. Nadolsky et al. “Implications of CTEQ global analysis for collider observables”. In: *Physical Review D* 78.1 (2008). ISSN: 1550-2368. DOI: 10.1103/physrevd.78.013004. URL: http://dx.doi.org/10.1103/PhysRevD.78.013004.

[15] R. D. Ball et al. “A first unbiased global NLO determination of parton distributions and their uncertainties”. In: *Nuclear Physics B* 838.1-2 (2010), pp. 136–206. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2010.05.008. URL: http://dx.doi.org/10.1016/j.nuclphysb.2010.05.008.

[16] R. D. Ball et al. “A determination of parton distributions with faithful uncertainty estimation”. In: *Nuclear Physics B* 809.1-2 (2009), pp. 1–63.

[17] A. Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *The European Physical Journal C* 75.3 (2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3318-8. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3318-8.

[18] D. Griffiths. *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008. ISBN: 9783527406012. URL: https://books.google.ca/books?id=w9Dz56myXm8C.

[19] M. Hashi and N. Kitazawa. “Signatures of low-scale string models at the LHC”. In: *Journal of High Energy Physics* 2012.2 (2012). ISSN: 1029-8479. DOI: 10.1007/jhep02(2012)050. URL: http://dx.doi.org/10.1007/JHEP02(2012)050.

[20] L. A. Anchordoqui et al. “String resonances at hadron colliders”. In: *Physical Review D* 90.6 (2014). ISSN: 1550-2368. DOI: 10.1103/physrevd.90.066013. URL: http://dx.doi.org/10.1103/PhysRevD.90.066013.

[21] T. Regge. “Introduction to complex orbital momenta”. In: *Il Nuovo Cimento (1955-1965)* 14 (1959), pp. 951–976.

[22] D. Lüst et al. “The LHC String Hunters Companion (II): Five-particle amplitudes and universal properties”. In: *Nuclear Physics B* 828.1-2 (2010), pp. 139–200. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2009.11.008. URL: http://dx.doi.org/10.1016/j.nuclphysb.2009.11.008.

[23] P. Gras et al. “Systematics of quark/gluon tagging”. English. In: *Journal of High Energy Physics* 2017.7 (2017). ISSN: 1029-8479. DOI: 10.1007/JHEP07(2017)091.

[24] L. A. Anchordoqui, H. Goldberg, and T. R. Taylor. “Decay widths of lowest massive Regge excitations of open strings”. In: *Physics Letters B* 668.5 (2008), pp. 373–377. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2008.09.003. URL: http://dx.doi.org/10.1016/j.physletb.2008.09.003.

[25] ATLAS Collaboration. *ATLAS Stand-Alone Event Displays*. 2012. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone.

[26] G. P. Salam. “Towards jetography”. In: *The European Physical Journal C* 67.3-4 (2010), pp. 637–686. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1314-6. URL: http://dx.doi.org/10.1140/epjc/s10052-010-1314-6.

[27] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[28] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003.

[29] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[30] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[31] K. Noda and LHCf Collaboration. "The LHCf experiment". In: *AIP Conference Proceedings* 1492.1 (2012), pp. 182–188. DOI: 10.1063/1.4763514. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4763514. URL: https://aip.scitation.org/doi/abs/10.1063/1.4763514.

[32] G. Apollinari et al. "High Luminosity Large Hadron Collider HL-LHC". In: (2015). DOI: 10.5170/CERN-2015-005.1.

[33] M. Schott and M. Dunford. "Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV". In: *The European Physical Journal C* 74.7 (2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-2916-1. URL: http://dx.doi.org/10.1140/epjc/s10052-014-2916-1.

[34] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *Journal of Instrumentation* 15.10 (2020), P10004–P10004. DOI: 10.1088/1748-0221/15/10/p10004. URL: https://doi.org/10.1088/1748-0221/15/10/p10004.

[35] G. Soyez. "Pileup mitigation at the LHC: A theorists view". In: *Physics Reports* 803 (2019), pp. 1–158. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2019.01.007. URL: http://dx.doi.org/10.1016/j.physrep.2019.01.007.

[36] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. CERN, 2008. URL: https://cds.cern.ch/record/1099735.

[37] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[38] V. Giangiobbe. "Jet calibration in the ATLAS experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 617.1 (2010), pp. 90–91. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.09.117. URL: https://www.sciencedirect.com/science/article/pii/S0168900209018993.

[39] ATLAS Collaboration. *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*. Tech. rep. CERN, 2015. URL: https://cds.cern.ch/record/2044941.

[40] L. Kogan. "Determining the Jet Energy Scale uncertainty in the ATLAS detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 718 (2013), pp. 86–87. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2012.11.048. URL: https://www.sciencedirect.com/science/article/pii/S016890021201385X.

[41] ATLAS Collaboration. *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. 2020. eprint: 2007.02645. URL: https://cds.cern.ch/record/2722869.

[42] P. Vakilipourtakalou and D. M. Gingrich. *Monte Carlo event generator for the production and decay of string resonances in proton-proton collisions*. 2018. arXiv: 1811.07458 [hep-ph]. URL: https://strings.hepforge.org/.

[43] J. Alwall et al. "A standard format for Les Houches Event Files". In: *Computer Physics Communications* 176.4 (2007), pp. 300–304. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2006.11.010. URL: http://dx.doi.org/10.1016/j.cpc.2006.11.010.

[44] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. CERN, 2019. URL: https://cds.cern.ch/record/2677054.

[45] T. Sjostrand, S. Mrenna, and P. Skands. "A brief introduction to Pythia 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. eprint: 0710.3820.

[46] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419.

[47] S. Höche. "Introduction to parton-shower event generators". In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. 2014. DOI: 10.1142/9789814678766_0005. eprint: 1411.4085.

[48] Z. Marshall. *Re-re-defining the standard QCD di-jet samples: beginning to like event weights*. ATL-COM-PHYS-2015-417. 2015. URL: https://cds.cern.ch/record/2016630.

[49] V. Khachatryan et al. "Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev.* D91.5 (2015), p. 052009. DOI: 10.1103/PhysRevD.91.052009. eprint: 1501.04198.

[50] V. Khachatryan et al. "Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{(s)} = 13$ TeV". In: *Phys. Rev. Lett.* 116.7 (2016), p. 071801. DOI: 10.1103/PhysRevLett.116.071801. eprint: 1512.01224.

[51] A. M. Sirunyan et al. "Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles". In: *JHEP* 08 (2018), p. 130. DOI: 10.1007/JHEP08(2018)130. eprint: 1806.00843.

[52] T. Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. eprint: 1410.3012.

[53] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[54] R. D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys.* B867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003. eprint: 1207.1303.

[55] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model.* ATL-PHYS-PUB-2016-017. 2016. URL: https://cds.cern.ch/record/2206965.

[56] L. A. Beresford. "Dijet Invariant Mass Spectra". In: *Searches for Dijet Resonances: Using $\sqrt{s}$ = 13 TeV Proton–Proton Collision Data Recorded by the ATLAS Detector at the Large Hadron Collider*. Springer International Publishing, 2018, pp. 69–97. ISBN: 978-3-319-97520-7. DOI: 10.1007/978-3-319-97520-7_5. URL: https://doi.org/10.1007/978-3-319-97520-7_5.

[57] ATLAS Collaboration. "Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *JHEP* 03 (2020), p. 145. DOI: 10.1007/JHEP03(2020)145. eprint: 1910.08447 (hep-ex).

[58] ATLAS Collaboration. *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*. Tech. rep. CERN, 2015. URL: https://cds.cern.ch/record/2037702.

[59] M. Vanadia et al. *Search for New Phenomena in Dijet Events with the ATLAS Detector at $\sqrt{s}$=13 TeV with 2016 data*. Tech. rep. CERN, 2016. URL: https://cds.cern.ch/record/2151829.

[60] R. M. Harris and K. Kousouris. "Searches for dijet resonances at hadron colliders". In: *Int. J. Mod. Phys.* A26 (2011), pp. 5005–5055. DOI: 10.1142/S0217751X11054905. eprint: 1110.5302.

[61] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[62] ATLAS Collaboration. *Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector*. ATL-PHYS-PUB-2017-009. 2017. URL: https://cds.cern.ch/record/2263679.

[63] ATLAS Collaboration. "Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Phys. Lett. B* 754 (2016), p. 302. DOI: 10.1016/j.physletb.2016.01.032. eprint: 1512.01530.

[64] N. Nishu et al. *Search for New Phenomena in Dijet Events using the complete Run-2 dataset collected with the ATLAS Detector at $\sqrt{s}$ =13 TeV*. Tech. rep. ATL-COM-PHYS-2018-1538. CERN, 2018. URL: https://cds.cern.ch/record/2646455.

[65] ATLAS Collaboration. "Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Physical Review D* 96.5 (2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.052004. URL: http://dx.doi.org/10.1103/PhysRevD.96.052004.

[66] ATLAS Collaboration. *Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. CERN, 2015. URL: https://cds.cern.ch/record/2114853.

[67] ATLAS Collaboration. "Search for new phenomena in dijet events using 37fb[1] of *pp* collision data collected at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Physical Review D* 96.5 (2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.052004. URL: http://dx.doi.org/10.1103/PhysRevD.96.052004.

[68] ATLAS Collaboration. "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$ = 13TeV with the ATLAS detector". In: *Physical Review D* 96.7 (2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.072002. URL: http://dx.doi.org/10.1103/PhysRevD.96.072002.

[69] D. Börner. "A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale". In: *European Physical Journal Web of Conferences*. Vol. 137. European Physical Journal Web of Conferences. Mar. 2017, p. 11003. DOI: 10.1051/epjconf/201713711003.

[70] G. Choudalakis. *On hypothesis testing, trials factor, hypertests and the BumpHunter*. 2011. arXiv: 1101.0390 [physics.data-an].

[71] A. L. Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/0954-3899/28/10/313.

[72] M. Baak et al. "HistFitter software framework for statistical data analysis". In: *The European Physical Journal C* 75.4 (2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3327-7. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3327-7.

[73] K. Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York U., 2012. URL: https://cds.cern.ch/record/1456844.

Appendices

Appendix A STRINGS-2.00 Kinematics

STRINGS-1.00 treats all incoming and outgoing partons as massless particles. The kinematic variables have been modified in STRINGS-2.00 to include the masses of quarks in the final state. This requires the derivation of new equations for the energy and momentum of outgoing partons.

The momentum fractions of incoming partons can be written as functions of the masses of outgoing partons m_3 and m_4

$$x_a = \sqrt{\frac{\tau}{\zeta_3\zeta_4}}e^Y, \qquad x_b = \sqrt{\tau\zeta_3\zeta_4}e^{-Y}, \tag{A.1}$$

where

$$\zeta_3 = \frac{\hat{u} - m_3^2}{\hat{t} - m_3^2}, \qquad \zeta_4 = \frac{\hat{t} - m_4^2}{\hat{u} - m_4^2}. \tag{A.2}$$

We also calculate the energy and momentum of outgoing particles with nonzero mass. In the five subprocesses that we consider here the possible combinations of outgoing particles can both have zero mass (gluons), one massive and one zero mass (quark and gluon), or two massive particle of equal mass (quarks of the same flavour). The magnitude of transverse momentum of a particle depends on its mass. For two outgoing massless partons we have the definition in Equation 3.11. For two outgoing massive partons with $m_3 = m_4 = m$

$$p_T = \frac{1}{2\cosh y}\sqrt{M^2 - 4m^2\cosh^2 y}. \tag{A.3}$$

For one outgoing massless parton and one with nonzero mass m

$$\begin{aligned} p_T = \frac{1}{2}\Bigg(& \frac{\sqrt{2\cosh^2 2y(m^4\cosh 4y - m^4 + 2M^4)}}{\cosh^2 2y - 1} - \frac{2m^2\cosh^2 2y}{\cosh^2 2y - 1} \\ & + \frac{2m^2}{\cosh^2 2y - 1} - \frac{2M^2}{\cosh^2 2y - 1}\Bigg)^{\frac{1}{2}}. \end{aligned} \tag{A.4}$$

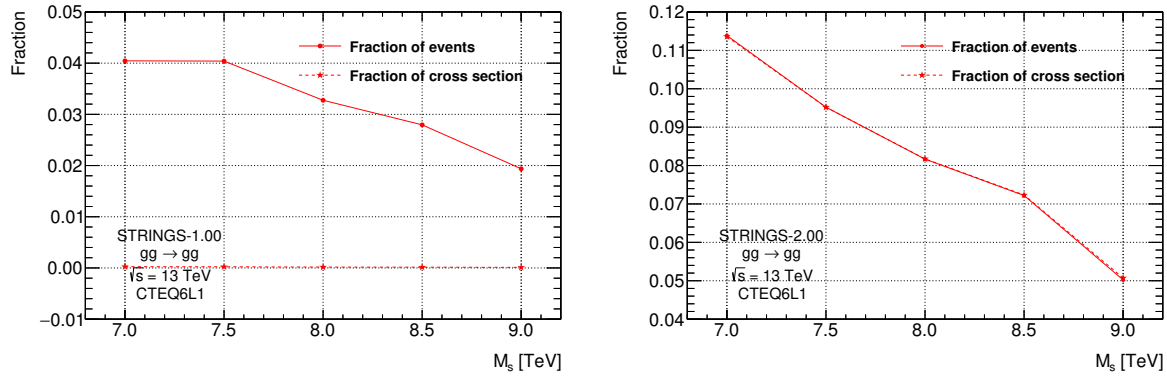
Outgoing partons have three-momentum as in Equation 3.13 and energies

$$E_3 = m_{T_3}\cosh y_3, \qquad E_4 = m_{T_4}\cosh y_4, \tag{A.5}$$

where $m_T = \sqrt{m^2 + p_T^2}$.

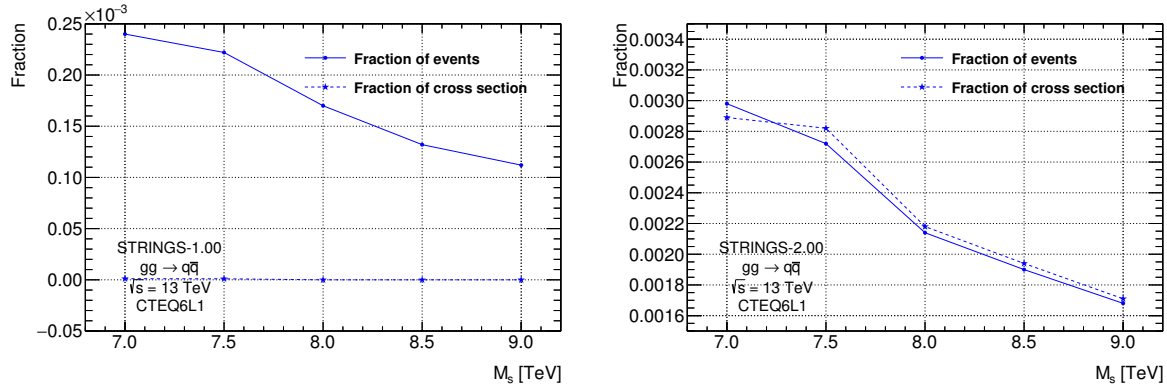
Appendix B STRINGS Process Fractions

The algorithm to calculate cross sections for each process in STRINGS-1.00 results in the fraction of each processes cross section to the total being different from its fraction of total events. This should not be the case, the fractional cross section should scale with the fraction of events since the cross section is a measure of the probability of a certain interaction. The method for calculating cross sections is modified in STRINGS-2.00 to correct this problem. The fraction of cross section and fraction of events for each process is shown below. These are calculated from a sample of 10000 events. The two fractions are very different when calculated with STRINGS-1.00 and are very similar with STRINGS-2.00. They are not exactly equal after correcting the algorithm due to rounding error but the difference between them decreases with more events generated.



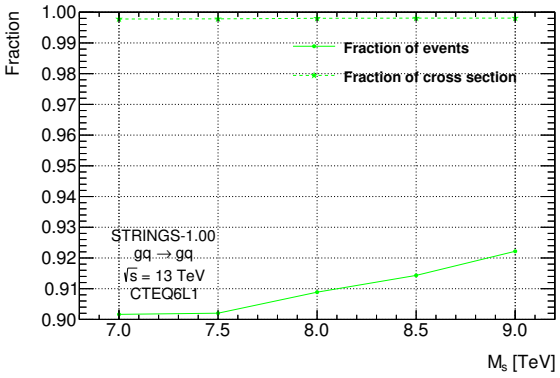
(a) STRINGS-1.00. (b) STRINGS-2.00.

Figure B.1: Fraction of total number of events (solid line) and total cross section (dashed line) for the $gg \to gg$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.

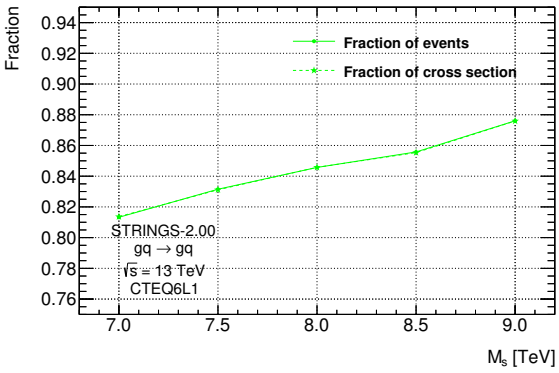


(a) STRINGS-1.00. (b) STRINGS-2.00.

Figure B.2: Fraction of total number of events (solid line) and total cross section (dashed line) for the $gg \to q\bar{q}$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.

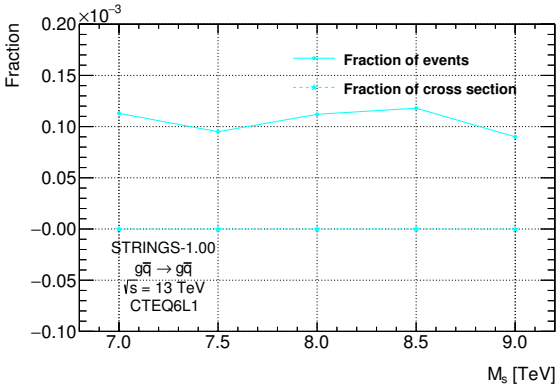


(a) STRINGS-1.00.

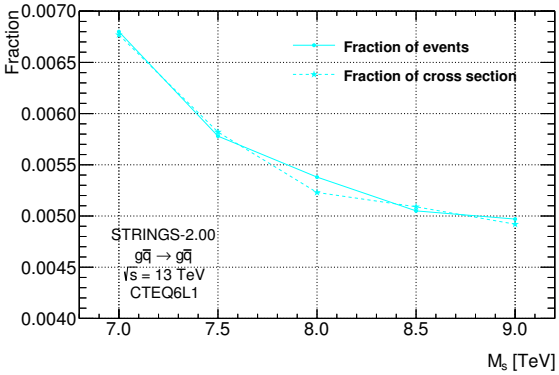


(b) STRINGS-2.00.

Figure B.3: Fraction of total number of events (solid line) and total cross section (dashed line) for the $gq \to gq$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.

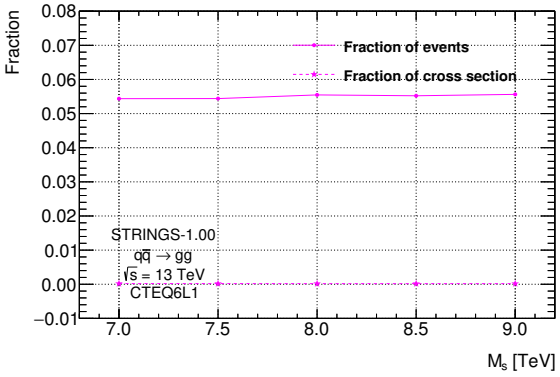


(a) STRINGS-1.00.

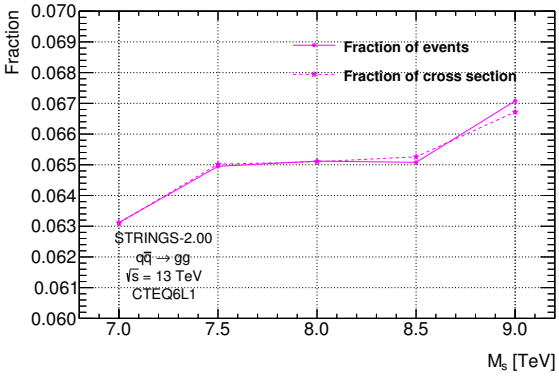


(b) STRINGS-2.00.

Figure B.4: Fraction of total number of events (solid line) and total cross section (dashed line) for the $g\bar{q} \to g\bar{q}$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.



(a) STRINGS-1.00.

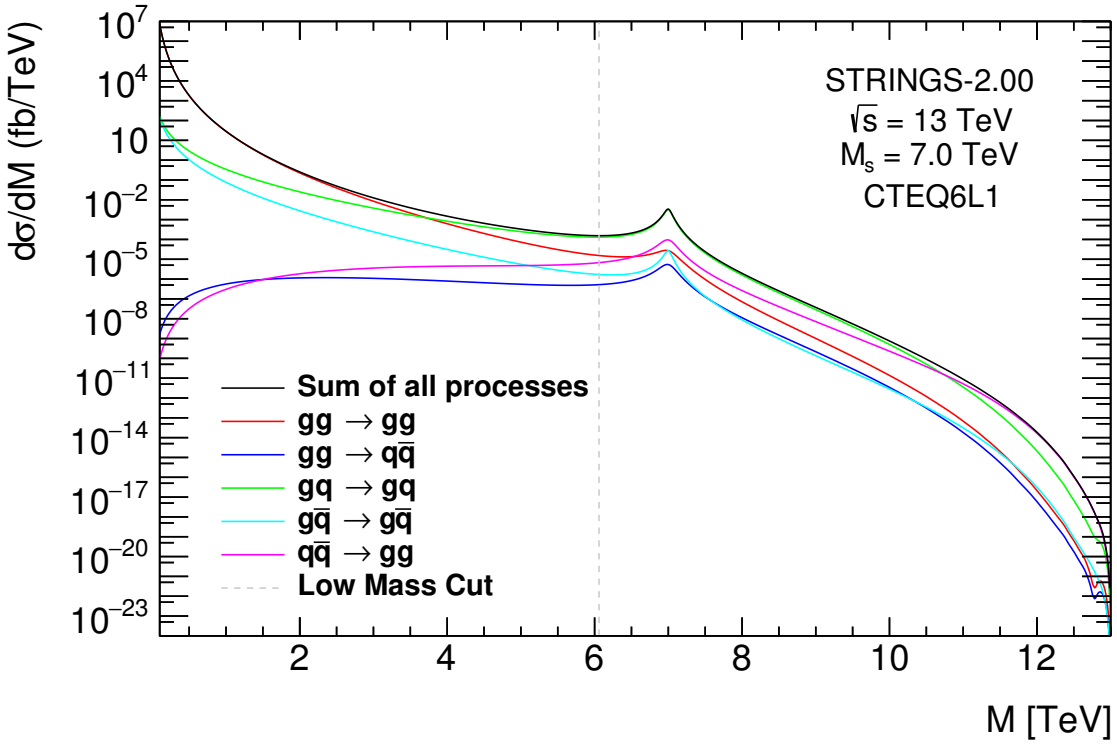


(b) STRINGS-2.00.

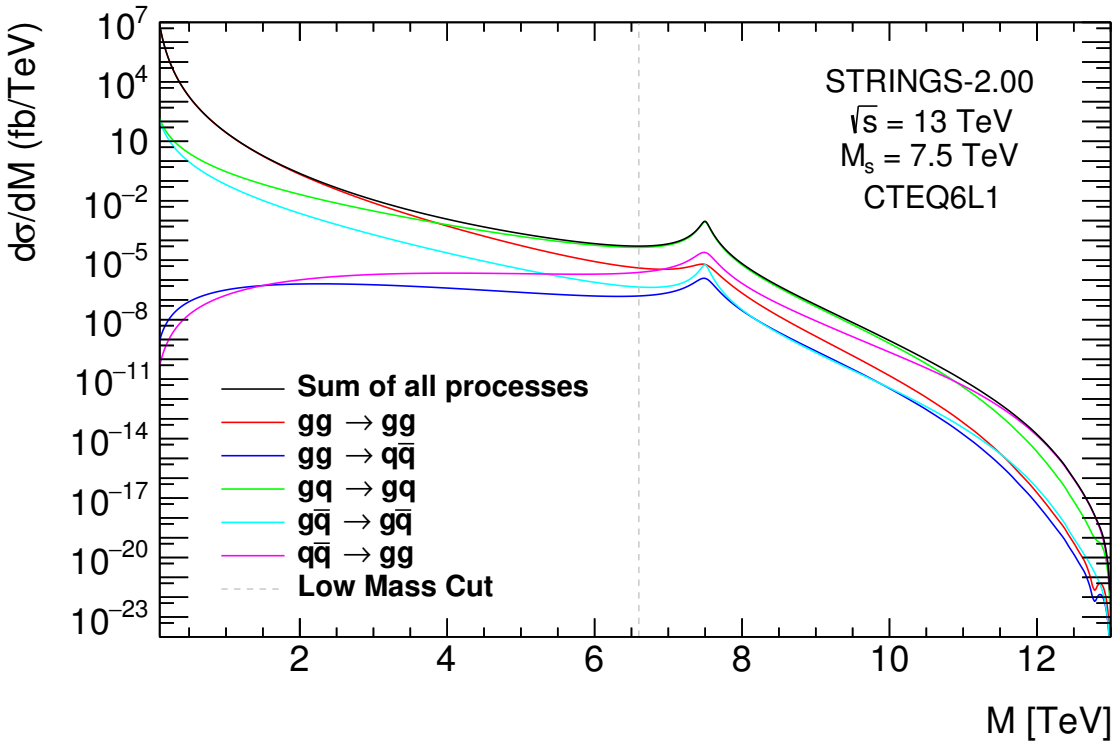
Figure B.5: Fraction of total number of events (solid line) and total cross section (dashed line) for the $q\bar{q} \to gg$ subprocess generated by STRINGS. A sample of 10000 events is generated with string scales $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.

Appendix C String Differential Cross Sections

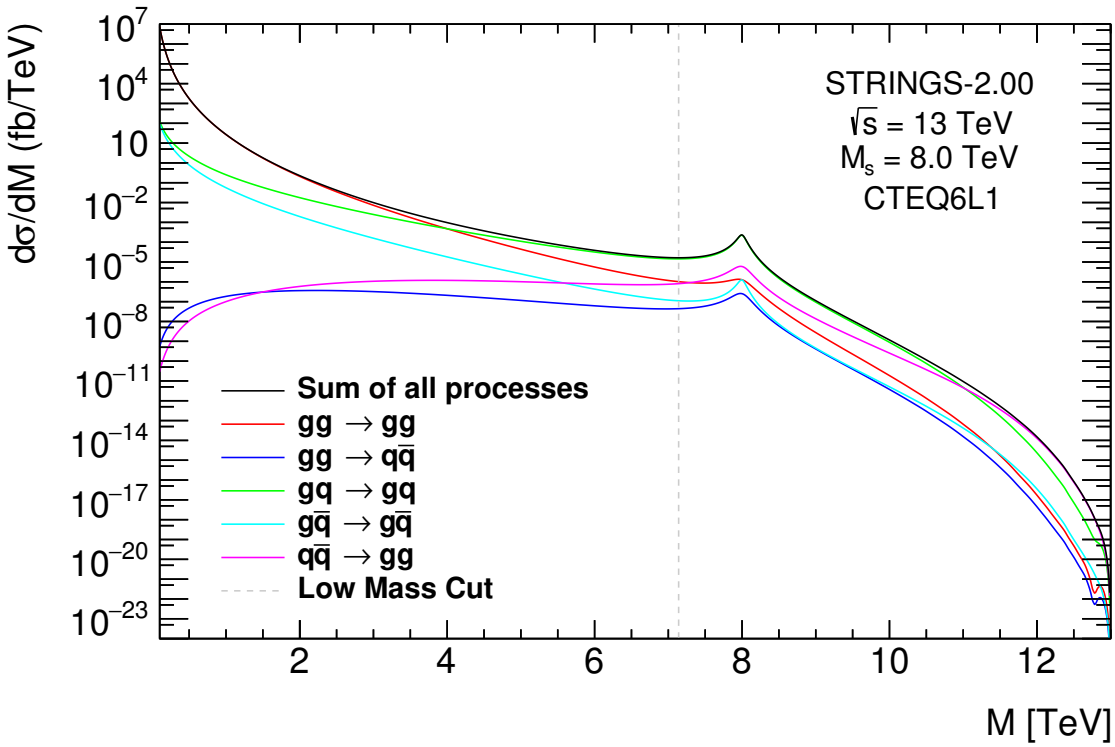
Differential cross sections as a function of M for various string scales are shown below. The entire m_{jj} spectrum is shown, but to reduce the contribution to the cross section of the low mass region samples are generated with $M_{Low\,cut} < M < 13$ TeV. The low mass cut for each sample is calculated using the larger of two possible ranges. The first possibility for the low mass bound is the M value corresponding to the lowest differential cross section to the left of the resonance peak. The second option is the M value such that the range $[M, 13]$ TeV covers 95% of the area under the curve in the range $[M_s - 1.0, 13]$ TeV. String samples with $M_s = 7.0$, 7.5, and 8.0 TeV use the first option while $M_s = 8.5$ and 9.0 use the second option.



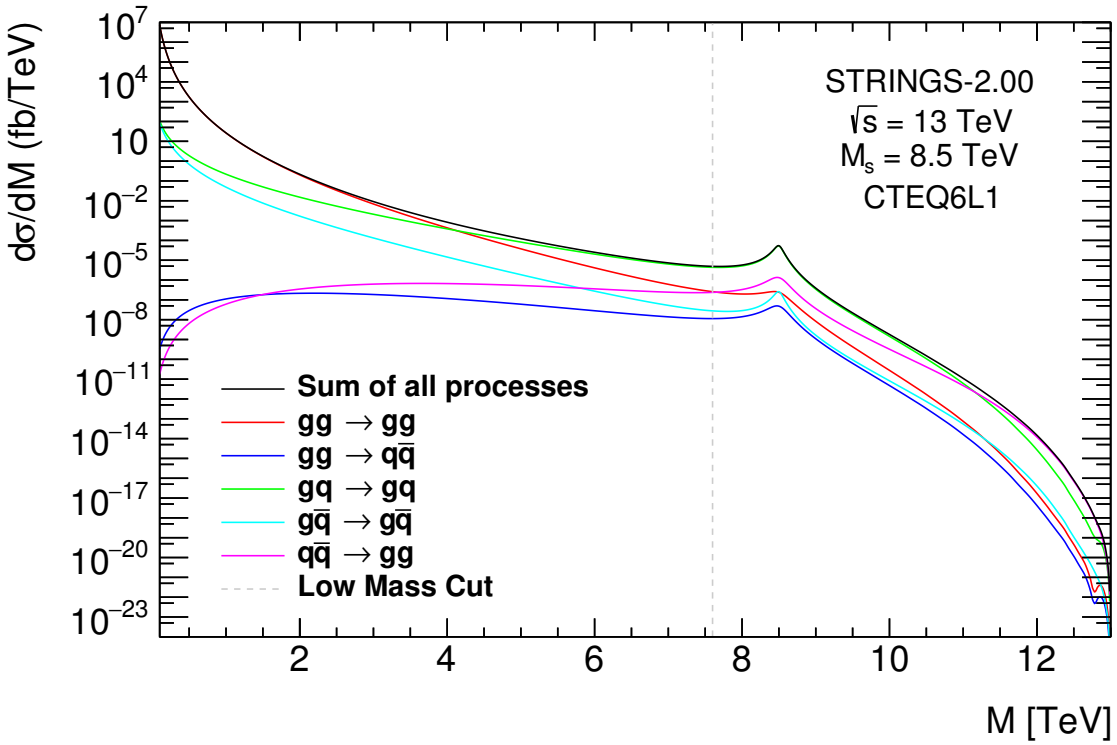
(a) $M_s = 7.0$ TeV, $M_{Low\,cut} = 6.06$ TeV.



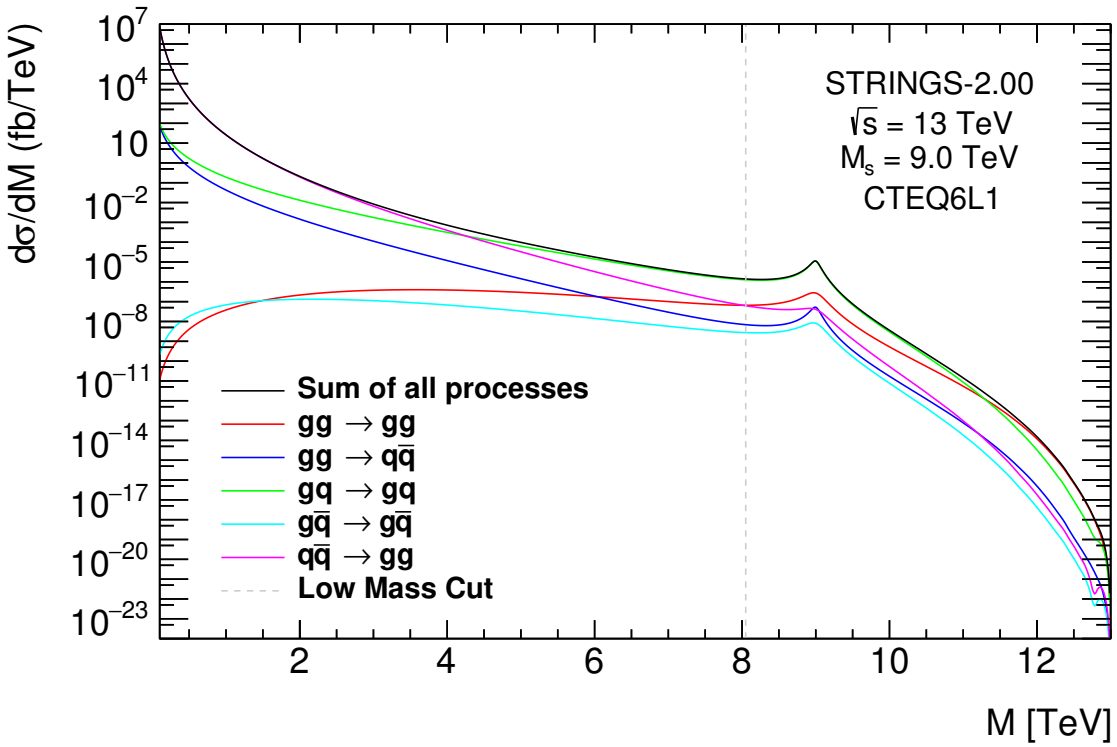
(b) $M_s = 7.5$ TeV, $M_{Low\,cut} = 6.60$ TeV.



(c) $M_s = 8.0$ TeV, $M_{Low\,cut} = 7.14$ TeV.



(d) $M_s = 8.5$ TeV, $M_{Low\,cut} = 7.60$ TeV.



(e) $M_s = 9.0$ TeV, $M_{Low\,cut} = 8.05$ TeV.

Figure C.1: Differential cross sections as a function of M for all five processes generated by STRINGS-2.00 using PDF set CETQ6L1 [14]. The black line is $M_{Low\,cut}$.

Appendix D Jet Reconstruction Parameters

The parameters used to reconstruct jets from topological clusters are listed below. These are applied to all samples used in this analysis including data and MC. The details of the jet reconstruction procedure are outlined in Section 4.4.

Jet reconstruction parameters	
Parameter	Value
Algorithm	anti-k_t
R-parameter	0.4
Input Constituent	EMTopo
Analysis Release Number	21.2.121
CalibArea tag	00-04-82
Calibration configuration	JES_MC16Recommendation_Consolidated_EMTopo_Apr2019_Rel21.config
Calibration sequence (Data)	JetArea_Residual_EtaJES_GSC_Insitu
Calibration sequence (MC)	JetArea_Residual_EtaJES_GSC_Smear
Calibration configuration (AFII)	JES_MC16Recommendation_AFII_EMTopo_Apr2019_Rel21.config
Calibration sequence (AFII)	JetArea_Residual_EtaJES_GSC_Smear

Table D.1: Jet reconstruction parameters used for data, QCD MC, and string samples.

Appendix E Data and MC Samples

The ntuples used to create the data, QCD MC, and string resonance distributions are listed below. These contain information on the kinematic variables of events that pass jet reconstruction and cleaning. Data samples are taken from the Good Run List, which includes only data recorded when all instruments in the detector were running properly. MC samples are generated in the MC16 campaign, divided into subcampaigns a, d, and e, that correspond to data15 and data16, data17, and data18, respectively. The QCD MC samples are divided into slices, and both QCD MC and string resonance samples must be weighted as described in Section 5.2.

Good Run List: Requirement that all relevant detectors were in a good state ready for physics. The GRLs used for this analysis are:

- 2015(3.2 fb^{-1}): data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown _PHYS_StandardGRL_All_Good_25ns.xml
- 2016(33 fb^{-1}): data16_13TeV.periodAllYear_DetStatus-v89-pro21-01 _DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
- 2017(44.2 fb^{-1}): data17_13TeV.periodAllYear_DetStatus-v99-pro22-01 _Unknown_PHYS_StandardGRL_All_Good_25ns_JetHLT_Normal2017.xml
- 2018(58.5 fb^{-1}): data18_13TeV.periodAllYear_DetStatus-v102-pro22-04 _Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

data15_13TeV.periodD.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016
data15_13TeV.periodE.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016
data15_13TeV.periodF.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016
data15_13TeV.periodG.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016
data15_13TeV.periodH.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016
data15_13TeV.periodJ.physics_Main.PhysCont.DAOD_EXOT2.grp15_v01_p4016

Table E.1: 2015 Data Samples.

data16_13TeV.periodA.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodB.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodC.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodD.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodE.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodF.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodG.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodI.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodK.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016
data16_13TeV.periodL.physics_Main.PhysCont.DAOD_EXOT2.grp16_v01_p4016

Table E.2: 2016 Data Samples.

data17_13TeV.periodB.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodC.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodD.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodE.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodF.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodH.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodI.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016
data17_13TeV.periodK.physics_Main.PhysCont.DAOD_EXOT2.grp17_v01_p4016

Table E.3: 2017 Data Samples.

data18_13TeV.periodB.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodC.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodD.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodF.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodI.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodK.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodL.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodM.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodO.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016
data18_13TeV.periodQ.physics_Main.PhysCont.DAOD_EXOT2.grp18_v01_p4016

Table E.4: 2018 Data Samples.

mc16_13TeV.364700.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364701.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364702.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364703.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364704.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364705.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364706.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364707.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364708.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364709.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364710.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364711.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364712.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006

Table E.5: MC16a Pythia QCD dijet samples.

mc16_13TeV.364700.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364701.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364702.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364703.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364704.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364705.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364706.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364707.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364708.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364709.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364710.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364711.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006
mc16_13TeV.364712.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10201_p4006

Table E.6: MC16d Pythia QCD dijet samples.

mc16_13TeV.364700.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364701.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364702.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364703.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364704.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364705.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364706.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364707.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364708.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364709.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364710.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364711.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006
mc16_13TeV.364712.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12WithSW.deriv.DAOD_EXOT2.e7142_s3126_r10724_p4006

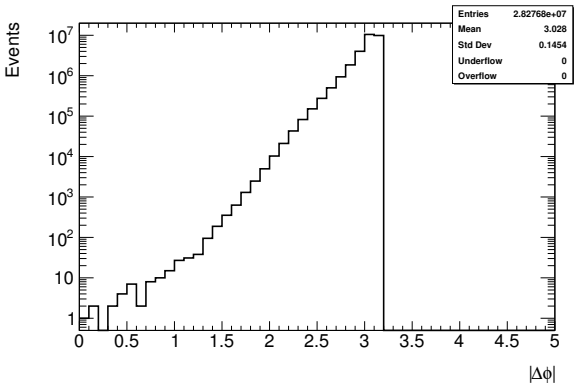
Table E.7: MC16e Pythia QCD dijet samples.

mc16_13TeV.312404.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07000.deriv.DAOD_EXOT2.e7655_a875_r9364_p3929/
mc16_13TeV.312404.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07000.deriv.DAOD_EXOT2.e7655_a875_r10201_p3929/
mc16_13TeV.312404.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07000.deriv.DAOD_EXOT2.e7655_a875_r10724_p3929/
mc16_13TeV.312405.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07500.deriv.DAOD_EXOT2.e7655_a875_r9364_p3929/
mc16_13TeV.312405.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07500.deriv.DAOD_EXOT2.e7655_a875_r10201_p3929/
mc16_13TeV.312405.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms07500.deriv.DAOD_EXOT2.e7655_a875_r10724_p3929/
mc16_13TeV.312406.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08000.deriv.DAOD_EXOT2.e7655_a875_r9364_p3929/
mc16_13TeV.312406.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08000.deriv.DAOD_EXOT2.e7655_a875_r10201_p3929/
mc16_13TeV.312406.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08000.deriv.DAOD_EXOT2.e7655_a875_r10724_p3929/
mc16_13TeV.312407.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08500.deriv.DAOD_EXOT2.e7655_a875_r9364_p3929/
mc16_13TeV.312407.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08500.deriv.DAOD_EXOT2.e7655_a875_r10201_p3929/
mc16_13TeV.312407.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms08500.deriv.DAOD_EXOT2.e7655_a875_r10724_p3929/
mc16_13TeV.312408.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms09000.deriv.DAOD_EXOT2.e7655_a875_r9364_p3929/
mc16_13TeV.312408.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms09000.deriv.DAOD_EXOT2.e7655_a875_r10201_p3929/
mc16_13TeV.312408.STRINGSPythia8EvtGen_A14CTEQ6L1_STR_Ms09000.deriv.DAOD_EXOT2.e7655_a875_r10724_p3929/

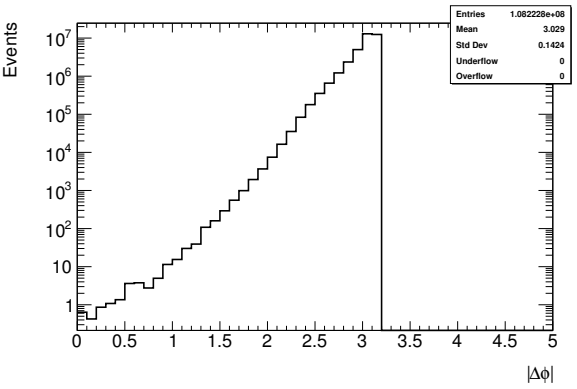
Table E.8: MC string resonance signal samples.

Appendix F Distribution of Event Selection Variables

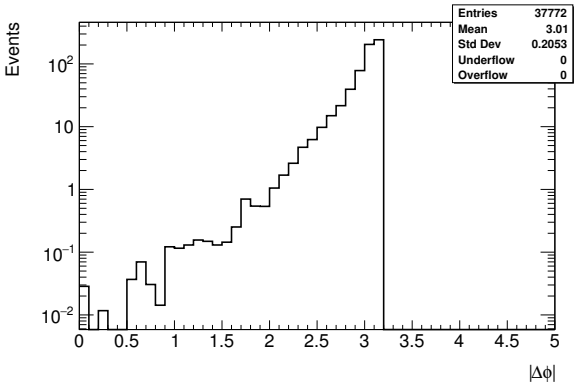
The distribution of kinematic variables used for event selection are shown below for data, QCD MC, and string resonance samples. Each variable in shown with all other selections made except for the criteria corresponding to the given variable. The QCD MC samples and string signal samples are weighted. The distributions are shown for $|\Delta\phi|$, y*, m_{jj} , and y.



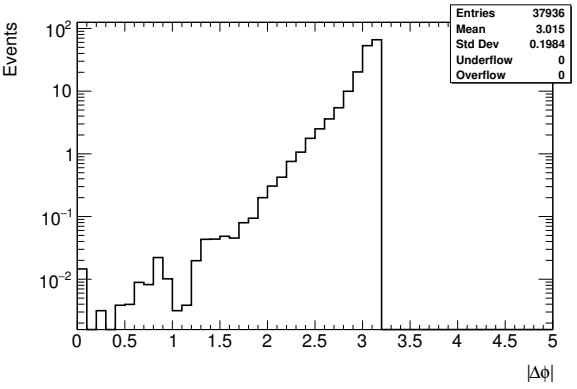
(a) Data.



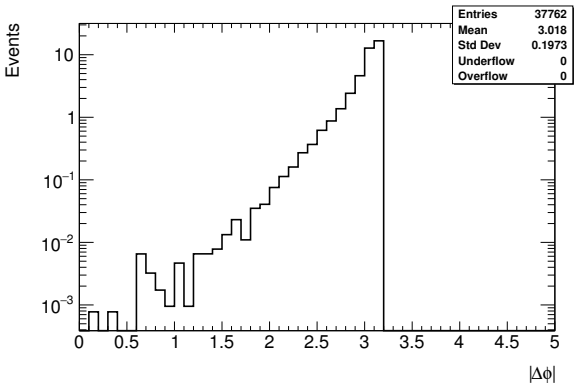
(b) QCD MC.



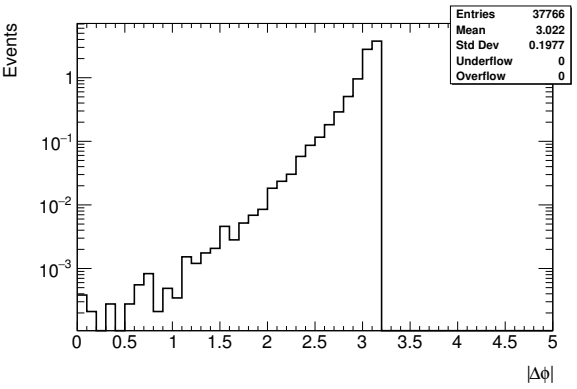
(c) Strings resonance, $M_s = 7.0$ TeV.



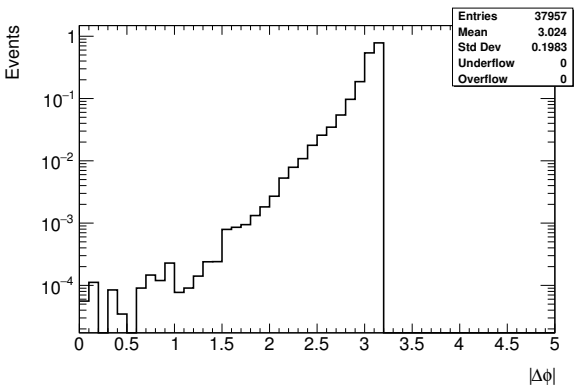
(d) Strings resonance, $M_s = 7.5$ TeV.



(e) Strings resonance, $M_s = 8.0$ TeV.

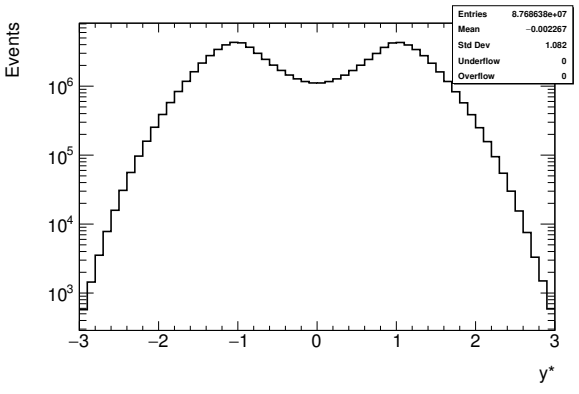


(f) Strings resonance, $M_s = 8.5$ TeV.

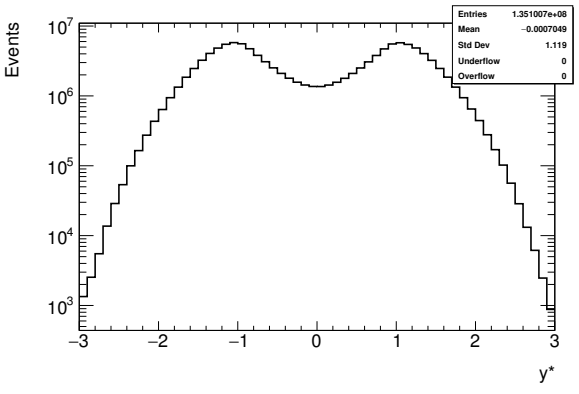


(g) Strings resonance, $M_s = 9.0$ TeV.

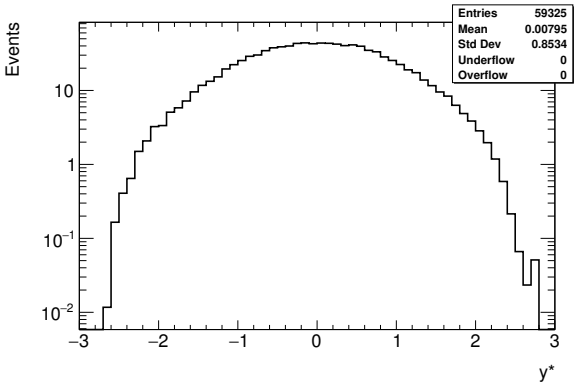
Figure F.1: $|\Delta\phi|$ for data, QCD MC, and string resonance with $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.



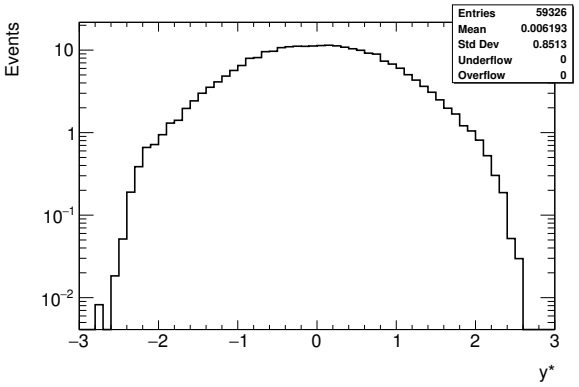
(a) Data.



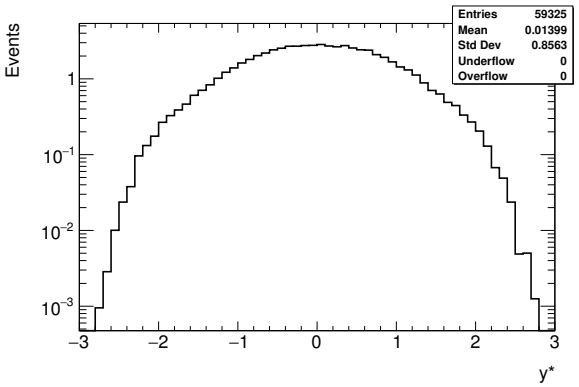
(b) Pythia MC QCD.



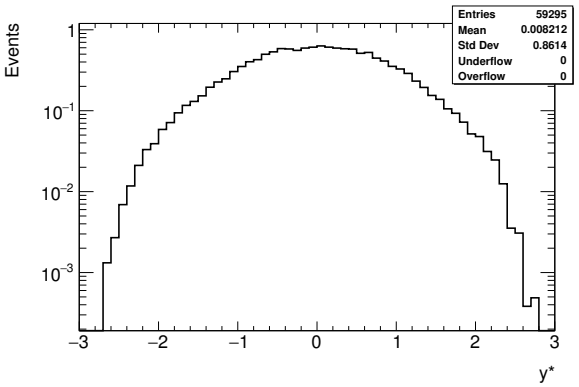
(c) Strings resonance, $M_s = 7.0$ TeV.



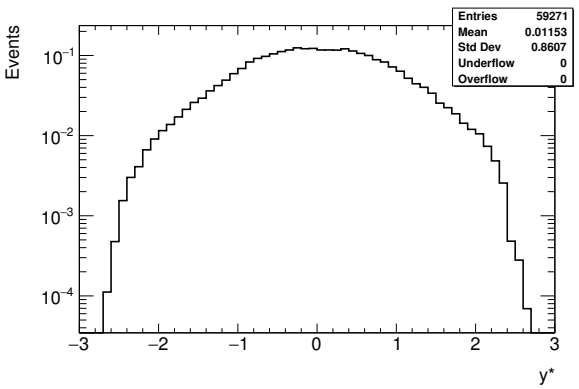
(d) Strings resonance, $M_s = 7.5$ TeV.



(e) Strings resonance, $M_s = 8.0$ TeV.

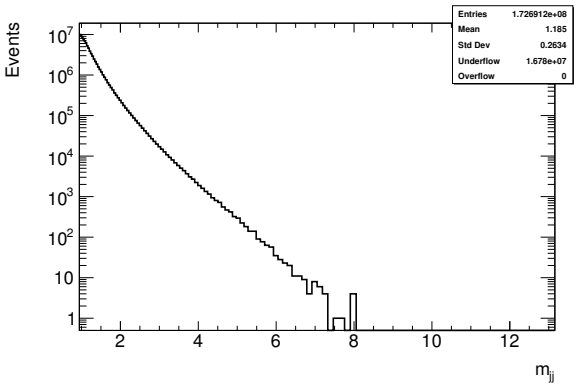


(f) Strings resonance, $M_s = 8.5$ TeV.

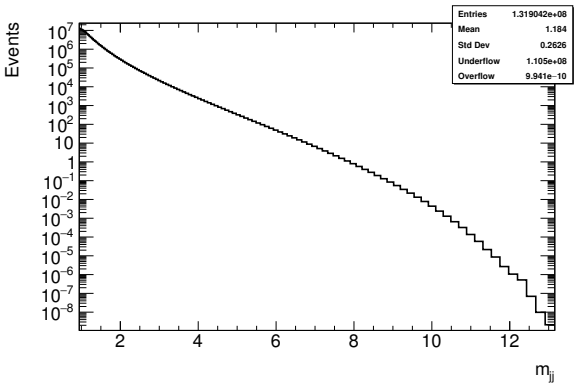


(g) Strings resonance, $M_s = 9.0$ TeV.

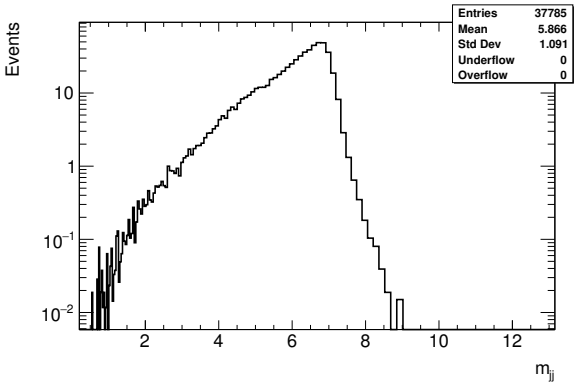
Figure F.2: y^* for data, QCD MC, and string resonance with $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.



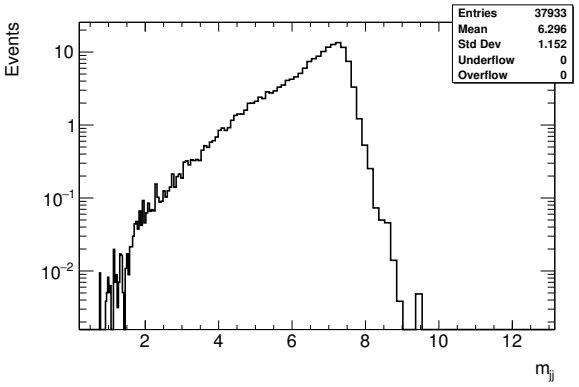
(a) Data.



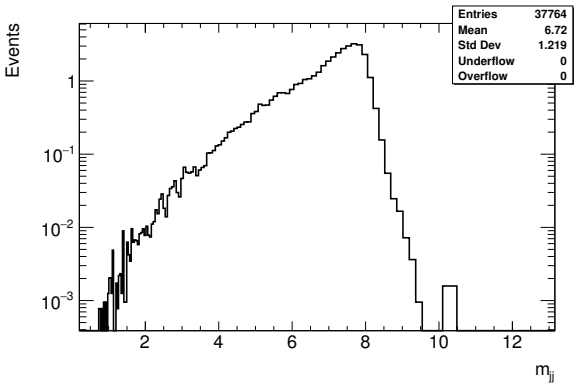
(b) Pythia MC QCD.



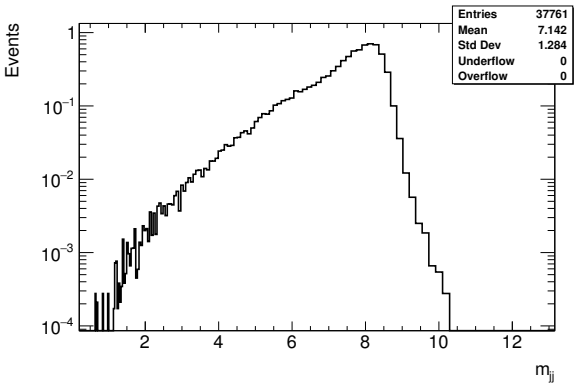
(c) Strings resonance, $M_s = 7.0$ TeV.



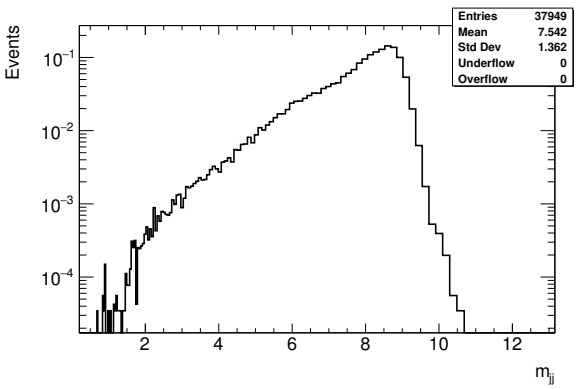
(d) Strings resonance, $M_s = 7.5$ TeV.



(e) Strings resonance, $M_s = 8.0$ TeV.

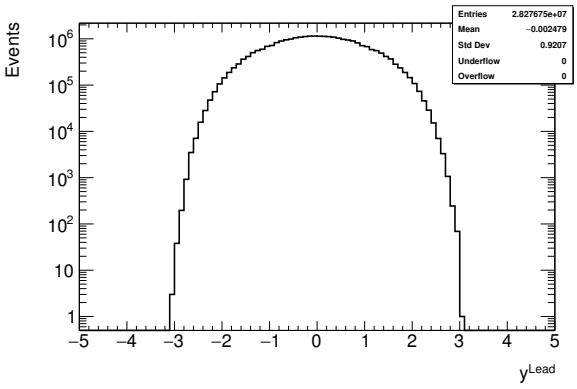


(f) Strings resonance, $M_s = 8.5$ TeV.

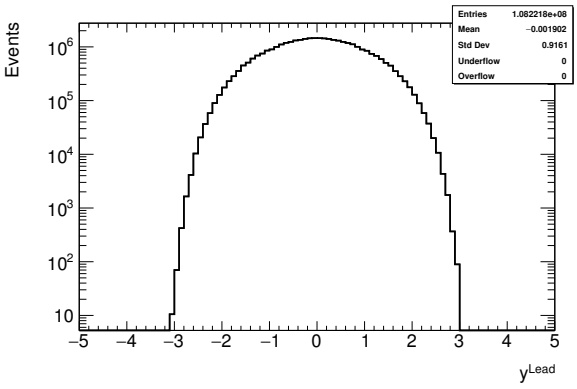


(g) Strings resonance, $M_s = 9.0$ TeV.

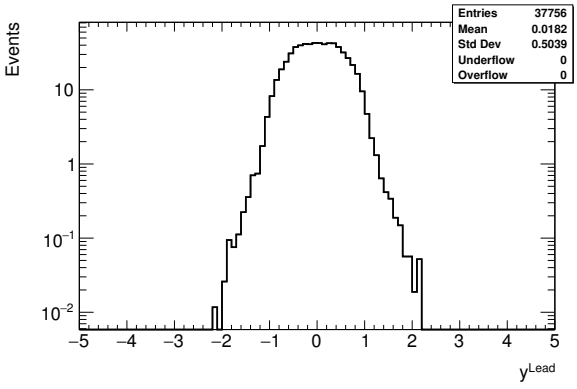
Figure F.3: m_{jj} for data, QCD MC, and string resonance with $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.



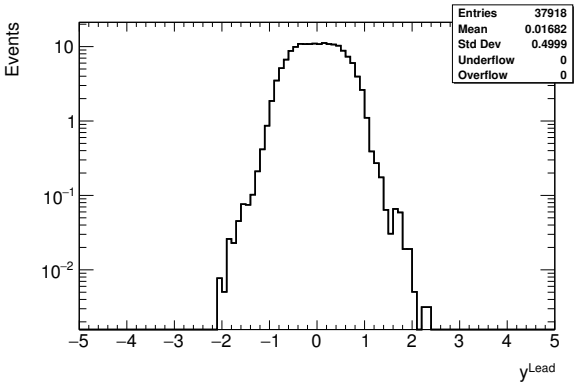
(a) Data.



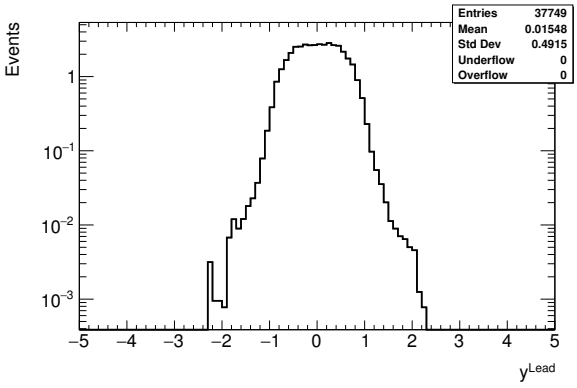
(b) Pythia MC QCD.



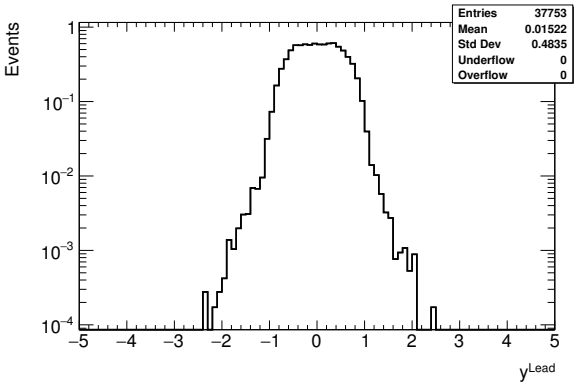
(c) Strings resonance, $M_s = 7.0$ TeV.



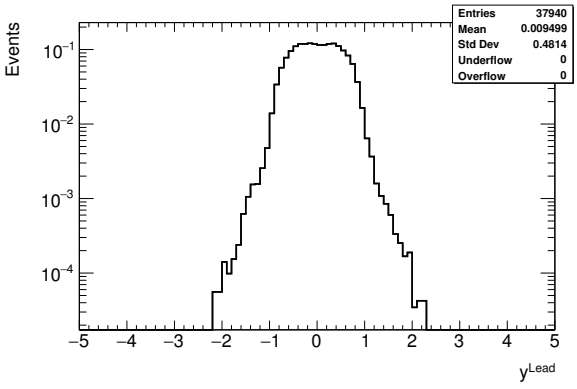
(d) Strings resonance, $M_s = 7.5$ TeV.



(e) Strings resonance, $M_s = 8.0$ TeV.

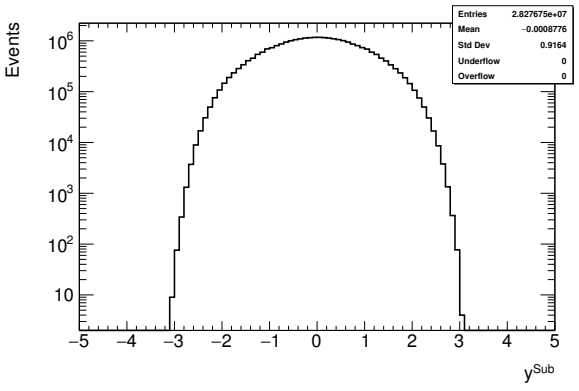


(f) Strings resonance, $M_s = 8.5$ TeV.

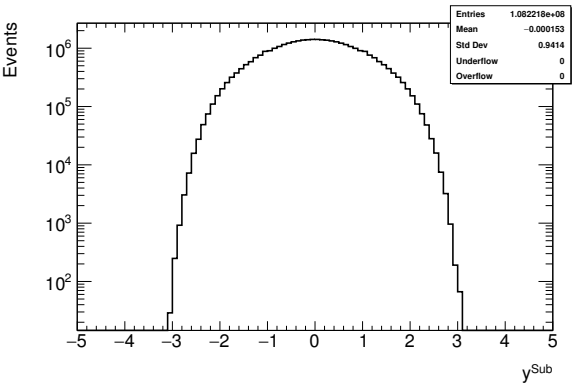


(g) Strings resonance, $M_s = 9.0$ TeV.

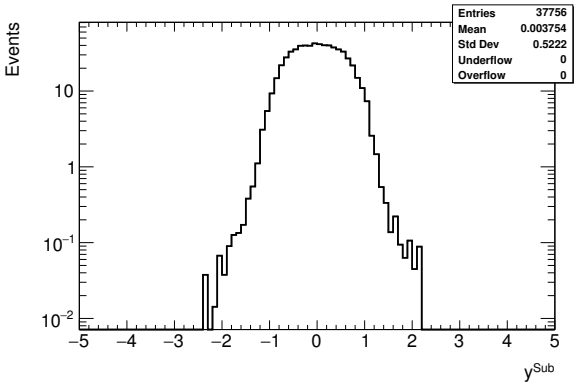
Figure F.4: Leading jet y for data, QCD MC, and string resonance with $M_s = 7.0$, 7.5, 8.0, 8.5, and 9.0 TeV.



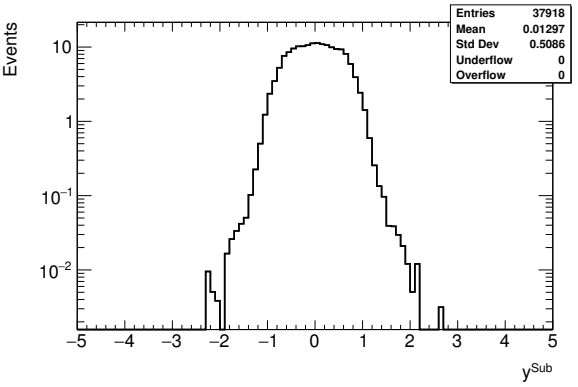
(a) Data.



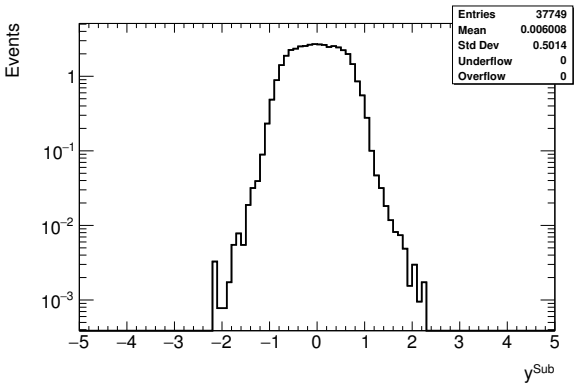
(b) Pythia MC QCD.



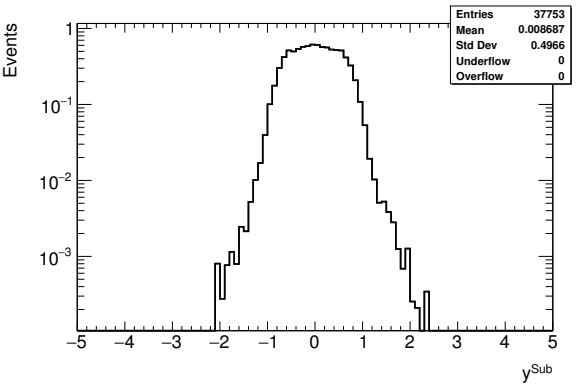
(c) Strings resonance, $M_s = 7.0$ TeV.



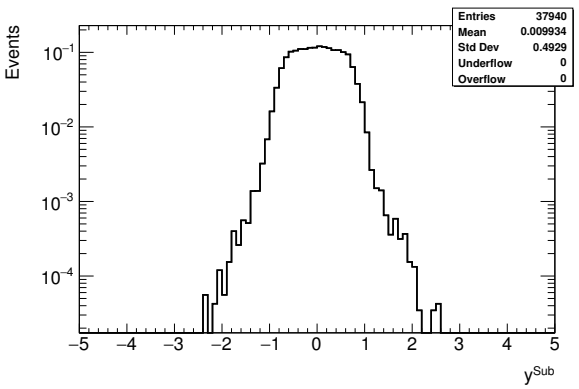
(d) Strings resonance, $M_s = 7.5$ TeV.



(e) Strings resonance, $M_s = 8.0$ TeV.



(f) Strings resonance, $M_s = 8.5$ TeV.



(g) Strings resonance, $M_s = 9.0$ TeV.

Figure F.5: Subleading jet y for data, QCD MC, and string resonance with $M_s =$ 7.0, 7.5, 8.0, 8.5, and 9.0 TeV.

Appendix G Event Selection

The selection criteria that are used for data samples are listed below. The number of events remaining after each selection has been made are shown for data, QCD MC (raw and weighted), and string resonance (raw and weighted) samples. The selection criteria used are:

- LAr: Liquid Argon Calorimeter error rejected (errorState(xAOD::EventInfo::LAr))
- Tile: Tile Calorimeter error rejected (errorState(xAOD::EventInfo::Tile))
- SCT: SCT single event upsets rejected (errorState(xAOD::EventInfo::SCT))
- Core: Incomplete event build rejected (isEventFlagBitSet(xAOD::EventInfo::Core, 18))
- Jet Cleaning: All jets with $p_T > 150$ GeV pass LooseBad cleaning cuts
- NPV: the highest $\sum {p_T}^2$(trk) (xAOD::VxType::VertexType::PriVtx) vertex has at least two tracks associated with it
- Trigger: Passes one of the single-jet triggers, HLT_j360 or HLT_j380 or HLT_j400 or HLT_j420 or HLT_j440
- Jet Selection: Leading jet $p_T > 380$ GeV, subleading jet $p_T > 150$ GeV, and Jet multiplicity ≥ 2
- Trigger: HLT_j420
- $|\Delta\phi| > 1$
- $|y^*| < 0.8$
- $m_{jj} > 1.2$ TeV

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	326660326		
GRL	318820170	97.60	97.60
LAr	318725744	99.97	97.57
Tile	318725744	100.00	97.57
SCT	318702682	99.99	97.56
Core	318702682	100.00	97.56
Jet Cleaning	315244974	98.92	96.51
NPV	315237388	100.00	96.50
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	12973559	4.12	3.97
JetSelect	12642978	97.45	3.87
Trigger HLT_j420	5748609	45.47	1.76
$\|\Delta\phi\| > 1$	5746945	99.97	1.76
$\|y^*\| < 0.8$	3932395	68.43	1.20
$m_{jj} > 1.2$ TeV	685898	17.44	0.21

Table G.1: Cutflow of Data 2015.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	1316851559		
GRL	1287179227	97.75	97.75
LAr	1285484431	99.87	97.62
Tile	1285483788	100.00	97.62
SCT	1285415377	99.99	97.61
Core	1285415377	100.00	97.61
Jet Cleaning	1272097086	98.96	96.60
NPV	1272096188	100.00	96.60
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	105647542	8.30	8.02
JetSelect	104058580	98.50	7.90
Trigger HLT_j420	58917913	56.62	4.47
$\|\Delta\phi\| > 1$	58902324	99.97	4.47
$\|y^*\| < 0.8$	40364691	68.53	3.07
$m_{jj} > 1.2$ TeV	7071879	17.52	0.54

Table G.2: Cutflow of Data 2016.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	1427254378		
GRL	1373686458	96.25	96.25
LAr	1372286418	99.90	96.15
Tile	1372240640	100.00	96.15
SCT	1372143768	99.99	96.14
Core	1372143768	100.00	96.14
Jet Cleaning	1352722412	98.58	94.78
NPV	1352722313	100.00	94.78
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	124537730	9.21	8.73
JetSelect	120705428	96.92	8.46
Trigger HLT_j420	86572256	71.72	6.07
$\|\Delta\phi\| > 1$	86548827	99.97	6.06
$\|y^*\| < 0.8$	59778764	69.07	4.19
$m_{jj} > 1.2$ TeV	9513250	15.91	0.67

Table G.3: Cutflow of Data 2017.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	1332899485		
GRL	1312042568	98.44	98.44
LAr	1311194122	99.94	98.37
Tile	1311194121	100.00	98.37
SCT	1311128315	99.99	98.37
Core	1311128315	100.00	98.37
Jet Cleaning	1290746318	98.45	96.84
NPV	1290746029	100.00	96.84
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	127739031	9.90	9.58
JetSelect	126472646	99.01	9.49
Trigger HLT_j420	99029518	78.30	7.43
$\|\Delta\phi\| > 1$	99003749	99.97	7.43
$\|y^*\| < 0.8$	68615359	69.31	5.15
$m_{jj} > 1.2$ TeV	11005721	16.04	0.83

Table G.4: Cutflow of Data 2018.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	4403665748		
GRL	4291728423	97.46	97.46
LAr	4287690715	99.91	97.37
Tile	4287644293	100.00	97.37
SCT	4287390142	100.00	97.36
Core	4287390142	100.00	97.36
Jet Cleaning	4230810790	98.68	96.10
NPV	4230801918	100.00	96.10
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	370897862	8.77	8.42
JetSelect	363879632	98.11	8.26
Trigger HLT_j420	250268296	68.78	5.68
$\|\Delta\phi\| > 1$	250201845	99.97	5.68
$\|y^*\| < 0.8$	172691209	69.02	3.92
$m_{jj} > 1.2$ TeV	28276748	16.37	0.64

Table G.5: Cutflow of full Run 2 Data (2015-2018).

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	59315550		
Jet Cleaning	59156575	99.73	99.73
NPV	59156575	100.00	99.73
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	43934795	74.27	74.07
JetSelect	43891188	99.90	74.00
MC Cleaning	43890581	100.00	74.00
Trigger HLT_j420	40573706	92.44	68.40
$\|\Delta\phi\| > 1$	40563151	99.97	68.39
$\|y^*\| < 0.8$	33316909	82.14	56.17
$m_{jj} > 1.2$ TeV	27291783	81.92	46.01

Table G.6: Cutflow of Pythia QCD MC16a, raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	74336300		
Jet Cleaning	74019931	99.57	99.57
NPV	74019931	100.00	99.57
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	55103433	74.44	74.13
JetSelect	55046982	99.90	74.05
MC Cleaning	55046226	100.00	74.05
Trigger HLT_j420	50990364	92.63	68.59
$\|\Delta\phi\| > 1$	50977274	99.97	68.58
$\|y^*\| < 0.8$	41826697	82.05	56.27
$m_{jj} > 1.2$ TeV	34150298	81.65	45.94

Table G.7: Cutflow of Pythia QCD MC16d, raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	99461184		
Jet Cleaning	99054259	99.59	99.59
NPV	99054258	100.00	99.59
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	74446383	75.16	74.85
JetSelect	74372709	99.90	74.78
MC Cleaning	74371630	100.00	74.77
Trigger HLT_j420	69089849	92.90	69.46
$\|\Delta\phi\| > 1$	69072288	99.97	69.45
$\|y^*\| < 0.8$	56760612	82.18	57.07
$m_{jj} > 1.2$ TeV	46779744	82.42	47.03

Table G.8: Cutflow of Pythia QCD MC16e, raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	233113034		
Jet Cleaning	232230765	99.62	99.62
NPV	232230764	100.00	99.62
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	173484611	74.70	74.42
JetSelect	173310879	99.90	74.35
MC Cleaning	173308437	100.00	74.35
Trigger HLT_j420	160653919	92.70	68.92
$\|\Delta\phi\| > 1$	160612713	99.97	68.90
$\|y^*\| < 0.8$	131904218	82.13	56.58
$m_{jj} > 1.2$ TeV	108221825	82.05	46.42

Table G.9: Cutflow of Pythia QCD MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	84942397		
$\|\Delta\phi\| > 1$	84923903	99.98	99.98
$\|y^*\| < 0.8$	56986192	67.10	67.09
$m_{jj} > 1.2$ TeV	9226827	16.19	10.86

Table G.10: Cutflow of Pythia QCD MC16a, weighted events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	105310345		
$\|\Delta\phi\| > 1$	105287974	99.98	99.98
$\|y^*\| < 0.8$	54725401	51.98	51.97
$m_{jj} > 1.2$ TeV	10808280	19.75	10.26

Table G.11: Cutflow of Pythia QCD MC16d, weighted events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	138772257		
$\|\Delta\phi\| > 1$	138742146	99.98	99.98
$\|y^*\| < 0.8$	92531998	66.69	66.68
$m_{jj} > 1.2$ TeV	14826236	16.02	10.68

Table G.12: Cutflow of Pythia QCD MC16e, weighted events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	329024999		
$\|\Delta\phi\| > 1$	328954023	99.98	99.98
$\|y^*\| < 0.8$	204243591	62.09	62.08
$m_{jj} > 1.2$ TeV	34861343	17.07	10.60

Table G.13: Cutflow of Pythia QCD MC16 (a+d+e), weighted events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	60000		
Jet Cleaning	59420	99.03	99.03
NPV	59420	100.00	99.03
Trigger $HLT\_j360\|j380\|j400\|j420\|j440$	59420	100.00	99.03
JetSelect	59420	100.00	99.03
Trigger HLT_j420	59410	99.98	99.02
$\|\Delta\phi\| > 1$	59354	99.91	98.92
$\|y^*\| < 0.8$	37785	63.66	62.98
$m_{jj} > 1.2$ TeV	37756	99.92	62.93

Table G.14: Cutflow of string 7.0 TeV MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	60000		
Jet Cleaning	59398	99.00	99.00
NPV	59398	100.00	99.00
Trigger $HLT\_j360\|j380\|j400\|j420\|j440$	59398	100.00	99.00
JetSelect	59398	100.00	99.00
Trigger HLT_j420	59392	99.99	98.99
$\|\Delta\phi\| > 1$	59342	99.92	98.90
$\|y^*\| < 0.8$	37933	63.92	63.22
$m_{jj} > 1.2$ TeV	37918	99.96	63.20

Table G.15: Cutflow of string 7.5 TeV MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	60000		
Jet Cleaning	59382	98.97	98.97
NPV	59382	100.00	98.97
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	59382	100.00	98.97
JetSelect	59382	100.00	98.97
Trigger HLT_j420	59380	100.00	98.97
$\|\Delta\phi\| > 1$	59340	99.93	98.90
$\|y^*\| < 0.8$	37764	63.64	62.94
$m_{jj} > 1.2$ TeV	37749	99.96	62.92

Table G.16: Cutflow of string 8.0 TeV MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	60000		
Jet Cleaning	59361	98.94	98.94
NPV	59361	100.00	98.94
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	59361	100.00	98.94
JetSelect	59361	100.00	98.94
Trigger HLT_j420	59360	100.00	98.93
$\|\Delta\phi\| > 1$	59303	99.90	98.84
$\|y^*\| < 0.8$	37761	63.67	62.94
$m_{jj} > 1.2$ TeV	37753	99.98	62.92

Table G.17: Cutflow of string 8.5 TeV MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
All	60000		
Jet Cleaning	59324	98.87	98.87
NPV	59324	100.00	98.87
Trigger $HLT\_j360 \| j380 \| j400 \| j420 \| j440$	59324	100.00	98.87
JetSelect	59324	100.00	98.87
Trigger HLT_j420	59323	100.00	98.87
$\|\Delta\phi\| > 1$	59280	99.93	98.80
$\|y^*\| < 0.8$	37949	64.02	63.25
$m_{jj} > 1.2$ TeV	37940	99.96	63.23

Table G.18: Cutflow of string 9.0 TeV MC16 (a+d+e), raw MC Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	984.6980		
$\|\Delta\phi\| > 1$	984.0963	99.94	99.94
$\|y^*\| < 0.8$	630.5890	64.08	64.04
$m_{jj} > 1.2$ TeV	630.2987	99.95	64.01

Table G.19: Cutflow of string 7.0 TeV MC16 (a+d+e), weighted Events.

Selection criteria	N_{events}	Relative percentage	Cummulative percentage
Trigger HLT_j420	258.3895		
$\|\Delta\phi\| > 1$	258.2474	99.95	99.95
$\|y^*\| < 0.8$	165.9706	64.27	64.23
$m_{jj} > 1.2$ TeV	165.9192	99.97	64.21

Table G.20: Cutflow of string 7.5 TeV MC16 (a+d+e), weighted Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	63.3469		
$\|\Delta\phi\| > 1$	63.3196	99.96	99.96
$\|y^*\| < 0.8$	40.4869	63.94	63.91
$m_{jj} > 1.2$ TeV	40.4754	99.97	63.89

Table G.21: Cutflow of string 8.0 TeV MC16 (a+d+e), weighted Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	13.8899		
$\|\Delta\phi\| > 1$	13.8819	99.94	99.94
$\|y^*\| < 0.8$	8.8673	63.88	63.84
$m_{jj} > 1.2$ TeV	8.8662	99.99	63.83

Table G.22: Cutflow of string 8.5 TeV MC16 (a+d+e), weighted Events.

Selection criteria	N_{events}	Relative percentage	Cumulative percentage
Trigger HLT_j420	2.7646		
$\|\Delta\phi\| > 1$	2.7632	99.95	99.95
$\|y^*\| < 0.8$	1.7728	64.16	64.13
$m_{jj} > 1.2$ TeV	1.7727	99.99	64.12

Table G.23: Cutflow of string 9.0 TeV MC16 (a+d+e), weighted Events.

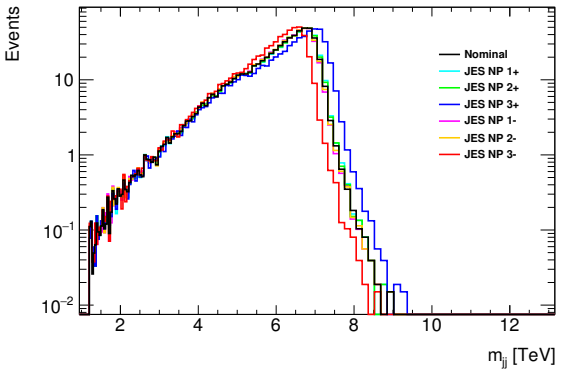
Appendix H m_{jj} Binning

The bins used to create m_{jj} histograms for data, background, and signal samples are shown below. The same binning has been used in previous ATLAS dijet analyses [57, 64] and follows the m_{jj} distribution with wider bins at higher energies.

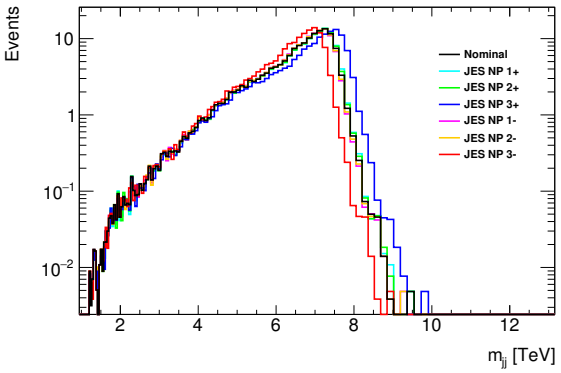
[0.946, 0.976, 1.006, 1.037, 1.068, 1.100, 1.133, 1.166, 1.200, 1.234, 1.269, 1.305, 1.341, 1.378, 1.416, 1.454, 1.493, 1.533, 1.573, 1.614, 1.656, 1.698, 1.741, 1.785, 1.830, 1.875, 1.921, 1.968, 2.016, 2.065, 2.114, 2.164, 2.215, 2.267, 2.320, 2.374, 2.429, 2.485, 2.542, 2.600, 2.659, 2.719, 2.780, 2.842, 2.905, 2.969, 3.034, 3.100, 3.167, 3.235, 3.305, 3.376, 3.448, 3.521, 3.596, 3.672, 3.749, 3.827, 3.907, 3.988, 4.070, 4.154, 4.239, 4.326, 4.414, 4.504, 4.595, 4.688, 4.782, 4.878, 4.975, 5.074, 5.175, 5.277, 5.381, 5.487, 5.595, 5.705, 5.817, 5.931, 6.047, 6.165, 6.285, 6.407, 6.531, 6.658, 6.787, 6.918, 7.052, 7.188, 7.326, 7.467, 7.610, 7.756, 7.904, 8.055, 8.208, 8.364, 8.523, 8.685, 8.850, 9.019, 9.191, 9.366, 9.544, 9.726, 9.911, 10.100, 10.292, 10.488, 10.688, 10.892, 11.100, 11.312, 11.528, 11.748, 11.972, 12.200, 12.432, 12.669, 12.910, 13.156] TeV

Appendix I Signal JES and JER Distributions

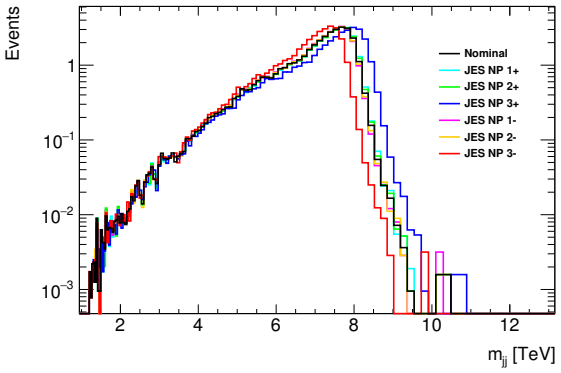
The shift in raw events for each string resonance sample caused by JES and JER are shown below. The JES systematic uncertainty is treated as two histograms, one is the nominal distribution with the mean m_{jj} shifted higher and one with the mean m_{jj} shifted lower. The JER uncertainty is also treated as two histograms, one is the nominal distribution with m_{jj} shifted wider and one with m_{jj} shifted narrower.



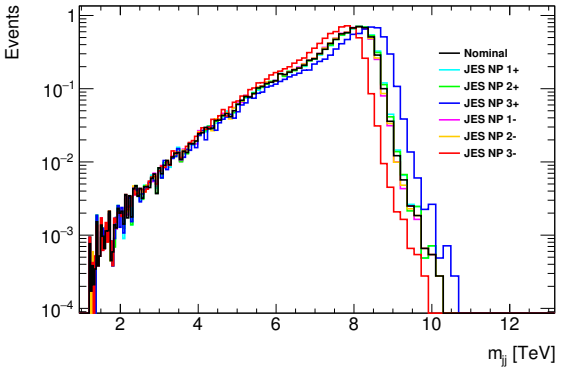
(a) $M_s = 7.0$ TeV.



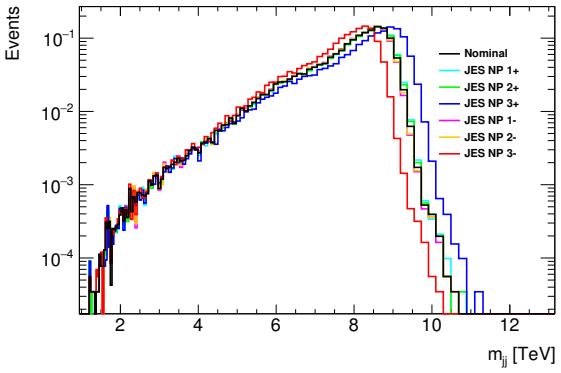
(b) $M_s = 7.5$ TeV.



(c) $M_s = 8.0$ TeV.

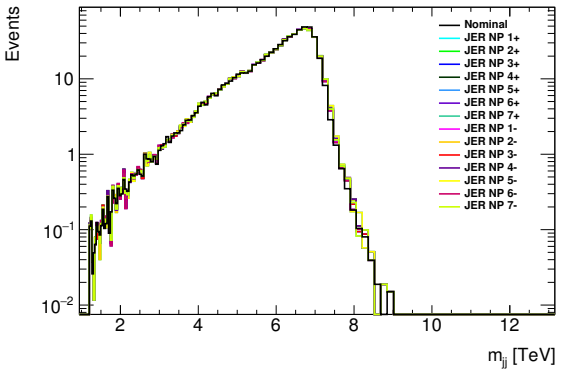


(d) $M_s = 8.5$ TeV.

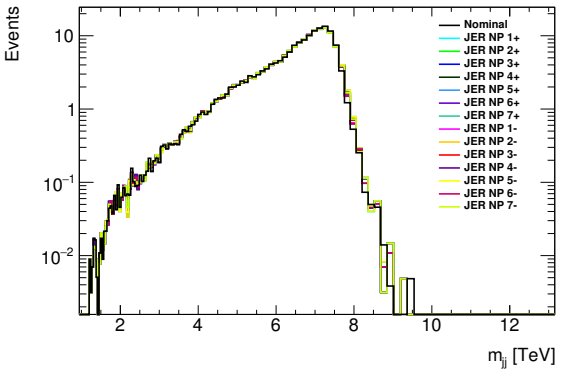


(e) $M_s = 9.0$ TeV.

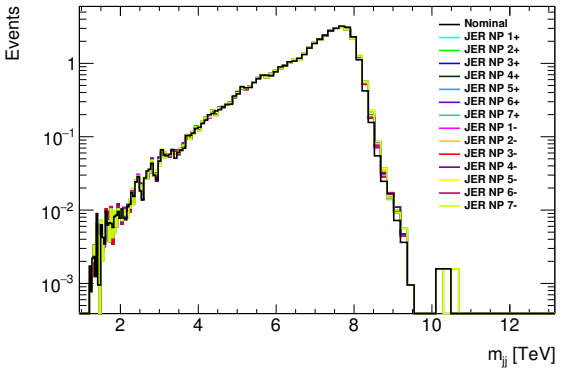
Figure I.1: Nominal m_{jj} distribution for string signal samples with ± JES uncertainty.



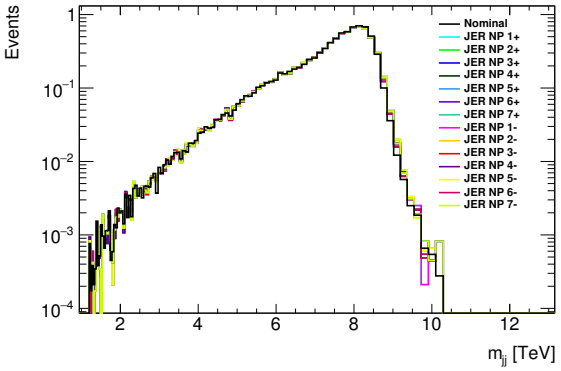
(a) $M_s = 7.0$ TeV.



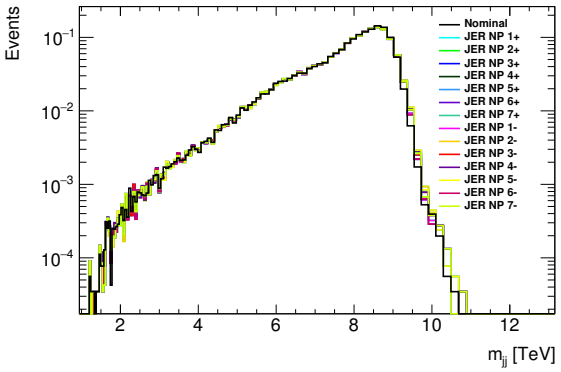
(b) $M_s = 7.5$ TeV.



(c) $M_s = 8.0$ TeV.



(d) $M_s = 8.5$ TeV.

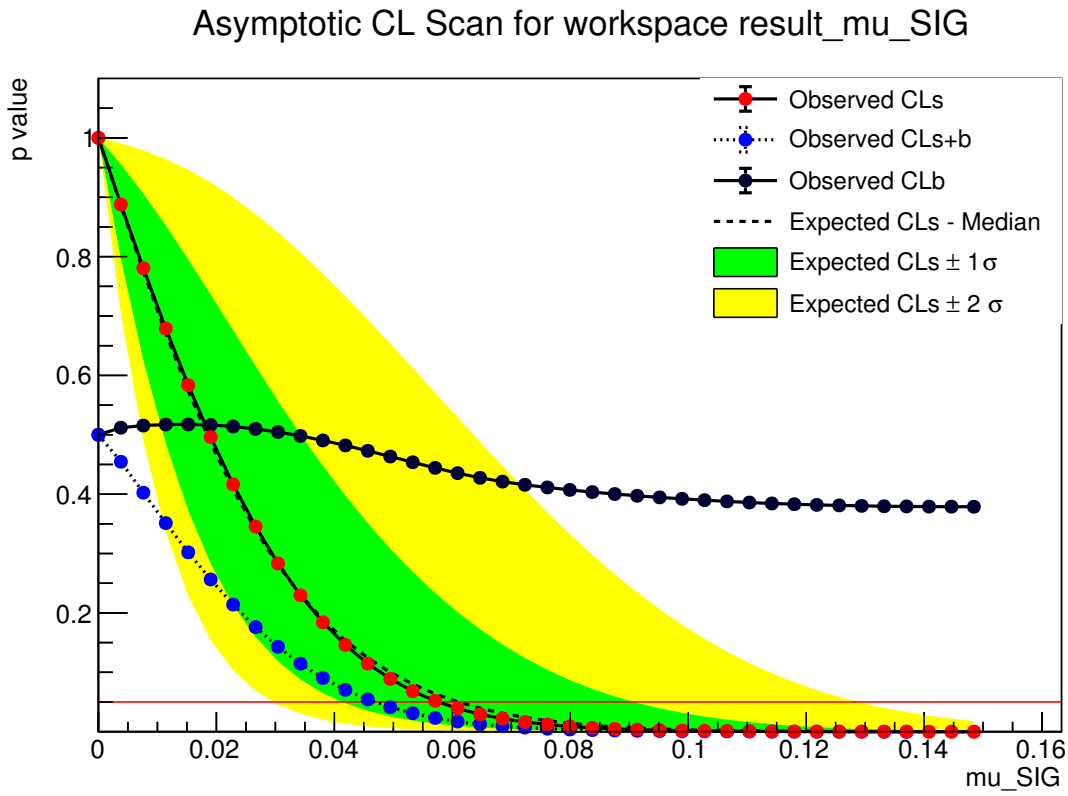


(e) $M_s = 9.0$ TeV.

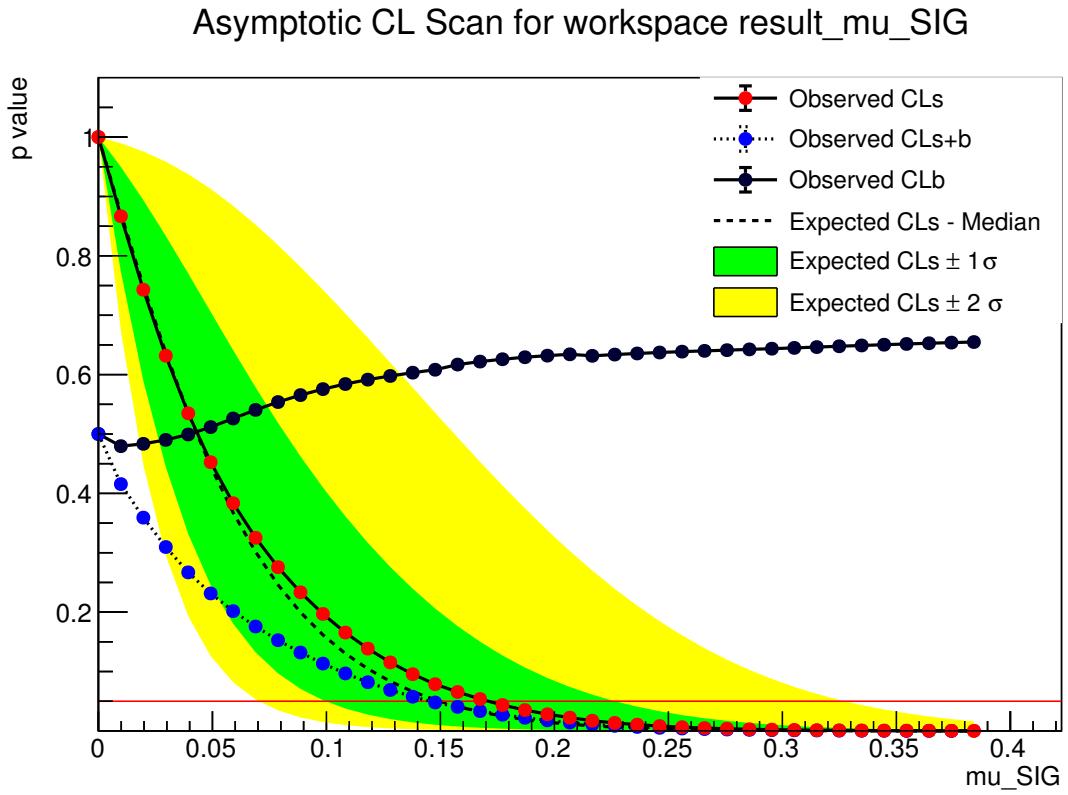
Figure I.2: Nominal m_{jj} distribution for string signal samples with ± JER uncertainty.

Appendix J Limit Scans

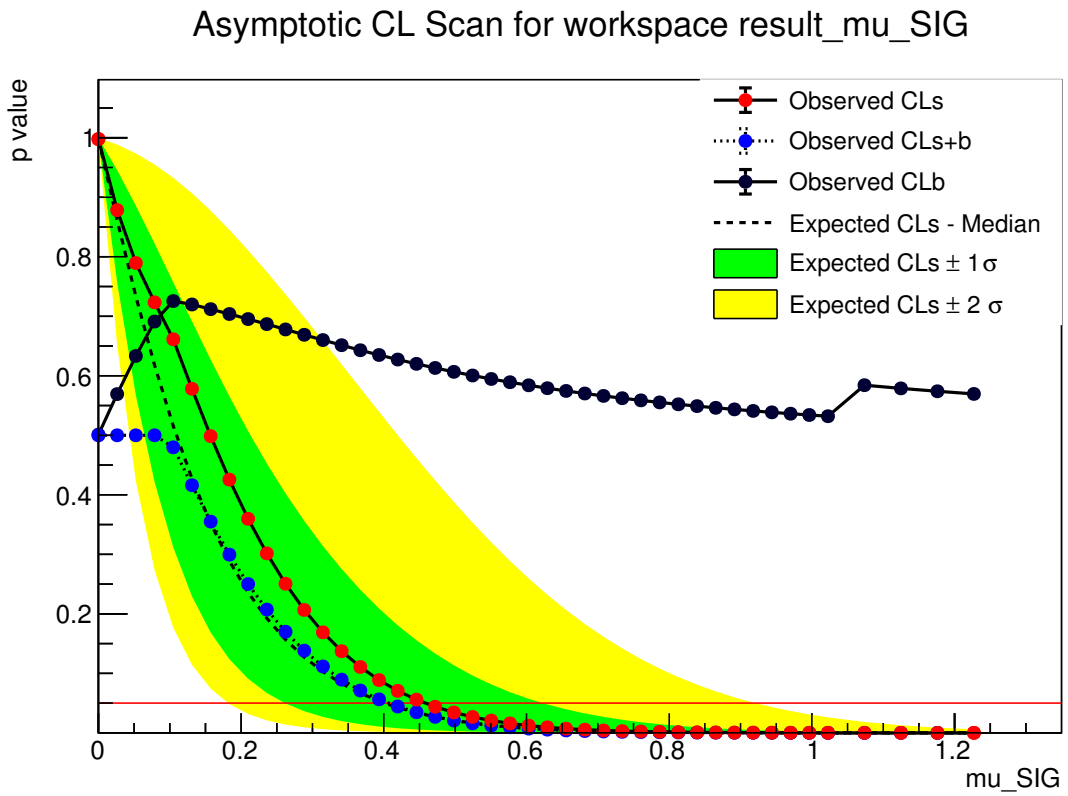
Limit scans for signal samples are shown below. These are used to calculate the upper limit on signal strength μ at a 95% CL. Upper limits on $\sigma \times A \times \epsilon$ can be computed using the results of these scans by comparing to theoretical predictions.



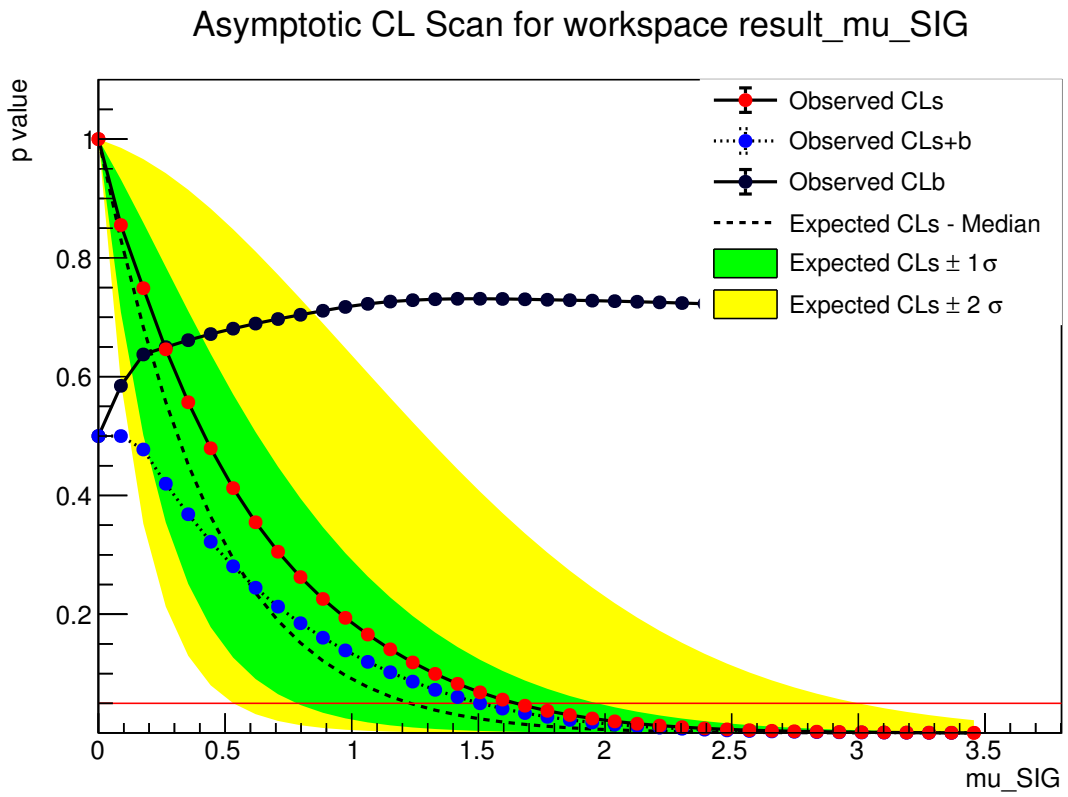
(a) $M_s = 7.0$ TeV.



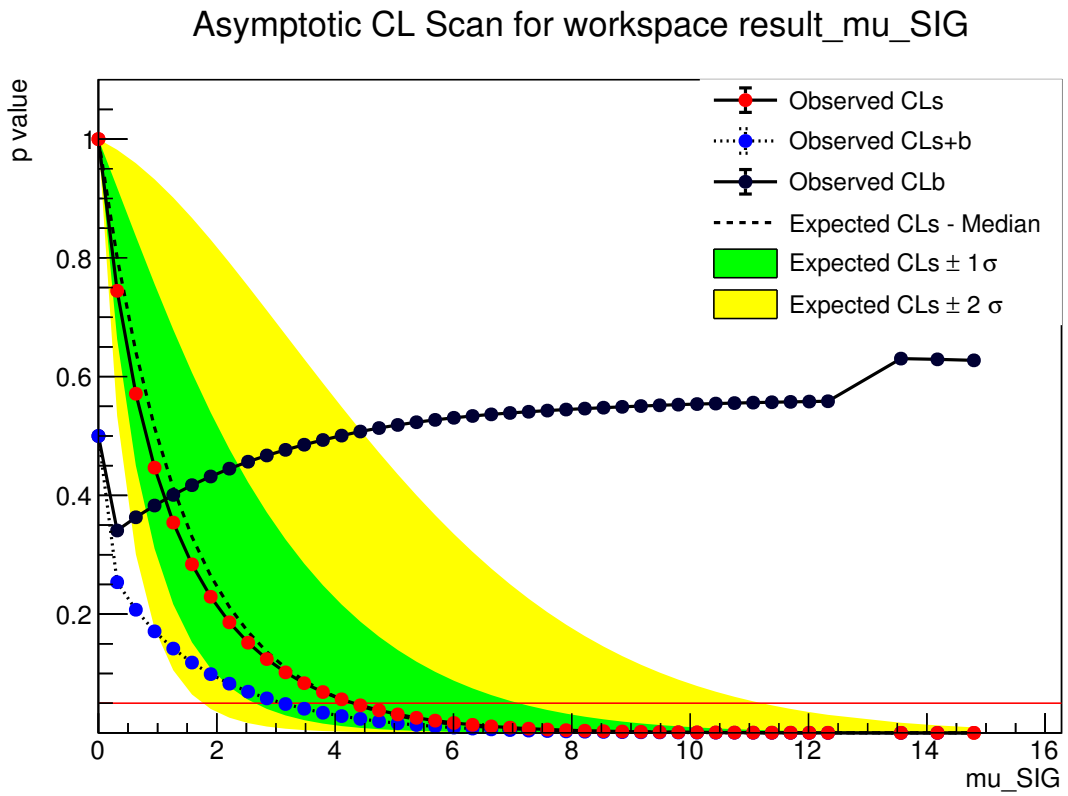
(b) $M_s = 7.5$ TeV.



(c) $M_s = 8.0$ TeV.



(d) $M_s = 8.5$ TeV.



(e) $M_s = 9.0$ TeV.

Figure J.1: Limit scan results for string resonance signal models.