Presented at DIS2022: XXIX International Workshop on Deep-Inelastic Scattering and Related Subjects, Santiago de Compostela, Spain, May 2-6 2022

# Identification of *b*-jets using QCD-inspired observables <sup>a</sup>

Oleh Fedkevych<sup>1,†</sup>

<sup>1</sup>Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146, Italy

In this talk I consider the issue of separating hadronic jets that contain bottom quarks (*b*-jets) from jets featuring light partons only. I discuss a recently proposed *b*-tagging approach that exploits the application of the QCD-inspired jet substructure observables such as one-dimensional jet angularities and the two-dimensional primary Lund plane and demonstrate that these quantities can be used as inputs to modern machine-learning algorithms to efficiently separate *b*-jets from light ones.

#### Introduction

In the last few years several measurement campaigns dedicated to the jet substructure physics were performed at the LHC. In particular, so-called jet angularities [4–6], as well as the primary Lund Plane (pLP) [7], were measured by the ATLAS [8, 9], CMS [10] and ALICE [11] collaborations. On the other hand, these observables have been the target of detailed theoretical investigation performed by means of the Monte Carlo (MC) and resummation techniques see, for example, Refs. [2, 3, 12–17]. Whereas the comparison of the collected data against theoretical predictions clearly indicates the necessity to improve the exiting tools, the fair description of the observed data in certain domains of the phase space was achieved (see, for example, Refs. [3, 8–11]). Therefore, it is interesting to investigate various possible applications of the collected high-quality data. For example, one possible direction to go would be the application of the jet angularities and pLP to the tagging purposes to investigate a flavor of a particle seeding a given hadronic jet. Various tagging applications of jet substructure observables exist in the literature: for example, in Ref. [18] it was proposed to use jet angularities as quark-gluon taggers to increase the initial-state gluon contribution to the Z + jet production which could allow to probe the gluonic degrees of freedom of the colliding protons. Also, the pLP already found many applications, *e.g.* in the context of W boson, top quark, quark-gluon and Higgs boson tagging [7, 17, 19–23].

In this talk I discuss how one can use jet angularities and pLP to tell *b*-jets from jets produced by light flavors which is crucial for studies of the Higgs boson properties [24-27], measurements of Standard Model (SM) processes [28-30], and searches for beyond SM (BSM) signals [31-33]. In particular, I demonstrate how one can use jet angularities and pLP as inputs to train a deep neural network (DNN) and a convolutional neural network (CNN) to achieve a discrimination power comparable with the state-of-the-art taggers used by the ATLAS collaboration [34].

### Definition of observables and the MC setup

Let me start with a definition of the observables I use through this talk. For the jet of the radius R the jet angularities are defined as

$$\lambda_{\alpha}^{\kappa} = \sum_{i \in jet} \left( \frac{p_{\mathrm{T,i}}}{\sum_{j \in jet} p_{\mathrm{T,j}}} \right)^{\kappa} \left( \frac{\Delta_i}{R} \right)^{\alpha}, \text{ where } \Delta_i = \sqrt{(y_i - y_{jet})^2 + (\phi_i - \phi_{jet})^2} \tag{1}$$

and the sum  $\sum_{i \in jet}$  runs over all jet constituents. The requirement of infrared and collinear safety implies  $\kappa = 1$  and  $\alpha > 0$ . Here, I consider three commonly used cases namely,  $\lambda_{1/2}^1$  (LHA),  $\lambda_1^1$  (Jet Width) and  $\lambda_2^1$  (Jet Thrust). To keep the notation simple, I skip the  $\kappa$  label in the rest of this talk.

In Fig. 1 I present the comparison between recent CMS measurements of the Jet Thrust [10] compared against theoretical predictions from Refs. [2, 3] for the high- $p_T$  jets before (left) and after (right) application of the SoftDrop grooming procedure [35] to reduce the impact of non-perturbative corrections. As one can see, the theoretical predictions

<sup>&</sup>lt;sup>a</sup> Based upon results of Ref. [1] and, partially, of Ref. [2, 3].

give a fairly well description of the experimental data in the whole range of binning which can be considered as a motivation to use jet angularities as an input for the machine learning algorithms.



FIG. 1. Comparison against the recent CMS measurements [10] for the ungroomed (left) and groomed (right) Jet Thrust distributions. The theoretical predictions are taken from Refs. [2, 3].

To construct the pLP one first recluster a selected jet using the Cambridge-Aachen (C/A) algorithm [36, 37] and then follow the hardest branch of the declustering tree and evaluate at each splitting:

$$\Delta_{ab} \equiv \sqrt{\left(y_a - y_b\right)^2 + \left(\phi_a - \phi_b\right)^2}, \quad k_t \equiv p_{\mathrm{T}b} \,\Delta_{ab},\tag{2}$$

where  $y_{a,b}$ ,  $\phi_{a,b}$  are rapidity and azimuthal angle of subjets a and b correspondingly.

To produce distributions of  $\lambda_{1/2}$ ,  $\lambda_1$ ,  $\lambda_2$  and pLP were used the leading-order (LO) MC event simulations interfaced to a parton-shower algorithm and followed by parton hadronization. Two alternative MC setups were considered to produce the pseudo data to test the tagging algorithms: the first setup is given by the PYTHIA event generator [38] using the NNPDF2.3 LO set [39] of proton Parton Distribution Functions (PDFs) in the five flavor number scheme (FNS) and the transverse-momentum ordered parton shower (PS) [40]. The second MC setup is represented by the HERWIG event generator [41, 42] using the CT14 LO PDF set [43] in the five FNS as well. The PS emissions in HERWIG are simulated according to the coherent branching algorithm from Ref. [44]. Both PYTHIA and HERWIG are using their default tunes [42, 45].

The PS effects due to the mass of emitting quarks are taken into account by PYTHIA and HERWIG in different ways [46, 47]. Also, the two programs have different approaches to simulation of non-perturbative contributions due to the underlying event (UE) [48–50] and hadronization [51–53]. Therefore, here PYTHIA is used as the main MC tool and HERWIG is used to check the stability of the obtained results with respect to different models of PS and non-perturbative effects.

In this talk I consider a typical setup for measurements of a Z boson produced in association with hadronic high $p_{\rm T}$  jets, similarly to the recent  $\sqrt{s} = 13$  TeV CMS measurements [10] of jet angularities with slightly adjusted leading jet cuts (rapidity  $|y_{\rm jet}| < 2.5$  and transverse momentum  $p_{\rm T,jet} > 500$  GeV) to be consistent with the ATLAS selection cuts from Ref. [34], which is used as a benchmark for comparison of the tagging performance. The event selection and analysis were performed with the help of the RIVET [54, 55] and FASTJET [56] libraries. The grooming is performed with the SoftDrop class from the fjcontrib library.

The flavor of each selected jet (b-, c- or light-jet) is assigned using the standard ATLAS approach [34]. More precisely, a jet is considered as a b-jet if it can be matched to at least one weakly decaying b-hadron with  $p_T \ge 5$  GeV located within a  $\Delta R = 0.3$  cone around the jet axis; if no b-hadrons are found, then the same selection procedure is applied to check if a jet can be matched to at least one c-hadron. A jet matched to a b-hadron or a c-hadron is labelled as a b-jet or c-jet, respectively. After assigning b- and c-jet labels the remaining jets are labelled as the light-jets. Since it is a common approach (see, for example, Refs. [34, 57]) to test the quality of b-tagging algorithms comparing

3



FIG. 2. An example of jet substructure observables used for *b*-tagging. The top row shows for hadron-level jets: (a) the histogram of the LHA angularity  $\lambda_{1/2}$ , including the underflow events in the first bin; the pLP for *b*-jets (b) and light-jets (c). The bottom row shows for hadron-level jets after the application of the SoftDrop (SD) grooming algorithm [35] with  $\beta = 0$  and  $z_{\text{cut}} = 0.1$  parameters: (d) the histogram of the LHA angularity  $\lambda_{1/2}$ , including the underflow events in the first bin; the pLP for *b*-jets (e) and light-jets (f).

b-jet vs. light-jet discrimination, the jets carrying the c-labels are discarded from the produced MC samples. In this talk two 100k samples of b- and light-jets events are considered.

Finally, I would like to note that here the jet angularities and pLP are calculated for the jet before and after application of the SoftDrop groomer with  $\beta = 0$  and  $z_{\text{cut}} = 0.1$  parameters. By applying the SoftDrop groomer one can reduce the sensitivity of the observables to soft radiation and hence make the groomed observables more resilient against difficult to model non-perturbative effects. In the following, the SD label is used to distinguish between groomed and ungroomed jets. Also, to avoid complications due to the possible bin migration caused by SoftDrop the value of the transverse momentum of a groomed jet is always referring to the transverse momentum of a jet *before* application of the SoftDrop grooming.

### Results

In the previous section I provided the definitions of the jet substructure observables and of the MC setup used in this study. The corresponding differential distributions are shown in Fig. 2 (note that unlike Fig. 1 here different binning is used for the jet angularities). Namely, in Figs. 2 (a) and (d) I compare the LHA distribution for *b*- and light-jet samples produced with PYTHIA, before and after jet grooming. The *b*-jet flavor discrimination performance is estimated using the receiver operating characteristic (ROC) curves in the inserts of Figs. 2 (a) and (d). In this case, the ROC curve is computed by plotting the efficiency of the signal selection  $\varepsilon_B$  versus the efficiency of the background

selection  $\varepsilon_L$  calculated according to:

$$\varepsilon_{B/L} = \frac{1}{N_{B/L}} \int_{\lambda_{\text{cut}}}^{1} \frac{dN_{B/L}}{d\lambda} d\lambda.$$
 (3)

The efficiency of the tested algorithm is represented by the area under the ROC curve (AUC): it is equal to one for the ideal discrimination between signal and background events and is equal to one half if there is the same probability to pick a signal or background event. The complete set of ROC curves and corresponding AUC values for  $\lambda_{1/2}$ ,  $\lambda_1$  and  $\lambda_2$  angularities is given in Fig. 3 (see the blue, green and magenta lines). In case of the PYTHIA dataset the most efficient tagger based on a single jet angularity gives about 60% of the signal selection versus about 40% of the background selection corresponding to the AUC value of about 0.64. Similar efficiency is achieved after the application of the SoftDrop groomer. By comparing the shapes of the ROC curves corresponding to PYTHIA (left) and HERWIG (right) MC we see that the obtained results demonstrate a good stability with respect to change of the UE and hadronization models.

One can, however, improve the performance of the aforementioned cut-based approach by using a DNN which as an input takes values of  $\lambda_{1/2}$ ,  $\lambda_1$  and  $\lambda_2$  jet angularities. The DNN architecture considered in this talk consists out of two hidden layers each containing five nodes. The intermediate and output layers are using the "ReLu" and "SoftMax" activation functions, respectively. The DNN model is using the Adam optimizer [58] and the cross-entropy loss function. The data was split into the training, validation and test datasets each containing 60%, 20% and 20% of the initial 100k set correspondingly. The output score returned by the DNN was used to produce the ROC curves and evaluate the associated AUC values to estimate the tagging efficiency.

The results are shown in Fig. 3 where one can see a noticeable improvement in the performance as comparing to the taggers based upon single angularity distributions (see the red lines). For example, in case of the PYTHIA dataset the ROC reaches a signal selection efficiency of  $\varepsilon_B \simeq 64\%$  and a background selection efficiency of  $\varepsilon_L \simeq 40\%$ , the AUC reaches a value of 0.67, with almost no difference between ungroomed and groomed jets.



FIG. 3. ROC curves, obtained using PYTHIA (left) and HERWIG (right) MC data, for one-dimensional angularity distributions, multivariable DNN classifier and pLP CNN classifier. The single points correspond to ATLAS JetFitter, IP3D and DL1 *b*-tagging performance from [34].

Now, let me consider the *b*-tagging algorithm based on pLP. Since pLP is a two-dimensional observable, here it is more convenient to use a CNN, which is better-suited for image recognition purposes. The CNN is built out of four convolutional layers followed by the flat layer. First and second convolutional layers contain twenty and ten filters whereas the third and fourth layers have eight filters, respectively. The flat layer consist out of two hundred neurons. The CNN model is using the same activation function and optimizers as the DNN model.

The ROC curves and the corresponding AUC values for the CNN algorithm are given in Fig. 3, showing a relevant

improvement over the DNN results (see the black curves). For example, for the PYTHIA MC data the ROC curve is characterized by  $\varepsilon_B \simeq 70\%$ ,  $\varepsilon_L \simeq 40\%$  and the AUC value equal to 0.71 for the ungroomed jets (for the groomed jets the AUC value decrease by 0.02). Also, by comparing the CNN predictions obtained with the PYTHIA and HERWIG MC datasets one can see that the differences between corresponding AUC values are staving within the 5% range which give a confidence about the robustness of the proposed CNN taggers against different models of PS and non-perturbative effects.

Finally, the performance of the DNN and CNN discriminants is compared against the state-of-the-art b-tagging algorithms used by the ATLAS experiment. Here the two low-level b-tagging algorithms are compared against the DNN and CNN results. The fist algorithm is called JetFitter [59] and is based upon the reconstruction of the bto c-hadron decay chains, whereas the second one is called IP3D [60] and is using the analysis of the tracks of the charged particles. Additionally, the deep feed-forward neural network DL1 [60] (based upon the information obtained from the JetFitter and IP3D taggers plus some additional input based on jet and secondary vertex reconstruction) is considered. The scatter points in Fig. 3 showing the signal vs. background efficiency for the JetFitter, IP3D, and DL1 algorithms are taken from Ref. [61]. By comparing the performance of the JetFitter, IP3D and DL1 taggers against results presented in this talk one can see that the DNN using jet angularities show better performance than the IP3D tagger and somewhat worse performance than the JetFitter tagger. However, as one can see, the CNN discriminator improves the b-tagging performance and makes it comparable with the performance of the JetFitter algorithm. Finally, I would like to note that the DL1 tagger, which is trained upon multiple features, leads to a better performance than presented here DNN and CNN models. Nevertheless, one can argue that the set of simple input features considered in this talk, which rely on QCD phenomenological ideas and are not directly based on charged particle track reconstruction, can be used in in parallel with the aforementioned ATLAS taggers, to improve existing multivariate *b*-tagging algorithms.

#### Conclusions

A novel b-tagging approach based upon one-dimensional jet angularities and two-dimensional primary Lund plane was proposed. It was found that the DNN and CNN discriminators trained upon these observables for the high- $p_{\rm T}$ jets ( $p_{\rm T} \geq 500 \text{ GeV}$ ) reach accuracy similar to the JetFitter and IP3D b-tagging algorithms used by the ATLAS collaboration, but smaller than the accuracy of the DL1 tagger based upon combination of the aforementioned trackbased taggers with some additional features. However, the main advantage of the presented approach is given by its simplicity, since all information needed by the DNN and CNN discriminators is contained in the jet clustering history. Moreover, the considered here ATLAS taggers rely on the b-hadron lifetime whereas the jet substructure observables, as dictated by QCD, are sensitive to the kinematics of the jet constituents and their dynamics (in particular to the b-quark mass). Therefore, it would be very interesting to try to combine the input used to train the DNN and CNN from this talk with the one used by the DL1 tagger, in order to check if the further discriminating accuracy can be achieved. Such study is left to the future work.

## Acknowledgments

My work was supported by the Università di Genova under the curiosity-driven grant "Using jets to challenge the Standard Model of particle physics". The main results presented in this talk were obtained in Ref. [1] in collaboration with C. K. Khosa, S. Marzani and F. Sforza. Also in Fig. 1 I used the results from Refs. [2, 3] obtained in collaboration with S. Caletti, S. Marzani, D. Reichelt, S. Schumann, G. Sovez and V. Theeuwes.

<sup>&</sup>lt;sup>†</sup> oleh.fedkevych@ge.infn.it

<sup>[1]</sup> O. Fedkevych, C. K. Khosa, S. Marzani, and F. Sforza, (2022), arXiv:2202.05082 [hep-ph].

<sup>[2]</sup> S. Caletti, O. Fedkevych, S. Marzani, D. Reichelt, S. Schumann, G. Soyez, and V. Theeuwes, JHEP 07, 076, arXiv:2104.06920 [hep-ph].

<sup>[3]</sup> D. Reichelt, S. Caletti, O. Fedkevych, S. Marzani, S. Schumann, and G. Soyez, JHEP 03, 131, arXiv:2112.09545 [hep-ph]. [4] C. F. Berger, T. Kucs, and G. F. Sterman, Phys. Rev. D 68, 014012 (2003), arXiv:hep-ph/0303051.

<sup>[5]</sup> L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, Phys. Rev. D 79, 074017 (2009), arXiv:0807.0234 [hep-ph].

<sup>[6]</sup> A. J. Larkoski, J. Thaler, and W. J. Waalewijn, JHEP 11, 129, arXiv:1408.3122 [hep-ph].

<sup>[7]</sup> F. A. Dreyer, G. P. Salam, and G. Soyez, JHEP 12, 064, arXiv:1807.04758 [hep-ph].

<sup>[8]</sup> G. Aad et al. (ATLAS), Phys. Rev. D 101, 052007 (2020), arXiv:1912.09837 [hep-ex].

- [9] G. Aad et al. (ATLAS), Phys. Rev. Lett. 124, 222002 (2020), arXiv:2004.03540 [hep-ex].
- [10] A. Tumasyan *et al.* (CMS), JHEP **01**, 188, arXiv:2109.03340 [hep-ex].
- [11] S. Acharya et al. (ALICE), JHEP 05, 061, arXiv:2107.11303 [nucl-ex].
- [12] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig, and C. Lee, JHEP 11, 101, arXiv:1001.0014 [hep-ph].
- [13] A. Hornig, Y. Makris, and T. Mehen, JHEP 04, 097, arXiv:1601.01319 [hep-ph].
- [14] Z.-B. Kang, K. Lee, and F. Ringer, JHEP 04, 110, arXiv:1801.00790 [hep-ph].
- [15] Z.-B. Kang, K. Lee, X. Liu, and F. Ringer, Phys. Lett. B 793, 41 (2019), arXiv:1811.06983 [hep-ph].
- [16] A. Lifson, G. P. Salam, and G. Soyez, JHEP 10, 170, arXiv:2007.06578 [hep-ph].
- [17] F. Dreyer, G. Soyez, and A. Takacs, (2021), arXiv:2112.09140 [hep-ph].
- [18] S. Caletti, O. Fedkevych, S. Marzani, and D. Reichelt, Eur. Phys. J. C 81, 844 (2021), arXiv:2108.10024 [hep-ph].
- [19] F. A. Dreyer and H. Qu, JHEP 03, 052, arXiv:2012.08526 [hep-ph].
- [20] F. A. Dreyer, R. Grabarczyk, and P. F. Monni, (2022), arXiv:2203.06210 [hep-ph].
- [21] C. K. Khosa and S. Marzani, Phys. Rev. D **104**, 055043 (2021), arXiv:2105.03989 [hep-ph].
- [22] C. K. Khosa, in 50th International Symposium on Multiparticle Dynamics (2021) arXiv:2110.15135 [hep-ph].
- [23] L. Cavallini, A. Coccaro, C. K. Khosa, G. Manco, S. Marzani, F. Parodi, D. Rebuzzi, A. Rescia, and G. Stagnitto, Eur. Phys. J. C 82, 493 (2022), arXiv:2112.09650 [hep-ph].
- [24] G. Aad et al. (ATLAS), Eur. Phys. J. C 81, 178 (2021), arXiv:2007.02873 [hep-ex].
- [25] A. M. Sirunyan et al. (CMS), Phys. Rev. Lett. 121, 121801 (2018), arXiv:1808.08242 [hep-ex].
- [26] M. Aaboud et al. (ATLAS), Phys. Lett. B 784, 173 (2018), arXiv:1806.00425 [hep-ex].
- [27] A. M. Sirunyan et al. (CMS), Phys. Rev. Lett. 120, 231801 (2018), arXiv:1804.02610 [hep-ex].
- [28] G. Aad et al. (ATLAS), JHEP 07, 044, arXiv:2003.11960 [hep-ex].
- [29] G. Aad et al. (ATLAS), Phys. Lett. B 810, 135797 (2020), arXiv:2006.13076 [hep-ex].
- [30] A. M. Sirunyan et al. (CMS), Eur. Phys. J. C 79, 368 (2019), arXiv:1812.10505 [hep-ex].
- [31] R. Aaij et al. (LHCb), Nature Phys. 18, 277 (2022), arXiv:2103.11769 [hep-ex].
- [32] G. Aad et al. (ATLAS), JHEP 04, 174, arXiv:2012.03799 [hep-ex].
- [33] A. M. Sirunyan et al. (CMS), Eur. Phys. J. C 81, 3 (2021), arXiv:2008.05936 [hep-ex].
- [34] G. Aad et al. (ATLAS), Eur. Phys. J. C 79, 970 (2019), arXiv:1907.05120 [hep-ex].
- [35] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, JHEP 05, 146, arXiv:1402.2657 [hep-ph].
- [36] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, JHEP 08, 001, arXiv:hep-ph/9707323.
- [37] M. Wobisch and T. Wengler, in Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting) (1998) pp. 270–279, arXiv:hep-ph/9907280.
- [38] 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].
- [39] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF), Nucl. Phys. B 877, 290 (2013), arXiv:1308.0598 [hep-ph].
- [40] T. Sjostrand and P. Z. Skands, Eur. Phys. J. C 39, 129 (2005), arXiv:hep-ph/0408302.
- [41] J. Bellm et al., Eur. Phys. J. C 76, 196 (2016), arXiv:1512.01178 [hep-ph].
- [42] J. Bellm et al., Eur. Phys. J. C 80, 452 (2020), arXiv:1912.06509 [hep-ph].
- [43] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, Phys. Rev. D 93, 033006 (2016), arXiv:1506.07443 [hep-ph].
- [44] S. Gieseke, P. Stephens, and B. Webber, JHEP 12, 045, arXiv:hep-ph/0310083.
- [45] P. Skands, S. Carrazza, and J. Rojo, Eur. Phys. J. C 74, 3024 (2014), arXiv:1404.5630 [hep-ph].
- [46] E. Norrbin and T. Sjostrand, Nucl. Phys. B 603, 297 (2001), arXiv:hep-ph/0010012.
- [47] M. Bahr et al., Eur. Phys. J. C 58, 639 (2008), arXiv:0803.0883 [hep-ph].
- [48] T. Sjostrand and M. van Zijl, Phys. Rev. D 36, 2019 (1987).
- [49] M. Bahr, S. Gieseke, and M. H. Seymour, JHEP 07, 076, arXiv:0803.3633 [hep-ph].
- [50] S. Gieseke, F. Loshaj, and P. Kirchgaeßer, Eur. Phys. J. C 77, 156 (2017), arXiv:1612.04701 [hep-ph].
- [51] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, Phys. Rept. 97, 31 (1983).
- [52] B. R. Webber, Nucl. Phys. B 238, 492 (1984).
- [53] A. Kupco, in Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting) (1998) pp. 292–300, arXiv:hep-ph/9906412.
- [54] A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lonnblad, J. Monk, H. Schulz, and F. Siegert, Comput. Phys. Commun. 184, 2803 (2013), arXiv:1003.0694 [hep-ph].
- [55] C. Bierlich et al., SciPost Phys. 8, 026 (2020), arXiv:1912.05451 [hep-ph].
- [56] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C 72, 1896 (2012), arXiv:1111.6097 [hep-ph].
- [57] A. M. Sirunyan et al. (CMS), JINST 13 (05), P05011, arXiv:1712.07158 [physics.ins-det].
- [58] D. P. Kingma and J. Ba (2014) arXiv:1412.6980 [cs.LG].
- [59] (2018).
- [60] (2017).
- [61] G. Aad et al. (ATLAS), Eur. Phys. J. C 70, 823 (2010), arXiv:1005.4568 [physics.ins-det].