

**PROJECT 3:**  
**SEARCH FOR GRAVITON RESONANCES AT  $\sqrt{s} = 13\text{TeV}$**   
**USING ATLAS OPEN DATA**  
FYS5555 - SPRING 2020

ELISABETH CHRISTENSEN

ABSTRACT. ...

## I. INTRODUCTION

In the search for a unified *theory of everything*, one must overcome the question of gravitation and the relation between the two grand theories derived within physics thus far. One of which, dedicated to describe the universe on a macroscopic level and the gravitational forces within, known as general theory of relativity, and the other, dedicated to describe the fundamental forces at a quantum level, known as the Standard Model.

The general theory of relativity, as proposed by Albert Einstein [1], emerged in 1915 in a (highly succesful) attempt to describe the gravitational interactions of matter. The Standard Model [2] on the other hand, emerged as a result of the tedious calculations of quantum field theory, and has for the past 50 years proven a valid model describing the interactions of elementary particles and the forces acting between them. So far, the Standard Model has achieved in describing, with great accuracy, three of the four fundamental forces of nature. This concludes the electromagnetic force, the weak force, and the strong force which came into place during the 1970s upon the experimental evidence of quarks. The latest contribution to the Standard Model being the Higgs boson, upon its discovery in 2012. However, it has so far not been able to account for the fourth fundamental force, i.e. gravity. This is due to the infinities arising from the singularities in Feynman diagrams [3], which, unlike in QED and QCD, cannot be renormalized away.

However, there are multiple theories devoted to solve such a problem. One of which is based on string theory, and will be the theory which allows us to search for not only graviton resonances in a proton-proton collision, but also higher dimensions apart from time and the three spatial dimensions familiar to us today. We will in this report study the production and decay of gravitons to two leptons ( $q\bar{q} \rightarrow G^* \rightarrow l^+l^-$ ), at a resonance range between 750-4000GeV, by the use of Monte Carlo simulations of the graviton based on the Randall-Sundrum model and observed data obtained

---

*Date:* June 1, 2020.

from the *ATLAS Open Data* source. The ATLAS Open Data source provides both observed and simulated data events, using Monte Carlo, of pp collisions at an integrated luminosity of  $10\text{fb}^{-1}$ .

## II. THEORY

### A. Graviton resonances

(Kaluza-Klein towers, RS-model emerging from Kaluza-Klein theory)

### B. Collider physics

In order to test a theory within particle physics we must know what to look for and how to look for it. A tool that helps us tremendously along the way is a particle detector. A particle detector allows for the detection, identification and track of the decay products emerging from a collision. An example of this is the ATLAS detector, which will provide the dataset for our analysis.

**ATLAS Detector.** The ATLAS detector consists mainly of an inner tracking detector, hadronic and electromagnetic calorimeters measuring the energy lost by the charged particles as they move through matter, and a muon spectrometer. The layout within ATLAS can be seen in the 2D-scheme from figure 1. Here, the pixel/SCT detector and the transition radiation tracker constitutes the inner parts of ATLAS, and are the first components to track the decay products of the p-p collisions. They measure, among other things, the momentum and charge of the particles, with the solenoid magnet assisting in the momentum measurements. As the particles move further out from the inner detector they reach the electromagnetic (EM) calorimeter. The EM calorimeter measures the energy deposited by any charged particle, except muons due to the density of the liquid argon used within the detector. Hadrons, such as the proton and neutron, carry enough energy to make it past the EM calorimeter and move on to the hadronic calorimeter. The hadronic calorimeter consists of iron plates and scintillating tiles [4], where the scintillation light, produced by the energy deposited from hadrons, is read out by photomultiplier tubes. Lastly, any particles that makes it through the calorimeters, i.e. muons and neutrinos, reach the outer layers of the detector, that is the muon spectrometer. The muon spectrometer consists of a large array of muon chambers each dedicated to identify and measure the momentum of the passing muons.

**Kinematic properties.** A few definitions of the parameters used in our dataset are in order. To start, we operate with polar angle  $\theta$  and azimuthal angle  $\phi$  to describe the trajectories of the particles within the detector. In a p-p collision, the center-of-mass frame is between the initial hadrons, i.e. the protons, and *not* between the colliding partons themselves.

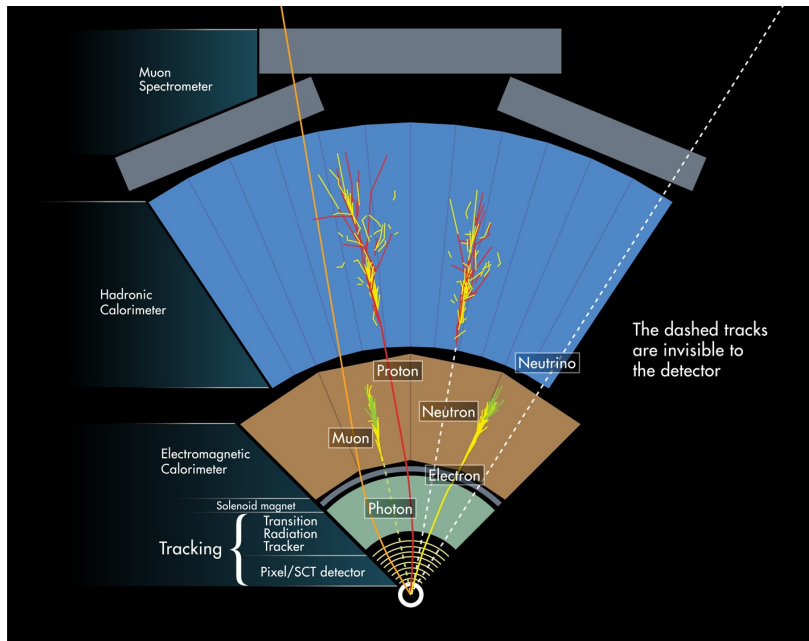


FIGURE 1. 2D layout of the ATLAS detector. [5]

Because of this, we can introduce a variable known as the rapidity,  $y$ , where the differences of rapidity are invariant under boosts along the beam direction [2]. (This insures that any longitudinal boost in a parton-parton system will not have an effect on the distribution of rapidity differences.) The rapidity is defined as

$$(1) \quad y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right)$$

where the measured quantities  $E$  and  $p_z$  represents respectively the energy and the z-component of momentum of a jet. For high-energy jets, where the jet masses can be neglected, we can use what is known as the pseudorapidity,  $\eta$ , where  $p_z \approx E \cos \theta$ , such that

$$(2) \quad \eta = -\ln \left( \tan \frac{\theta}{2} \right).$$

There are several factors involved in order to establish the energy measurements in a detector. Any missing energy may indicate either a mis-measurement, poor detector coverage or particles escaping the detector before a proper identification is made, such as neutrinos. The missing energy is often based on the missing transverse energy,  $E_T^{miss}$  as any longitudinal energy that goes undetected are generally poorly measured [6].

In order to accurately describe any physics at the interaction point of our collision, we must restrict ourselves to partons directly interacting with one another. This is done by measuring the transverse momentum,  $p_T$ , as any partons not involved in a collision will continue along the beam line, with the initial longitudinal momentum as before the interaction point. Thus,

high transverse momentum may be an indication of high mass phenomenas, i.e. potential new physics.

**Precision of detector.** The resolution of the detector determines the sensitivity for new discoveries. The tracking and energy resolution are defined as

$$(3) \quad \frac{dP}{P} = cP \oplus d, \quad \frac{dE}{E} = a/\sqrt{E} \oplus b$$

where  $c$  corresponds to the deflection angle of a particle in the magnetic field, while the factor  $d$  describes multiple scattering.  $a$  describes the stochastic fluctuations in the energy, while  $b$  describes the non-uniformity of the medium.

### C. Statistical physics

In order to study the correlation between our dataset and the theory we can make use of Bayesian statistics. Here, we can treat a parameter  $\theta$  as random. This thus allows us to compute a probability, after observing the data  $x$ , corresponding to the credibility of observing  $\theta$  within a certain interval. As an example, if our probability to observe  $\theta$  within an interval is 95% after observing the data, then we refer to this as a 95% credible interval. A credible interval of size  $1 - \alpha$  is referred to as an interval between  $a$  and  $b$ , where

$$(4) \quad P(a \leq \theta \leq b|x) = 1 - \alpha$$

with  $\alpha$  corresponding to the one-sided tail of a distribution. In the case of a two-sided tail, our credible interval is defined by

$$(5) \quad P(\theta < a|x) = \alpha/2, \quad P(\theta > b|x) = \alpha/2.$$

Bayesian statistics allows us to produce an upper limit on the amount of signal events,  $s_{up}$ , required in order for the true parameter to lie within our credible interval. In doing so, we can test the hypothesis  $H_1$  (signal+background), vs the null hypothesis  $H_0$  (background).

Bayes theorem states that

$$(6) \quad p(\theta|x) = \frac{P(x|\theta)p(\theta)}{\int P(x|\theta)p(\theta)d\theta}$$

Here,  $p(\theta|x)$  corresponds to the *posterior density function*,  $P(x|\theta)$  is the likelihood function based on a given dataset and  $p(\theta)$  corresponds to the *prior probability function*. The integral in the denominator is included for normalization. The prior probability function depends on our prior knowledge of the system, which, in our case, will correspond to a Poisson distribution. The posterior density function thus tells us how probable our hypothesis  $\theta$  given the data and our prior knowledge of the system. The more informative our prior is, the greater its influence on the posterior. The greater our sample size is, the greater the effect of the likelihood function on the posterior.

For our analysis, we will make use of the Bayesian single-bin limit setting. The likelihood function can then be expressed as

$$(7) \quad P(x|\theta) = L(s) = P(n_{obs}|s) = \frac{\lambda^{n_{obs}} e^{-\lambda}}{n_{obs}!}, \quad \lambda = s + b$$

where  $s$  is the expected signal,  $b$  is the expected background and  $n_{obs}$  is the actual number of observed events. In the case of background uncertainty we can use the *joint posterior*, where we include a prior,  $p(b)$ , on the uncertainty. The joint posterior is expressed as

$$(8) \quad p(s, b|n_{obs}) = \frac{P(n_{obs}|s, b)p(s)p(b)}{P(n_{obs})}$$

while  $\lambda$  can be expressed as

$$(9) \quad \lambda = L_{int}(\varepsilon_{sig}\sigma + \sigma_{bkg}^{eff})$$

where  $L_{int}$  is the integrated luminosity,  $\sigma$  is the signal cross-section with  $\varepsilon_{sig}$  being the signal efficiency. The likelihood  $L(\sigma, L_{int}, \varepsilon_{sig}, \sigma_{bkg}^{eff})$  can still be expressed as eq. (7). But we must marginalize to find the posterior for  $\sigma$ , given the number of observed events. In such a case, we end up with a multi-dimensional integral defined as

$$(10) \quad p(\sigma|n_{obs}) = N \int L(\sigma, L_{int}, \varepsilon_{sig}, \sigma_{bkg}^{eff}) p(L_{int}) p(\varepsilon_{sig}) p(\sigma_{bkg}^{eff}) dL_{int} d\varepsilon_{sig} d\sigma_{bkg}^{eff}$$

which, without having to go through long and tedious calculations, can be solved by Monte Carlo.

## D. Event selection

The event selections for this dataset follows closely the criteria described in Ref. [7], and is summarised in table 2 along with the preselection requirements for the 13 TeV ATLAS Open Data in table 1. The primary cut is made on events containing exactly two same-flavour leptons. For a muon pair we require that the two leptons must be of opposite charge. However, this restriction does not apply to a dielectron pair as charge-misidentification increases for higher  $E_T$  electrons. Furthermore, we require the impact parameter  $z_0 \sin \theta > 0.5\text{mm}$  for both muons and electrons, whereas the transverse impact parameter  $|d_0/\sigma(d_0)|$  is required to be less than 5 for electrons, and less than 3 for muons. The isolation criteria made on the electron and muon candidates requires that the scalar sum of track  $p_T$ , in a cone of radius  $\Delta R$  around a lepton (described by lep\_ptcone30)<sup>1</sup>, must be less than 0.15. The calorimeter isolation on a lepton requires that the scalar sum of track  $E_T$ , in a cone of radius  $\Delta R'$  around the lepton (described by lep\_etcone20)<sup>2</sup>, must be within the range  $(-0.1, 0.2)$ .

<sup>1</sup> $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} = 0.3$  and decreases as a function of  $p_T$ .

<sup>2</sup> $\Delta R'$  has a maximum value of 0.2.

TABLE 1. Event preselection criteria for electron candidates and muon candidates within the 13TeV ATLAS Open Data

Preselection	
Electron ( $e$ )	Muon ( $\mu$ )
InDet & EMCAL rec.	InDet & MS rec.
loose identification	loose identification
loose isolation	loose isolation
$p_T > 7\text{GeV}$	$p_T > 7\text{GeV}$
$ \eta  < 2.47$	$ \eta  < 2.5$

TABLE 2. Event selection closely following the criteria as specified in Ref. [7].

Electron ( $e$ )	Muon ( $\mu$ )
$E_T > 30\text{GeV}$	$p_T > 30\text{GeV}$
$ \eta  < 2.47$ & $1.37 <  \eta  < 1.52$	$ \eta  < 2.5$ & $1.01 <  \eta  < 1.1$
$z_0 \sin \theta > 0.5\text{mm}$	$z_0 \sin \theta > 0.5\text{mm}$
$d_0/\sigma(d_0) < 5$	$d_0/\sigma(d_0) < 3$
$-0.1 < lep\_etcone30 < 1.1$	$lep\_ptcone30 < 0.06 \cdot p_T$

In order for an electron to pass through the fine-granularity of the EM calorimeter, it must have  $E_T > 30\text{GeV}$  and satisfy  $|\eta| < 2.47$ . It must also have  $1.37 < |\eta| < 1.52$  in order to pass the transition region between the barrel and endcap EM calorimeters. A muon candidate must have  $p_T > 30\text{GeV}$  and  $|\eta| < 2.5$ . In order for a muon to pass the barrel-endcap overlap region it must also be outside the region  $1.01 < |\eta| < 1.1$ . A muon candidate must also have that the summed scalar  $p_T$ , within a cone of radii  $\Delta R$  around the lepton, is less than 6% than the  $p_T$  of the muon itself. Lastly, we require that the invariant mass of a dilepton system,  $m_{ll}$ , is greater than  $225\text{GeV}$  in order to avoid the Z boson peak region.

### III. RESULTS

By performing the criteria as mentioned in table 2 on both the expected signal and background events as obtained by Monte Carlo we can obtain the  $E_T^{miss}$ ,  $m_{ll}$  and  $p_T$  for the individual dilepton pairs. The plots for these parameters are shown in figure 2, and illustrates the spectra for the data, expected background and five signal samples of masses ranging between  $750 - 4000\text{GeV}$ . It must be mentioned, however, that the simulated signal samples for the decay of  $G^* \rightarrow ee$  was unfortunately not correct as the majority of the decay products was of the wrong lepton type, and is thus not included in the spectra for the  $E_T^{miss}$ ,  $m_{ll}$  and  $p_T$  involving the dielectron channel. We can see that, for all parameters, the agreement between the data and expected background generated by MC overlaps well in the lower regions where the number of events are of the order  $\sim 10^4$ . We can see from the  $m_{ll}$  plot that as we move to higher regions, the number of events

TABLE 3. Inputs used for the limit calculation of  $s_{up}$  in the Monte Carlo integration of the posterior. The first column corresponds to the invariant mass of the graviton, with the mass region  $m_{ll}$  declared in the second column. For each channel we calculate the number of expected signal events,  $N_{sig}$ , the number of background events,  $N_{bkg}$  and the number of observed events,  $N_{obs}$ . The calculation is done by integrating over the number of events within a mass region. The uncertainty on  $N_{bkg}$  and  $N_{obs}$  corresponds to the uncertainty of this integration.

$m_{G^*}$ [GeV]	$m_{ll}$ region [GeV]	Channel	$\varepsilon_{sig}$	$N_{sig}$	$N_{bkg}$	$N_{obs}$
750	(680, 820)	$ee$	$1.0 \pm 0.0$	-	$337.45 \pm 38.79$	339
		$\mu\mu$	$1.0 \pm 0.0$	$11224.33 \pm 75.55$	$307.91 \pm 41.60$	251
1000	(800, 1200)	$ee$	$1.0 \pm 0.0$	-	$125.98 \pm 22.81$	105
		$\mu\mu$	$1.0 \pm 0.0$	$2622.94 \pm 17.81$	$121.20 \pm 28.40$	80
2000	(1500, 2500)	$ee$	$1.0 \pm 0.0$	-	$10.55 \pm 6.47$	6
		$\mu\mu$	$1.0 \pm 0.0$	$46.99 \pm 0.33$	$3.66 \pm 4.02$	5
3000	(2000, 4000)	$ee$	$1.0 \pm 0.0$	-	$6.92 \pm 5.12$	3
		$\mu\mu$	$1.0 \pm 0.0$	$2.77 \pm 0.02$	$4.16 \pm 4.28$	1
4000	(2500, 5500)	$ee$	$1.0 \pm 0.0$	-	$0.22 \pm 0.12$	0
		$\mu\mu$	$1.0 \pm 0.0$	$0.27 \pm 0.00$	$10.30 \pm 7.93$	0

drops according to the Poisson distribution, such that the uncertainty on the ratio of data/MC increases accordingly. It should also be noted that the background lies higher than the data in the regions between 600 – 900GeV before the data overtakes the MC samples. This is as mentioned due to the small sample sizes from 600GeV and up. Eventually, the contrast between the data and background events becomes high enough such that the ratio lies beyond the limit of 2. We can also see from the plot showing the invariant mass spectrum that there is a lack of data points beyond 2500GeV. This is simply due to the fact that there are no observed events beyond this limit.

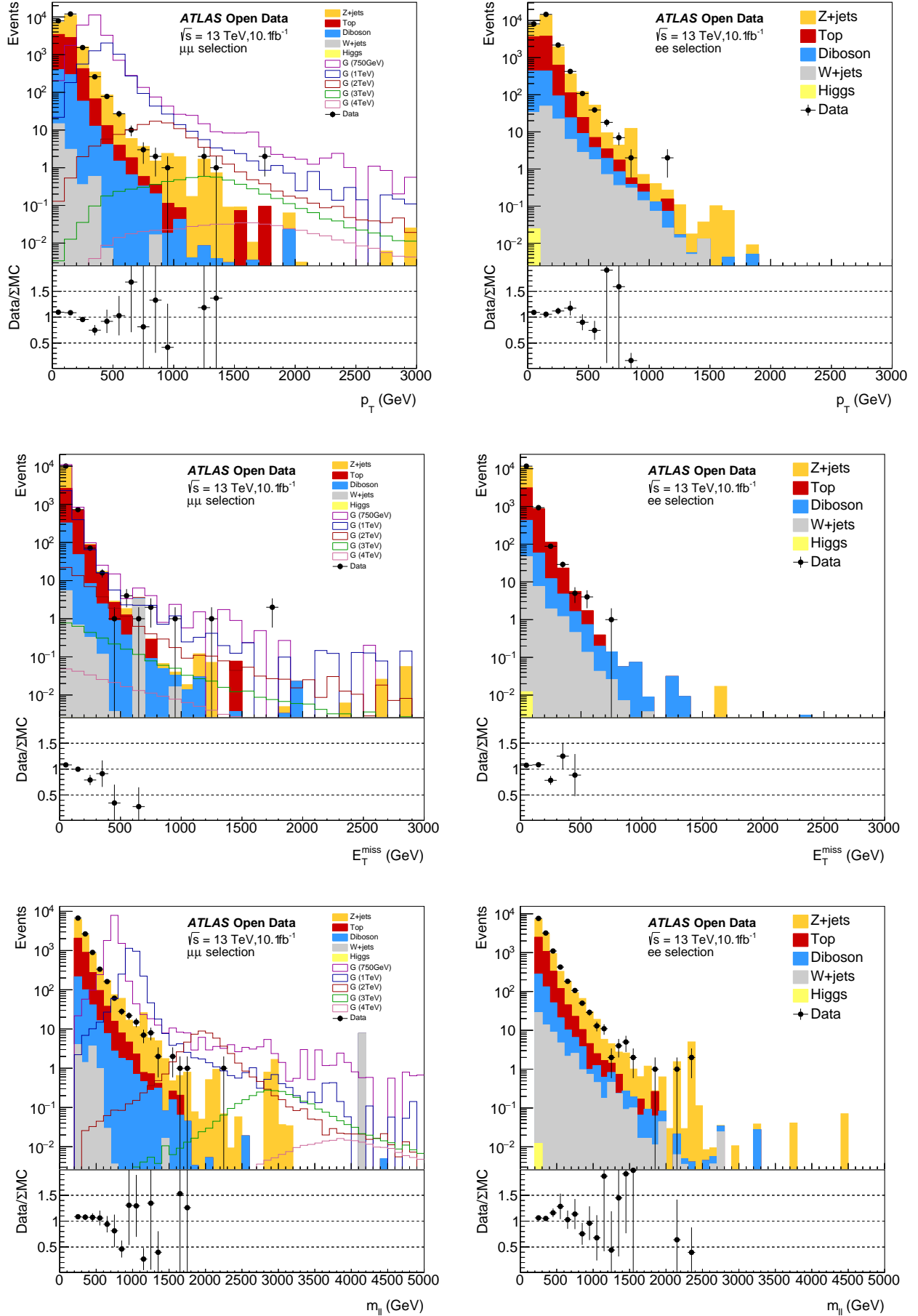
In order to find the posterior for the number of expected events given our data, we will make use of a 95% credible interval. This is done by solving eq. (10) through a Monte Carlo simulation, where we choose to set the integrated luminosity  $L_{int} = 1.0$  with zero uncertainty, and the signal efficiency  $\varepsilon_{sig} = 1.0$  with zero uncertainty, in order to obtain an upper limit,  $s_{up}$ , on the expected number of signal events. The inputs for the Monte Carlo integration can be seen in table 3 for both channels  $G^* \rightarrow ee$  and  $G^* \rightarrow \mu\mu$ .<sup>3</sup> The limits produced by the inputs from table 3 can be seen in table 4.

<sup>3</sup>The number of events and its uncertainty within a certain mass region was found using the function `IntegralAndError` in PyROOT.

TABLE 4. The first column corresponds to the invariant mass of the graviton with the corresponding decay products in column two. The third column  $s_{up}$  is the upper limit on the number events with a 95% certainty that  $s$  will lie below. The limits is shown from columns four to seven showing the corresponding credible levels with median.

$m_{G^*}$ [GeV]	Channel	$s_{up}$	$-2\sigma$	$-1\sigma$	Median	$+1\sigma$	$+2\sigma$
750	$ee$	79.39	36.67	52.88	78.02	116.20	164.68
	$\mu\mu$	46.23	36.70	54.04	80.76	121.07	174.39
	both	2.09	1.44	1.98	2.96	4.41	6.08
1000	$ee$	30.59	19.69	29.45	44.99	68.25	100.88
	$\mu\mu$	24.08	20.37	31.68	50.59	80.51	122.23
	both	0.89	0.77	1.15	1.72	2.57	3.88
2000	$ee$	7.23	4.22	6.42	11.80	20.11	36.68
	$\mu\mu$	8.34	3.01	3.98	7.20	12.98	29.21
	both	0.27	0.11	0.16	0.29	0.51	1.01
3000	$ee$	5.36	3.72	5.36	9.19	16.73	31.37
	$\mu\mu$	3.93	3.02	3.93	7.06	13.87	30.01
	both	0.13	0.11	0.16	0.27	0.48	0.86
4000	$ee$	3	3	3	4.548	6.066	-
	$\mu\mu$	3.01	3.70	5.97	11.51	23.04	44.97
	both	0.07	0.09	0.13	0.16	0.19	0.26




 FIGURE 2. Parameters  $p_T$ ,  $E_T^{miss}$  and  $m_{ll}$  for the electron ( $e$ ) and muon ( $\mu$ ) channel.

#### IV. DISCUSSION

#### V. RESULTS

#### REFERENCES

- <sup>1</sup>A. Einstein, “The Field Equations of Gravitation”, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys. ) **1915**, 844–847 (1915).
- <sup>2</sup>M. Thomson, *Modern Particle Physics* (Cambridge University Press, 2019).
- <sup>3</sup>E. Witten, “Reflections on the fate of spacetime”, in *Physics Meets Philosophy at the Planck Scale - Contemporary Theories in Quantum Gravity* (Cambridge University Press, 2001) Chap. 5, pp. 125–137.
- <sup>4</sup>A. Henriques, *The ATLAS Tile Calorimeter*, <https://cds.cern.ch/record/2004868/files/ATL-TILECAL-PROC-2015-002.pdf>, Accessed: 2020-05-26.
- <sup>5</sup><http://cds.cern.ch/record/1505342?ln=en>, Accessed: 2020-05-26.
- <sup>6</sup>D. Green, *High  $pT$  Physics at Hadron Colliders* (Cambridge University Press, 2004).
- <sup>7</sup>ATLAS Collaboration, “Search for high-mass dilepton resonances using  $139fb^{-1}$  of  $pp$  collision at  $\sqrt{s} = 13\text{TeV}$  with the ATLAS detector”, Physics Letters B **796**, 68 –87 (2019).