Report Summer Programme CERN 2017

FCChh: RSGraviton $\rightarrow$ WW analysis

Ine Arts

Supervisors
Michele Selvaggi
Clement Helsens
Filip Moortgat

8-9-2017

Contents

1 Introduction

First of all I want to start by saying I had an amazing time during the 10 weeks I've been in CERN. I got to be part of the exciting FCChh project and learned a lot about particle physics by working on this analysis, but also in the lectures. Therefore I would like to thank my supervisors Michele Selvaggi, Clement Helsens and Filip Moortgat and fellow summer student Rachel Smith.

1.1 FCC

The FCC-project is a study for the design of a new particle collider, the Future Circular Collider. It's a successor for the LHC, in order to continue the existing research. The FCC will push both the energy and intensity frontiers. It aims to reach a center of mass energy of 100TeV.[1]
This energy can be obtained by a machine that measures 100 km in circumference and 16 T dipoles, or 80 km with 20 T. To reach this, there need to be new technologies; the dipole magnets in the LHC, for example, can only reach 8.3 T, while double is needed.
Even though this collider isn't to be built in the near future, the design and construction can take more than 20 years to complete. So in order to be ready by 2035, the research has already started. The study focusses on very different aspects: different possibilities for a circular collider, new detector facilities and infrastructure, cost estimates, new technologies, physics and detector research, etc... [2]
There will be three different physics scenarios: hadronic collisions (pp and heavy ion, similar to the LHC), electron-positron collisions (like LEP) and proton-electron collisions. This last one is interesting for precise measurements of deep inelastic scattering and Higgs physics.

1.2 Physics at the FCC

At 100 TeV the physics will be a little different from the LHC. The differences in SM (standard model) physics compared to the LHC are explained by unknown particles or interactions. Because of the high energy it is expected that heavy BSM (beyond the standard model) particles will be produced, which will decay to lighter SM particles. BSM interactions will also influence production properties of SM particles. The SM final state can be a signature for BSM particles and can also show underlying BSM dynamics.
Unfortunately this makes 'normal' SM interactions both a signal and a background. Which is a challenge, since it can hide the signal and influence trigger strategies (because it reduces the ability to save interesting data for the research).
The production rates for SM physics at 100TeV will be huge. This not only increases the chances for rare phenomena to be discovered, but also gives us the possibility to perform a precise measurement of SM parameters and a possible test to discover deviations from the SM. [3]

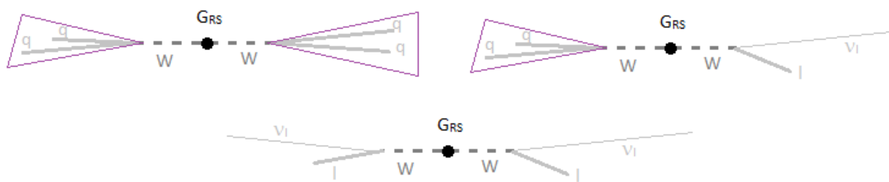


Figure 1: Full hadronic, half leptonic and full leptonic decay of G_{RS}

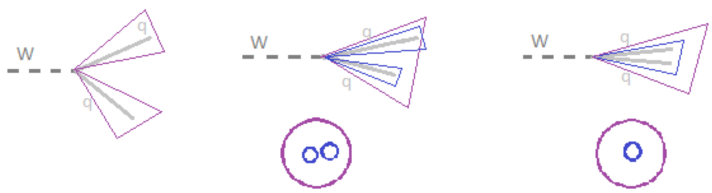


Figure 2: Full hadronic decay of a W boson. A boosted boson produces one *fat* jet and for very high transverse momentum the subjets overlap.

2 Analysis

2.1 RSGraviton to WW

The RSGraviton (G_{RS}) is a hypothetical elementary particle. Therefore some properties, like it's mass, are unknown. For the simulations different possible masses are used from 2 TeV to 40 TeV. G_{RS} can decay to two W bosons. Because of it's high mass, the W bosons that are produced by its decay will have a high momentum and are called *boosted bosons*.
Each of the bosons can then decay to either a quark-antiquark pair or to a lepton and a neutrino. The hadronic decay is more dominant, with branching ratio around 67%.
When looking at the event as a whole, there are three scenarios; a full hadronic (both bosons decay to quarks), a half leptonic (one of the bosons decays to a lepton neutrino pair) and a full leptonic decay. All are displayed in figure 1. The last one is quite rare and will be ignored in the analysis.
To determine whether an event has a W boson, we want to tag a boosted boson. The hadronic decay of the boson produces two quarks with an angular separation (ΔR) equal to $2\frac{M}{p_T}$. This means that the higher the p_T, the closer the two quarks to each other, illustrated in figure 2. A boosted boson will be seen in a detector as one, massive jet (a *fat jet*). This massive jet has an internal structure of two subjets (one for each quark). For very high momentum, the two subjets start to overlap.
An event will be considered full hadronic when there are two tagged boosted bosons and half leptonic when there is at least one lepton and a tagged boosted boson.
The possible backgrounds for $G_{RS} \to WW$ are: 2 jets, $t\bar{t}$, VV (this is WW or

ZZ) and W+jet(s). The dijet background (often referred to as *QCD jets*) is the most dominant one, so most cuts are based on differentiating W from QCD jets. When demanding half leptonic events, the QCD background will be eliminated and W+jet(s) becomes important.
To extract the signal from those backgrounds, we'll use different variables on which one can cut: jet p_T, soft dropped jet mass, N-subjettiness, flow variables, number of leptons, etc.

2.1.1 Softdropped jets

In order to measure more accurately the mass of the jet, one can apply a grooming method. This is a jet substructure algorithm that removes wide-angle soft radiation from a jet. It reduces the contamination from initial state radiation, underlying events and pileup. For a jet with only two constituents and radius R_0, the soft drop condition is as follows. [4]

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta} \tag{1}$$

The degree of jet grooming is controlled by z_{cut} and β. When $\beta \to \infty$, the algorithm returns an ungroomed jet.

2.1.2 N-subjettiness

Since both W and QCD jets can have the same mass (around 80 GeV), we have to use jet shape to differentiate them. The first one is by 'counting' the number of hard lobes of energy within a jet. The new variable is called *N-subjettiness*. After identifying N candidate subjets, the N-subjettiness (τ_N) is calculated:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\left(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\right) \tag{2}$$

With k running over constituent particles in the jet and $p_{t,k}$ their transverse momenta, $\Delta R_(J,k) = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the distance between constituent particle k and candidate subjet J, and $d_0 = \sum_k p_{T,k} R_0$ is the normalization factor (with R_0 the characteristic jet radius used in original jet clustering algorithm). [5]

The variable τ_N shows to what degree a jet can be regarded as a jet composed of N subjets. Jets with $\tau_N \approx 0$ have al their radiation aligned with candidate subjets and therefore have N or fewer subjets. Jets with $\tau_N \gg 0$ have at least $N+1$ subjets.
One might expect that a jet with small τ_2 and large τ_1 is more likely to be a W jet, but it is observed that QCD jets can also have small τ_2 and QCD jets with a diffuse spray of large angle radiation can also have large τ_1. However QCD with a large τ_1 typically have large values of τ_2 as well. So in fact, the ratio $\frac{\tau_2}{\tau_1}$ will be the discriminating variable.

2.1.3 Flow variables

Other jet substructure variables are flow variables. These show how well spread the energy is within the jet. It can be used as a tagging algorithm, since a W jet has almost all its energy in the center and it's quite empty outside of the center, while a QCD jet is more spread out. There are 5 flow quantities:

$$\text{Flow}_{n,5} = \sum_k \frac{|p_T^k|}{|p_T^{\text{jet}}|} \tag{3}$$

Here is $n = 1, \ldots, 5$. The summation goes over the values of k for which: $\frac{n-1}{5} R \leq \Delta R(k, \text{jet}) < \frac{n}{5} R$, where R is the jet size and $\Delta R(k, \text{jet}) = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ the angular distance between jet constituent k and the jet axis. [3]

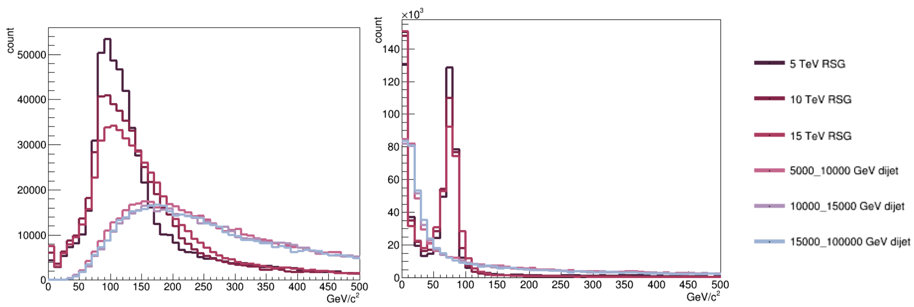
2.1.4 Boosted Decision Trees

All of the above variables can be used to construct one new variable, using multivariate analysis. There are a lot of methods available for machine learning in the MVA package in ROOT, the one that is used here is called a boosted decision tree (BDT). A decision tree (DT) exist of different questions (nodes) following each other. Each node has two options and at the end of the structure there is a decision (leaf): an event is regarded as either signal or background. This technique is fast, but not always accurate. To make the trees more reliable, boosting is applied. This method enhances the weight of misclassified events, so that the new DT's learn better. It is important not to iterate this too much as it may cause overtraining. [6]

2.2 Workflow

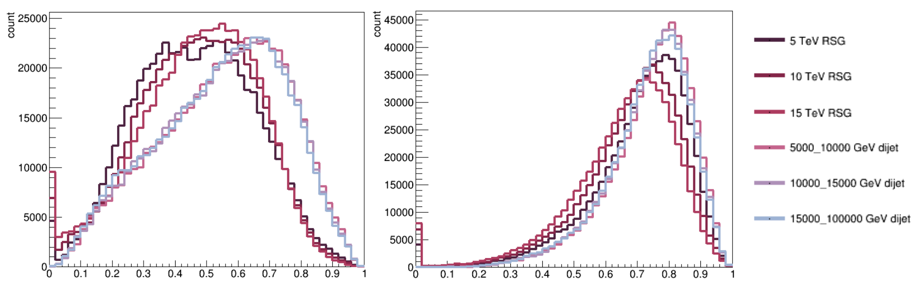
The full analysis goes as follows:

1. Simulate detector response with Delphes over LHE (Les Houche event) files previously produced with PYTHIA.
2. Analyse the output with Heppy and store the useful variables in FlatTrees.
3. Analyse those FlatTrees by applying cuts and creating plots.
4. Calculate confidence levels and significances for the obtained plots with FCCfitter.



(a) Ungroomed (b) Softdropped

Figure 3: Jet mass: W and QCD (different energies)



(a) τ_{21} (b) τ_{32}

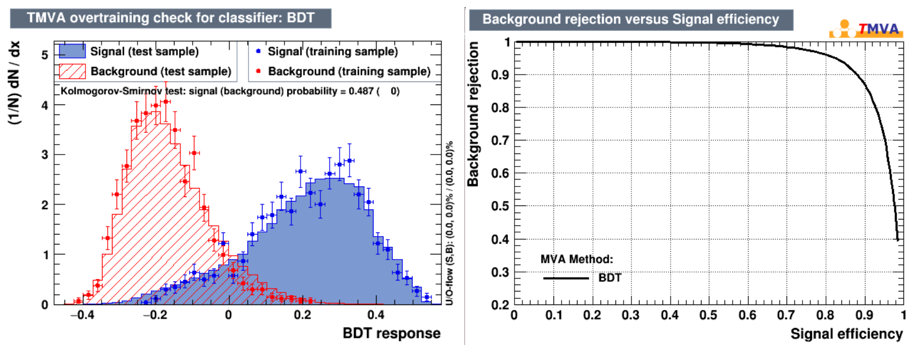
Figure 4: Ratios N-subjettiness: W and QCD (different energies)

2.3 Results

Soft drop Because of the soft drop algorithm, we can make a very good discrimination between W and QCD jets. By getting rid of the wide angle soft radiation, the jet mass plot for the signal shows a very fine peek at 80 GeV (in figure 3b). The background, however, has almost no mass after applying the soft drop method.

τ_{21} and τ_{32} In figure 4 $\tau_{21} = \frac{\tau_2}{\tau_1}$ and $\tau_{32} = \frac{\tau_3}{\tau_2}$ are drawn for W and QCD jets (three different p_T's). For QCD, both τ_{21} and τ_{32} are independent on the energy of the jet. τ_{21} offers a pretty clear discrimination between W and QCD jets (in contrast to τ_{32}), but the higher the G_{RS} mass, the more the signal overlaps with the background. This is expected; the more boosted the W boson is, the more the subjets overlap, until there is only one visible in the detector. Therefore τ_{21} becomes a less good discriminating variable for high G_{RS} masses.

BDT variable The BDT algorithm takes 9 variables as input and creates one new variable that can discriminate W and QCD jets. The input variables are: p_T, τ_{21}, τ_{32}, the 5 flow variables and $m_{\text{softdropped}}$. The BDT response for the



(a) BDT classifier

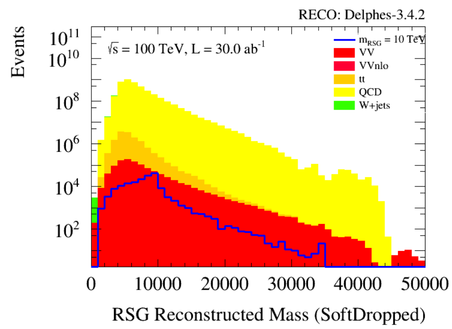
(b) ROC curve for BDT classifier

Figure 5: BDT response: W (signal) vs QCD jets (background)

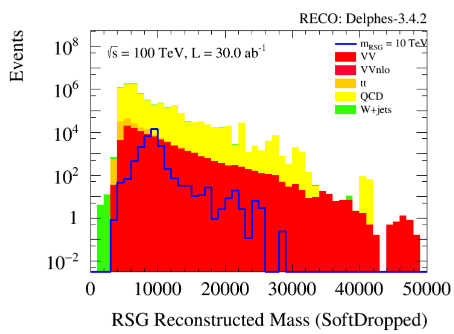
train and test samples (10 TeV signal vs QCD background) are displayed in figure 5a. The signal is very well separated from the background.
To illustrate how good a classifier is, one can draw a ROC (Receiver Operating Characteristic) curve. This curve shows the background rejection and signal efficiency by scanning over possible cuts on the variable. The best classifier is the one with the largest area under its curve. The ROC curve for the BDT variable is shown in figure 5b.

RSGraviton reconstructed mass Based on the four-momentum of the (soft-dropped) jet, the invariant mass of the RSGraviton can be reconstructed. In figure 6 is the reconstructed mass for a 10 TeV RSGraviton compared to the backgrounds. Figure 6a is without cuts, 6b full hadronic and 6c is half leptonic. Note that this is on a log scale, with the background stacked while the signal is not stacked. The most prominent background for full hadronic events is QCD and for half leptonic events it is W+jets. Contrary to the previous plots, these are weighted by the cross section and the number of events.
For the signal there will be a clear maximum at the mass used to simulate the data. The background should have its maximum near zero, but because of restrictions on the jets in the heppy analysis (minimum p_T) this is not the case in figure 6. For the half leptonic events the maximum at 10 TeV is not very clear, this is because the neutrino's energy can not be measured in the detector.

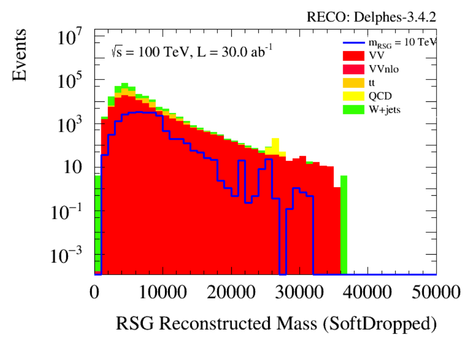
Limitplots The last step is calculating confidence levels and significances. Figure 7 shows the limit plots. The red line is the theory expectation and the dotted line shows the expected limet at 95% confidence level (CL). When the dotted line is below the red, the signal hypothesis can be excluded at 95% CL. For full hadronic cuts we can exclude up till 15 TeV and for half leptonic this is 20 TeV. This means that in real data it will be highly unlikely for a signal below this mass to exist.



(a) without cuts

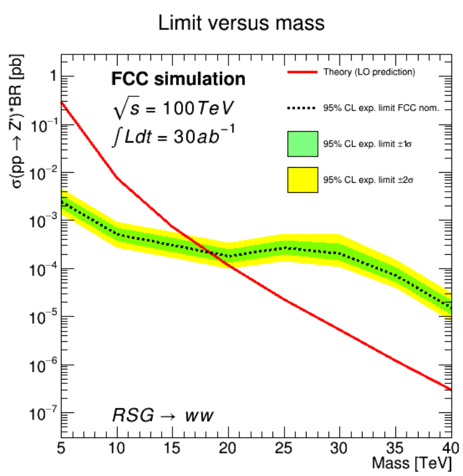


(b) full hadronic events

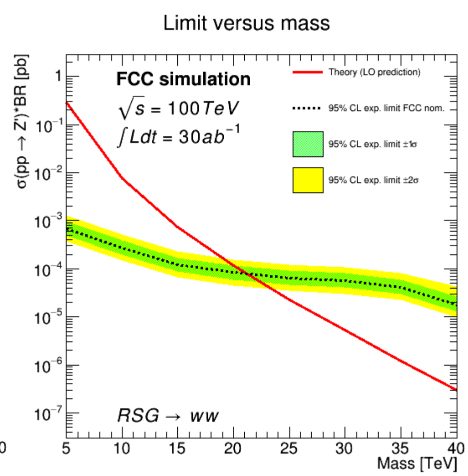


(c) half leptonic events

Figure 6: RSGraviton reconstructed mass (10 TeV)



(a) full hadronic events



(b) half leptonic events

Figure 7: limitplots

3 Conclusion

First of all we can conclude that for boosted w tagging, the jet substructure variables (τ_{21} and flow) separate signal and background quite well as long as the G_{RS} mass is not too high. For higher mass it is harder to discriminate W from QCD jets. The achieved BDT variable looks very promising as discriminator (figure 5a), but there was not enough time to implement this in the analysis.
Figure 7 shows that it is unlikely that we will measure an RSGraviton with mass below 20 TeV decaying to two W bosons in the FCC.
A lot of further research is possible. Within this analysis there is still the possibility to implement the MVA in heppy, to make a combined limit plot for full hadronic and half leptonic events and to use the CMS detector design instead of the FCC. But most importantly this workflow will be used in other (FCC) analyses.

References

[1] Jarlett H.K., *The future cicular collider*, internet article, http://cds.cern.ch/record/2138983, last visited 4-9-2017

[2] Benedikt M., *Future circular collider study*, presentation, http://cds.cern.ch/record/2206380/files, last visited 4-9-2017

[3] Mangano M.L., Zanderighi G. et al., *Physics at a 100 TeV pp collider: Standard model processes*, arXiv:1607.01831, 2016, 257 p.

[4] Larkoski A.J., Marzani S., Soyez G. and Thaler J., *Soft Drop*, arXiv:1402.2657, 2014, 46 p.

[5] Thaler J. and Van Tilburg K., *Identifying boosted objects with N-subjettiness*, arXiv:1011.2268, 2010, 27 p.

[6] Roe B.P., Yang H. et al., *Boosted decision trees as an alternative to artificial neural networks for particle identification*, arXiv:physics/0408124, 2004, 6 p.