Endcap Photon Fake Rate Study: Background Estimation for ADD and RS Graviton Searches

CILICIA UZZIEL M. PEREZ*
National Institute of Physics
University of the Philippines, Diliman
uzzie.perez@cern.ch

Abstract

We present a photon fake rate calculation for the endcaps that is used to estimate the reducible background components for high-mass diphoton resonances. Such resonances could be a signature of an excited state of either a Randall-Sundrum or ADD graviton. Quark and gluons may fragment in such a way that they produce fake photon signatures in the detector. A proper estimation of the reducible background is important in setting limits for the search for any excess over the total background. The data presented here are from proton-proton (pp) collisions at a centre-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 19.5 fb$^{-1}$. The same method is planned to be applied for data taken at a centre-of-mass energy of 13 TeV.

I. INTRODUCTION

The discovery of extra dimensions could be a solution to the Standard Model’s Hierarchy Problem. The ADD and RS Models both propose the existence of additional spatial dimensions where the latter involves a warped geometry with negative curvature. This leads to massive excitations of the graviton that could manifest themselves as peaks in the diphoton invariant mass spectrum. In this search, we look for an excess of events over the total background. The branching ratio for the diphoton channel is twice that of leptons and has a superior energy resolution compared with the dijet channel.

Different sources of background need to be taken into account:

- γγ: Events with prompt photons from Standard Model (SM) Born (quark annihilation) and Box (gluon fusion) processes. This is the only irreducible background for this search and is the dominant source of background.
- γ+jets: Events with a prompt photon and a jet that fakes a photon.
- QCD dijets: Events with two jets misidentified as photons.

In the next sections, we will discuss about how photons are identified in the CMS detector, the Fake Rate Method, a data-driven approach in estimating the background arising from misidentified photons, the Fake Rates for the CMS detector Endcaps and the software tools and libraries used for the summer project.

II. PHOTON IDENTIFICATION

In order to discriminate as much as possible between the signal and the sources of background, we apply photon ID criteria based on information from the CMS detector.

A detailed presentation of the CMS detector can be found elsewhere [3]. The main important subdetectors to the analysis are the electromagnetic calorimeter (ECAL) which measures the energy and position of electrons and photons while the hadronic calorimeter (HCAL) measures the energy and position of jets.

The following features are considered for
we employ a data-driven approach which is a very particular case of quark/gluon fragmentation leads to neutral mesons that subsequently decay to two highly boosted photons. The signatures of these two photons may overlap within the resolution of the calorimeter.

Estimating this reducible background from fake photons is not guaranteed to be accurately described by Monte Carlo predictions. Hence, we employ a data-driven approach which is based on the methodology described in [2]. The fake rate for the CMS detector barrel for the $\sqrt{s} = 8$ TeV data has been previously calculated. In this study, we present the fake rates calculated in different $p_t$ bins for the endcaps (1.56 < $|\eta_{\text{det}}|$ < 2.5) [3].

The fake rate method measures, in data, the relative rates of two types of jet fragmentation. The first type, the numerator, contains jets from quarks/gluons that fragment in such a way as to pass the photon identification. The denominator contains jets that pass criteria looser than the photon identification. The fake rate formula we used is defined below:

$$ f = \frac{J_{\gamma \text{LOOSE}}}{J_{\gamma \text{LOOSER}}} $$

where $J_{\gamma \text{LOOSE}}$ is the number of jets identified as photons and $J_{\gamma \text{LOOSER}}$ is the number of photon-like jets that pass criteria looser than the photon identification, and in addition failed at least one of the criteria typically associated with real photons. In particular, these are the CHIso, NHIso, and $\sigma_{i\eta\phi}$. The numerator in eqn. (1) is defined by selection criteria that admit both real and "fake photons". The real photon contamination must then be subtracted because our definition from the fake rate involves only jets.

To subtract the real photon contamination, we consider a photon ID variable which is sensitive to the difference between real and fake photons. In this study, we consider $\sigma_{i\eta\phi}$. We extract the fake photon templates, in data, by applying the criteria from Table 1 except for $\sigma_{i\eta\phi}$ and in addition, use a sideband in Charged Hadron Isolation (5 GeV < CHIso < 10 GeV). This sideband choice was designed to minimize the fraction of real photons in the fake templates.

A very particular case of quark/gluon fragmentation leads to neutral mesons that subsequently decay to two highly boosted photons. The signatures of these two photons may overlap within the resolution of the calorimeter.

The fake rate method measures, in data, the relative rates of two types of jet fragmentation. The first type, the numerator, contains jets from quarks/gluons that fragment in such a way as to pass the photon identification. The denominator contains jets that pass criteria looser than the photon identification. The fake rate formula we used is defined below:

$$ f = \frac{J_{\gamma \text{LOOSE}}}{J_{\gamma \text{LOOSER}}} $$

where $J_{\gamma \text{LOOSE}}$ is the number of jets identified as photons and $J_{\gamma \text{LOOSER}}$ is the number of photon-like jets that pass criteria looser than the photon identification, and in addition failed at least one of the criteria typically associated with real photons. In particular, these are the CHIso, NHIso, and $\sigma_{i\eta\phi}$.

The numerator in eqn. (1) is defined by selection criteria that admit both real and "fake photons". The real photon contamination must then be subtracted because our definition from the fake rate involves only jets.

To subtract the real photon contamination, we consider a photon ID variable which is sensitive to the difference between real and fake photons. In this study, we consider $\sigma_{i\eta\phi}$. We extract the fake photon templates, in data, by applying the criteria from Table 1 except for $\sigma_{i\eta\phi}$ and in addition, use a sideband in Charged Hadron Isolation (5 GeV < CHIso < 10 GeV). This sideband choice was designed to minimize the fraction of real photons in the fake templates. We also prepare the real photon templates using MC Pythia Born/Box samples. The real MC $\sigma_{i\eta\phi}$ templates have been found to have a significant disagreement between data and Monte-Carlo in other analyses. To overcome this discrepancy, we rescale as follows [4]

$$ \sigma_{i\eta\phi} \rightarrow 0.00003 + 0.99470 \times \sigma_{i\eta\phi}. $$

Once these are prepared, we then fit the numerator $\sigma_{i\eta\phi}$ distribution with these two templates to determine the fraction of real photons.
Table 1: Loose Photon ID

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Level Trigger</td>
<td>HLTDoublePhoton70</td>
</tr>
<tr>
<td>$p_t$</td>
<td>&gt;80GeV</td>
</tr>
<tr>
<td>Encap Pseudorapidity</td>
<td>Both photons $1.56 &lt;</td>
</tr>
<tr>
<td>Had/Em</td>
<td>Single tower Had/Em&lt;0.05</td>
</tr>
<tr>
<td>Rho-corrected charged hadron isolation (CHIso)</td>
<td>&lt;2.3</td>
</tr>
<tr>
<td>Rho-corrected neutral hadron Isolation (NHIso)</td>
<td>$&lt;2.9 + 0.04x p_t$</td>
</tr>
<tr>
<td>Shower Shaper</td>
<td>$\sigma_{\eta\eta} &lt; 0.034$</td>
</tr>
<tr>
<td>Electron Veto</td>
<td>Conversion-safe electron veto</td>
</tr>
</tbody>
</table>

after once again applying the cut on $\sigma_{\eta\eta}$ from Table 1.

The template fitting is done in different $p_t$ bins since templates are typically $p_t$-dependent. We use RooFit to fit and to extract the fraction of real photons in each $p_t$ bin. We then calculate the fake rate in each bin. The overall fake rate as a function of $p_t$ is fitted using the following function

$$f(x) = p_0 + \frac{p_1}{x^{p_2}}.$$ (3)

IV. Results

Fig. 1a presents the fake rate as a function of the transverse momentum of the photon candidate for a sideband choice between 5 and 10 GeV.

We have also examined the effects of different sideband choices on the shapes of the fake templates (Fig.1d). The shapes of the normalized templates show no significant difference from one another. The purpose of this examination is to determine the systematic uncertainty on the photon fake rate. The subtraction of the real photon contamination is the major source of the systematic uncertainty [1]. We choose the $5 < CHIso(GeV) < 10$ to be our default template sideband definition. The fake rates calculated per $p_T$ bin are summarized in Table 2.

<table>
<thead>
<tr>
<th>$p_t$ bin (GeV)</th>
<th>Fake Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-40</td>
<td>24.23 ± 0.16</td>
</tr>
<tr>
<td>40-60</td>
<td>16.10 ± 0.13</td>
</tr>
<tr>
<td>60-80</td>
<td>13.63 ± 0.14</td>
</tr>
<tr>
<td>80-100</td>
<td>13.18 ± 0.16</td>
</tr>
<tr>
<td>100-120</td>
<td>12.60 ± 0.18</td>
</tr>
<tr>
<td>120-200</td>
<td>12.17 ± 0.12</td>
</tr>
<tr>
<td>200-600</td>
<td>13.18 ± 0.13</td>
</tr>
</tbody>
</table>

V. Conclusion

Endcap photon fake rates are higher than the barrel fake rates that were previously calculated (Figure 1a - 1b). The fake rate goes from $12 - 13\%$ at $p_T$ range: $60 - 600GeV$ and it goes as high as $24\%$ at $p_T : 20 - 40GeV$. Different CHIso sideband definitions have also very little effect on the shape of the fake templates as well as the fake rates per $p_T$ bin.

VI. Appendix: Software Tools and Libraries

C++ macros were used with ROOT to implement the cuts required for real and fake templates as well as the denominator objects. The RooFit library was used out-of-the-box to do the template fits. The real photon templates
were prepared from Monte Carlo simulations using Pythia Born/Box samples.

VII. Acknowledgements

This Summer Studentship was fully supported by the CERN Summer Student program for Non-member States. The author would like to thank Otman Charaf for the two-month supervision, for the invaluable help in setting up all the required software and necessary files, as well as the great insights in physics, statistics and the art of coding. He has lent a great amount of time amidst his busy schedule, in assisting the author in making the most out of her two months here in CERN. He has taught a lot including technical details and the overall physics of the problem at hand. The author would also like to thank Conor Henderson for his co-supervision. His leadership in the weekly Alabama meetings gave clear directions on what this piece of work was for in the overall-scheme of things. The way he described the bigger picture has given more motivation for this work to get done. The author would also like to thank Andre David for his piercing physics questions during personal meetings. That made the author address very important loopholes in her overall understanding of this project. To Michael Andrews for helping the author go through the ROOT documentation to prepare templates in the early stages of the project. Finally, the author would like to thank Joey Magpantay for his physics discussions back in the Philippines that has made this CERN Summer Studentship possible.

References

[1] Charaf, O., Henderson, C. et al. (2015). Search for High-Mass Diphoton Resonances in pp Collisions at $\sqrt{s}$ TeV with the CMS Detector CMS AN-12-305 (For CMS internal use and distribution only).
a) EndCaps Fake Rate vs $p_T$. Eqn. 3 was fitted up to $p_T = 200$ GeV only so that the low $p_T$ regions are fitted correctly.

b) Fake Rate vs $p_T$ in the Barrel. All bins have been included in the fit.

c) Fake Rates for different CHIso Sidebands.

d) Overlayed Fake Templates ($p_T : 100 – 120$ GeV). There is no significant difference in the shape of the normalized fake templates in different CHIso Sidebands.

Figure 1: Fake Rates Comparison
Figure 2: Template Fit Results. RooFit was used to fit the templates to the data. For the Endcaps, the background peak does not exactly coincide with the signal. For the same $p_T$ bin, the fits look similar.