Modern Physics Letters A

IMPLEMENTATION OF THE CMS-EXO-17-030 ANALYSIS IN THE MADANALYSIS 5 FRAMEWORK

Yechan Kang

Department of Physics, University of Seoul, 13-3, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Republic of Korea

Jihun Kim

Department of Physics and Astronomy, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea

Jin Choi

Department of Physics and Astronomy, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea

Soohyun Yun

Department of Physics, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

Received (Day Month Year)
Revised (Day Month Year)

We present the MadAnalysis 5 implementation and validation of the CMS-EXO-17-030 search. The search targets pair-produced resonances, each of which decaying into three jets. The results are interpreted within an R-parity violating supersymmetric (RPV SUSY) model, that predicts that pair-produced gluinos decay into three jets. This leads to a six-jet event. For this study, proton-proton collision data which was collected with the CMS detector in 2016 at a center of energy of 13 TeV is used, with a corresponding luminosity of 35.9 fb^{-1}. In the search, the resonance mass is expected to range from 200 GeV to 2000 GeV so that the analysis comprises four signal regions (SRs). To validate the results, we have selected four gluino benchmark masses of 200 GeV, 500 GeV, 900 GeV, and 1600 GeV, each of which being representative of a given signal region (that are denoted SR1, SR2, SR3, and SR4). We have simulated signal events and calculated the signal acceptance within the MadAnalysis 5 framework in each signal region. To validate the recast, our predicted acceptances have been compared with the official values for those benchmark scenarios. An agreement at the level of about 10% has been obtained.

1. Introduction

Events associated with a multijet final state at hadron colliders provide a unique window to investigate various beyond standard model (BSM) physics. Typically, in the Standard Model, pair-produced heavy resonances each decaying into three jets

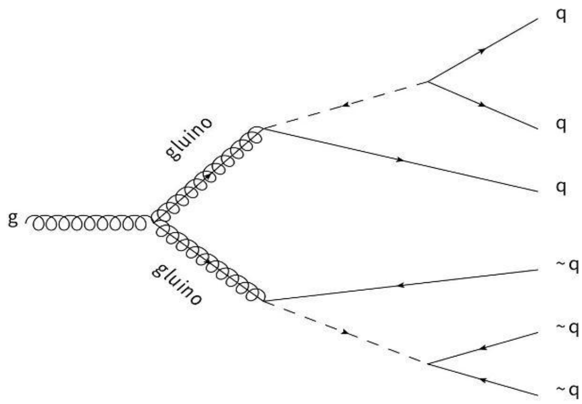

Fig. 1. Feynman diagram representative of pair-produced gluinos decaying into six jets.

only originate from the production of a pair of hadronically decaying top quarks. Therefore, if a particle heavier than the top quark exists, and manifests itself as a narrow resonance, then one should be able to see a clean high mass resonance peak in multijet invariant mass distributions.

We present the results of the recast of the CMS-EXO-17-030 three-jet analysis [1] which targets pair-produced resonances in proton-proton (pp) collisions, in a case where each resonance decays into three quarks. In this search, the RPV SUSY model [2] is used as a benchmark, with a varying gluino mass. This allows for the modeling of high mass resonances pair production, followed by subsequent gluino decays into three jets. Moreover, this leads to a final state comprising six quarks at the parton level. In this model, a new quantum number R is defined as

$$R = (-1)^{2S+3B+L},$$

where S is the spin, B is the baryon number, and L is the lepton number. In this search, we consider a model in which R-parity is broken via baryon number violation, so that squarks can decay into two quarks (Fig. 1). For our recast implementation and its validation, we follow the interpretation of the experimental analysis and the resonance is assumed to be a gluino.

The analysis is divided into four separate regions depending on the mass of the gluino. It exploits the geometrical event topology to discriminate signal events from background events. In order to improve the sensitivity to a wide range of resonance masses, the analysis includes signal regions that are each dedicated to a specific resonance mass, the associated topology and kinematics of the final-state jet activity. This separation is further necessary to manage the estimation of the background properly. In the low mass regions, the main background comes from top quark decays, whereas it comes from QCD events for the high mass regions. By defining different signal regions depending on the gluino mass, we can handle the background properly with different strategies. To perform the validation of our implementation, we select four benchmark gluino mass points representing each signal region, the gluino mass being respectively set to 200 GeV, 500 GeV, 900 GeV, and 1600 GeV. This enables the direct comparison between the recast and the result of the experimental publication in terms of acceptance and therefore allows us to validate our implementation.

In the rest of this note, we present the recast of the CMS-EXO-17-030 analysis in the MadAnalysis 5 framework [3–6], which is now available from the MadAnalysis 5 Public Analysis Database and the MadAnalysis 5 dataverse [7].

2. Description of the analysis

To identify pair-produced high mass resonances decaying into multiple jets in LHC events, the jet ensemble technique [8] is applied. This examines all possible combinatorial triplets that could be formed from a jet collection in each event. As a concrete example, we consider an event including 6 jets. First, we collect every possible set of 3 jets into a triplet. There should be 20 combinations of such triplets, and therefore 10 pairs of triplets in each event. All such triplet pairs and triplets are candidates for pair-produced gluinos and their decay. Then, to discriminate the 'correct' triplets (which originate from gluino decays) from wrongly combined triplets, and to reject the QCD background as well, we apply cuts on variables that embed the topology expected from the signal events. The cuts are categorized into three stages and applied step by step: event level, triplet pair level, and triplet level. The definition of each variable and the motivation to use them are described in section 2.2 in detail.

2.1. *Object definitions*

Jet candidates are reconstructed using the anti-k_T algorithm [9] with a radius parameter $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$. Jets in the detector are required to have a transverse momentum, p_T, larger than 20 GeV and an absolute value of the pseudorapidity, $|\eta|$, of at most 2.4. This analysis neither considers nor vetoes the presence of other objects like hard leptons or photons, so that their precise definition is irrelevant.

2.2. *Event selection*

Four separate signal regions have been defined to target all possible gluino masses in the range of 200 - 2000 GeV: SR1 (200-400 GeV), SR2 (400-700 GeV), SR3 (700-1200 GeV), and SR4 (1200-2000 GeV). The requirements in each signal region are described below.

First of all, each event is required to contain at least six reconstructed jets. From the entire set of jets, only the six jets with the highest p_T are considered. Then four selections based on event-level variables are applied. For the low mass regions targeting gluino masses below 700 GeV, all jets in the event must have a p_T larger than 30 GeV and the H_T variable, defined as the scalar sum of the p_Ts of all jets, is imposed to be larger than 650 GeV. For the high-mass regions dedicated to gluino masses beyond 700 GeV, the p_T of all jets must be larger than 50 GeV and the H_T variable must be greater than 900 GeV. Jets are arranged in descending order of p_T, and the p_T of the sixth jet is required to be larger than 40 GeV, 50 GeV, 125 GeV, or 175 GeV for the SR1, SR2, SR3, and SR4 signal region respectively.

To discriminate the signal from the QCD main background and wrongly combined triplets, Dalitz variables are adopted. Dalitz variables are effective discriminants for studying three-body decays. They were initially introduced by Dalitz in kaon to three pions decays [10]. The Dalitz variables for a triplet are defined as

$$\hat{m}(3,2)^2_{ij} = \frac{m^2_{ij}}{m^2_{ijk} + m^2_i + m^2_j + m^2_k},$$

where m_i, m_{ij} and m_{ijk} are respectively the invariant mass of the individual jet j_i, of the dijet system made of the jets j_i and j_j, and of the triplet. Here, indices refer to the jets in the triplet, where $i, j, k \in \{1, 2, 3\}$. These variables have good discriminating power as follows from our signal topology. In signal events for which a massive particle decays into three quarks, the angular distribution of the jets should be even in the center-of-mass frame. Therefore we expect the Dalitz variable to be close to 1/3 for each jet pair (m_{ij}).

By utilizing the above property of Dalitz variables, a new variable called the mass distance squared of a triplet is defined as

$$D^2_{[3,2]} = \sum_{i>j} \left(\hat{m}(3,2)_{ij} - \frac{1}{\sqrt{3}} \right)^2 .$$

This variable must be close to zero for symmetrically decaying signal triplets but deviates from zero for wrongly combined triplets and QCD backgrounds which may exhibit an asymmetric topology.

A generalized Dalitz variable is introduced as an extension of the original Dalitz variable for a six-jet topology, which should be close to 1/20 in the case of even angular distributions. It is defined from the normalized invariant mass of jet triplets,

$$\hat{m}(6,3)^2_{ijk} = \frac{m^2_{ijk}}{4m^2_{ijklmn} + 6\Sigma_i m^2_i}.$$

Here, m_{ijklmn} refers to the invariant mass of the leading six jets, where $i, j, k, l, m, n \in \{1, 2, 3, 4, 5, 6\}$.

Using the generalized Dalitz variables and the $D^2_{[3,2]}$ value associated with a triplet, the six-jet distance squared of an event is defined as

$$D^2_{[(6,3)+(3,2)]} = \sum_{i<j<k} \left(\sqrt{\hat{m}(6,3)^2_{ijk} + D^2_{[3,2]ijk}} - \frac{1}{\sqrt{20}} \right)^2 .$$

For signal events, each pair-produced gluino is expected to decay symmetrically, which leads to small values of $D^2_{[3,2]}$. Furthermore, each generalized Dalitz variable ($m(6,3)^2_{ijk}$) is expected to be close to 1/20. Therefore, signal events are likely to

Table 1. Selection criteria

Region	Gluino Mass	Events				Triplet Pairs	Triplets	
		Jet p_T	H_T	$p_T(j_6)$	$D^2_{[(6,3)+(3,2)]}$	A_m	Δ	$D^2_{[3,2]}$
1	200-400 GeV	>30 GeV	>650 GeV	>40 GeV	<1.25	<0.25	>250 GeV	<0.05
2	400-700 GeV	>30 GeV	>650 GeV	>50 GeV	<1.00	<0.175	>180 GeV	<0.175
3	700-1200 GeV	>50 GeV	>900 GeV	>125 GeV	<0.9	<0.15	>20 GeV	<0.2
4	1200-2000 GeV	>50 GeV	>900 GeV	>175 GeV	<0.75	<0.15	>-120 GeV	<0.25

feature $D^2_{[(6,3)+(3,2)]}$ close to zero. On the other hand, the events containing triplets originating from QCD multijet production will have an asymmetric angular distribution, and thus have values relatively far from zero. The official analysis has shown that the distribution of $D^2_{[(6,3)+(3,2)]}$ for QCD multijet events peaks at a farther point than the gluino events, as expected. The $D^2_{[(6,3)+(3,2)]}$ variable is used for the last selection at the event level and is required to be smaller than 1.25, 1.00, 0.9, or 0.75 for the SR1, SR2, SR3, and SR4 signal region respectively.

Furthermore, the masses of two distinct triplets are expected to be symmetric in the case of the signal, as originating from the decay of the same particle. Thus the mass asymmetry defined as

$$A_m = \frac{|m_{ijk} - m_{lmn}|}{m_{ijk} + m_{lmn}},$$

where m_{ijk} and m_{lmn} are the masses of the two distinct triplets in a triplet pair, is expected to be closer to zero for the correctly combined triplet pairs in the signal case. The mass asymmetry of a triplet pair is required to be smaller than 0.25 or 0.175 for the SR1 and SR2, or 0.15 for the SR3 and SR4.

Finally, selections at the triplet-level are applied. The variable Δ of a triplet is defined as the sum of the p_T of the jets in the triplet ($|p_T|_{ijk}$), after subtracting the triplet invariant mass (m_{ijk}):

$$\Delta = |p_T|_{ijk} - m_{ijk}.$$

In the official analysis, it has been shown that correctly combined triplets have a constant distribution in the mass vs p_T plane, whereas in cases of wrongly combined triplets and QCD backgrounds their p_T and mass are proportional to each other. Therefore the Δ observable has good discriminating power between wrongly combined triplets, QCD backgrounds and correctly combined triplets. This value is required to be larger than 250 GeV, 180 GeV, 20 GeV, or -120 GeV for the SR1, SR2, SR3, and SR4 region respectively. For the very last selection, the mass distance squared of a triplet ($D^2_{[3,2]}$) is required to be smaller than 0.05, 0.175, 0.2, or 0.25 for each region.

The actual cuts for each variable for the event, triplet pair, and triplet levels are summarized in Table 1.

3. Validation

3.1. *Event generation*

Simulation of double-trijet resonance events is done by making use of the MadGraph5_aMC@NLO version 2.7.3 Monte Carlo generator [11], using the RPVMSSM_UFO model file [12, 13]. For the parton distribution functions, the LO set of NNPDF3.0 [14] parton densities with $\alpha_s = 0.130$, as implemented in LHAPDF6 [15], is used. To avoid any squark contribution to gluino production, all the masses of squarks are set to be 2.5 TeV, and the masses of gluinos are set to be 200, 500, 900, and 1600 GeV to target the signal regions resulting from the cuts described in section 2.2. Based on the pair production of gluinos, we used MadSpin [16] and MadWidth [17] without spin correlations to simulate the gluino decays into three jets. We compared the acceptance resulting from the cuts described in the next section, using signal samples with and without spin correlation, and found that there is negligible difference in the final acceptance. Here, we thus present the results without any spin correlation.

After the simulation of the hard-scattering process, Pythia8 [18] is used for parton showering and hadronization, followed by Delphes3 [19] for the fast simulation of the CMS detector response.

3.2. *Comparison with the official results*

As using combined triplets of jets for the final selection, the analysis suffers from two major backgrounds, irreducible QCD backgrounds and a unique background not originating from a specific physical process: wrongly combined triplets. Since the invariant mass distribution is similar for QCD backgrounds and wrongly combined triplets [1], the CMS collaboration made signal and background fitting templates from those distributions and proceed with signal to background fitting directly to the data to calculate the final signal significance. Therefore, the number of triplets that pass all cuts is used indirectly for the final result. To see how many correct triplets survive in each signal region, the signal acceptance has been defined based on the triplet selection described in section 2:

$$\text{Acc.} = \frac{\text{Number of surviving triplets}}{\text{Number of generated events}}.$$

Here, the acceptance is defined as the ratio of the number of triplets and the number of total events, and not the number of events passing the selections and the total number of events. We hence collect all possible combinations of triplets out of 6 jets and have 20 triplets (or 10 triplet pairs) per single event. In this analysis, we have cuts at the event level, triplet-pair level, and triplet level.

Since the analysis has a distinctive definition of acceptance based on the number of triplets, one of the major difficulties in using the MadAnalysis 5 framework was the implementation of counting the triplets passing the different cuts in each signal region, as there are diverse triplet-level cut thresholds for each region. In

MadAnalysis 5, the framework provides cutflows based on the event selection, which makes it hard to count the number of surviving triplets in each signal region. To overcome this problem, we made four collections of triplets, *i.e.* one for each signal region, and updated each collection with the different cuts for each signal region. Finally, we multiplied each event weight by the number of triplets (for each region), which makes MadAnalysis 5 generating cutflows on triplet level.

The acceptance numbers officially calculated by CMS are of 0.00024, 0.084, 0.17 for SR1, SR3, and SR4. There is no result provided for SR2. For the purpose of recast, we define the difference as

$$\text{Diff.} = \frac{\text{Acc.(recast)} - \text{Acc.(CMS official)}}{\text{Acc.(CMS official)}}$$

to compare the recast values with the official results.

Comparing with the official results, the recast showed a large discrepancy. Acceptances (differences) we calculated are $6.25 \times 10^{-4}(140\%), 6.5 \times 10^{-1}(674\%)$ and $1.71(906\%)$ for SR1, SR3 and SR4. We found out that many wrongly combined triplets not originating from the same gluino still pass the final selection. Since there is no way to calculate the acceptance of the correctly matched triplets as originally performed through the template fit to the CMS data, we chose an alternative approach, using generator level information to check how many triplets from the same gluino can survive after all cuts. Therefore, we require that the correct triplets should be matched to their mother gluino as

- All jets should be matched to generator level partons within a distance in the transverse plane of $\Delta R(j, q) < 0.3$, where q generically stands for u, c, d, s and the corresponding antiparticles.
- Matched partons in a triplet should all be quarks, or all be antiquarks.
- All matched (anti)quarks in the triplet should have the same gluino as their mother.

Here, we required the jets to be matched to their mother gluino using the truth level information. For the purpose of generalization, any recasting analysis that wishes to use the truth information should change the Particle Data Group identifier (PID) of the mother particle. We defined the PID of this mother particle by using the #DEFINE preprocessor method, so the user can change the value of the EXO_17_030_PID variable to any other value relevant for the signal of their interest. Therefore, this implementation can be further tested with various other BSM models that allow resonance with three jet decay signature, *e.g.* searches based on composite quark model [20] or extra dimensional model [21].

The final acceptances that we obtain, for the considered benchmark scenarios, are of 2.8×10^{-4}, 7.3×10^{-2}, and 1.55×10^{-1} for the SR1, SR3 and SR4 regions. Our predictions show good agreements with the CMS official results, at the level of 8%, 13%, and 8.8% for the SR1, SR3 and SR4 regions. For the SR2 region, the final acceptance is 1.5×10^{-2}. This value has no comparison target because the official

Table 2. Cutflows in the Low-Mass Regions. The initial number of triplets that could be reconstructed from each event is assumed to be 20. All triplets are matched to their mother particle. Since there is no official CMS result for SR2, we did not calculate the difference for that region.

	Signal Region 1		Signal Region 2	
Cut	Events	Triplets	Events	Triplets
Initial events	400,000	8,000,000	400,000	8,000,000
Njets≥6	231,863	4,637,261	367,491	7,349,821
preselection	148,090	2,961,800	341,054	6,821,079
HT	38,434	768,680	329,561	6,591,218
Sixth jet p_T	29,611	592,220	242,511	4,850,220
$D^2_{[(6,3)+(3,2)]}$	23,296	465,920	186,731	3,734,618
A_m	3,982	4,630	89,853	118,285
Δ	187	199	5,534	6,501
$D^2_{[3,2]}$	108	112	5,145	5,995
Acc.		0.028%		1.50%
Acc.(CMS official)		0.026%		
Diff.		8%		

Table 3. Cutflows in the High Mass Region. The initial number of triplets reconstructed from each event is assumed to be 20. All triplets are matched to their mother particle.

	Signal Region 3		Signal Region 4	
Cut	Events	Triplets	Events	Triplets
Initial events	400,000	8,000,000	400,000	8,000,000
Njets≥6	388,119	7,762,382	394,516	7,890,321
preselection	340,320	6,806,404	373,669	7,473,380
HT	339,303	6,786,064	373,661	7,473,221
Sixth jet p_T	120,141	2,402,821	166,877	3,337,540
$D^2_{[(6,3)+(3,2)]}$	100,349	2,006,981	113,436	2,268,721
A_m	52,205	72,806	69,080	100,637
Δ	25,465	31,320	49,767	62,731
$D^2_{[3,2]}$	23,948	29,025	49,309	61,959
Acc.		7.3%		15.5%
Acc.(CMS official)		8.4%		17.0%
Diff.		-13%		-8.8%

acceptance for the SR2 region has not been provided by the CMS collaboration. Detailed results are provided in tables 2 and 3.

4. Conclusion

A recast of the CMS-EXO-17-030 double-three-jet analysis has been performed within the MadAnalysis 5 framework. To validate our implementation, we choose four gluino RPV SUSY scenario with masses ranging from 200 to 2000 GeV. The four masses that we selected are 200 GeV, 500 GeV, 900 GeV, and 1600 GeV, and represent each signal region. In this note, the event selection is described in detail, and corresponding cutflows for each benchmark point are presented. We exhibit the difficulties that are inherent to the usage of MadAnalysis 5 for the CMS-EXO-17-030 recast, as non-event based acceptance calculations are in order. We moreover explain our method to overcome them. The signal events are simulated under the same condition as for the official CMS result, which corresponds to an integrated luminosity of 35.9 fb^{-1} of collisions at a center-of-mass energy of 13 TeV, but with a CMS detector configuration based on Delphes 3. The validation is performed in terms of the acceptance for each signal region. The recast and the official results show good agreement, resulting in differences from a minimum of 8% to a maximum of 13%.

The code is available online from the MadAnalysis 5 dataverse [7], at https://doi.org/10.14428/DVN/GAZACQ, on which we also provide cards that were relevant for the validation of this implementation.

Acknowledgments

We congratulate all our colleagues who have participated in the second MadAnalysis 5 workshop on LHC recasting in Korea and thank the organizers and the tutors of the workshop for their sincere support. In addition, Soohyun Yun is grateful for financial support from Hyundai Motor Chung Mong-Koo Foundation.

References

1. CMS Collaboration, A. M. Sirunyan *et al.*, *Phys. Rev. D* **99**, 012010 (2019), arXiv:1810.10092 [hep-ex].
2. R. Barbier *et al.*, *Phys. Rept.* **420**, 1 (2005), arXiv:hep-ph/0406039.
3. E. Conte, B. Fuks and G. Serret, *Comput. Phys. Commun.* **184**, 222 (2013), arXiv:1206.1599 [hep-ph].
4. E. Conte, B. Dumont, B. Fuks and C. Wymant, *Eur. Phys. J. C* **74**, 3103 (2014), arXiv:1405.3982 [hep-ph].
5. B. Dumont, B. Fuks, S. Kraml, S. Bein, G. Chalons, E. Conte, S. Kulkarni, D. Sengupta and C. Wymant, *Eur. Phys. J. C* **75**, 56 (2015), arXiv:1407.3278 [hep-ph].
6. E. Conte and B. Fuks, *Int. J. Mod. Phys. A* **33**, 1830027 (2018), arXiv:1808.00480 [hep-ph].
7. Y. Kang, J. Kim, J. Choi and S. Yun, *10.14428/DVN/GAZACQ* (2020).
8. CDF Collaboration, T. Aaltonen *et al.*, *Phys. Rev. Lett.* **107**, 042001 (2011), arXiv:1105.2815 [hep-ex].
9. M. Cacciari, G. P. Salam and G. Soyez, *JHEP* **04**, 063 (2008), arXiv:0802.1189 [hep-ph].
10. R. Dalitz, *Phys. Rev.* **94**, 1046 (1954).

11. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, *JHEP* **07**, 079 (2014), arXiv:1405.0301 [hep-ph].
12. C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *Comput. Phys. Commun.* **183**, 1201 (2012), arXiv:1108.2040 [hep-ph].
13. B. Fuks, *Int. J. Mod. Phys. A* **27**, 1230007 (2012), arXiv:1202.4769 [hep-ph].
14. NNPDF Collaboration, R. D. Ball *et al.*, *JHEP* **04**, 040 (2015), arXiv:1410.8849 [hep-ph].
15. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr and G. Watt, *Eur. Phys. J. C* **75**, 132 (2015), arXiv:1412.7420 [hep-ph].
16. P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *JHEP* **03**, 015 (2013), arXiv:1212.3460 [hep-ph].
17. J. Alwall, C. Duhr, B. Fuks, O. Mattelaer, D. G. Öztürk and C.-H. Shen, *Comput. Phys. Commun.* **197**, 312 (2015), arXiv:1402.1178 [hep-ph].
18. T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015), arXiv:1410.3012 [hep-ph].
19. DELPHES 3 Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens and M. Selvaggi, *JHEP* **02**, 057 (2014), arXiv:1307.6346 [hep-ex].
20. M. Redi, V. Sanz, M. de Vries and A. Weiler, *JHEP* **08**, 008 (2013), arXiv:1305.3818 [hep-ph].
21. K. S. Agashe, J. Collins, P. Du, S. Hong, D. Kim and R. K. Mishra, *JHEP* **05**, 078 (2017), arXiv:1612.00047 [hep-ph].