17 October 2022
jonathan.charles.barrett@cern.ch

Performance of Isotropy for Jet Tagging

Jonathan Barrett
Supervisors: Matt LeBlanc & Maximilian Swiatlowski
CERN, CH-1211 Geneva, Switzerland

Keywords: Jet substructure; Jet tagging; Energy Mover's Distance; Isotropy.

Summary

We use the newly proposed Energy Mover's Distance as a measure of jet isotropy to define new jet substructure observables for quark/gluon discrimination and identifying hadronically-decaying top quarks with large transverse momentum. We assess their effectiveness by comparing them with other classifiers. The quark/gluon study is conducted at hadron level while the top quark study is conducted at detector-level in events reconstructed with a simulated version of the ATLAS detector implemented in GEANT4.

Contents

1 Introduction

1.1 Jet Substructure

Jets are collimated hadron showers produced in most events at the Large Hadron Collider. The energy of the protons accelerated in the LHC is sufficient for the partons (quarks and gluons) in the colliding protons to interact to produce many additional partons in a collimated parton shower. The parton showers hadronize to form jets that can be detected by particle detectors [1]. The internal structure (substructure) of a jet varies depending on the type of parton (quarks/gluons) which initiated the jet [1]. For instance, gluon jets tend to have a greater proportion of soft radiation than hard radiation and greater constituent multiplicity than quark jets while quark jets are usually more collimated than gluon jets [2].

Jets can be used to identify the particles which decayed to produce them, so understanding jet substructure is important in searches for beyond the standard model (BSM) physics. For example, a massive BSM particle could decay and produce a standard model (SM) W, Z, or H resonance (which can decay into quarks) or top quark with a high transverse momentum (p_T) such that the products of the subsequent decay of the SM resonance are collimated and can be reconstructed as a jet [1]. Specifically, when such a top quark decays, it tends to form jets with a three-pronged substructure. I.e. the decay may form a jet in which three constituent subjets can be identified.

Algorithms to reconstruct jets and identify their origin (tag) typically use the following processes:

1. **Recombination.** Recombines particles resulting from the parton shower to reconstruct jets. We used the anti-k_t algorithm [3] implemented in FastJet [4] with $R = 0.4$ in the quark/gluon study and $R = 1.0$ in the top quark study.

2. **Grooming.** This process is optional. The goal is to filter out as much of the background radiation as possible by removing soft radiation that is not correlated with the signal particle decay [1]. We used the soft-drop grooming algorithm [5] with z_{cut}=0.1 and =1.0 on the jets used for top quark reconstruction (these jets have R=1.0).

3. **Classification.** Classifiers such as jet shape observables (The observables in this report are of this type) or prong-finders (such as N-subjettiness [6]) are used to discriminate between the the remaining signal radiation and background radiation [1].

The Energy Mover's Distance of section 1.2 is used to define new jet shape observables that may be useful as classifiers for jet tagging algorithms.

1.2 Energy Mover's Distance (EMD)

Energy Mover's Distance (EMD), first defined in Ref. [7] and inspired by the Earth Mover's Distance, is a measure of the minimum work required to rearrange the energy distribution of one event, denoted E, into the energy distribution of another event, denoted E'. Mathem-

atically, it is defined in Ref. [7] by equation 1:

$$EMD(\mathsf{E};\mathsf{E}') = \min_{f_{ij}\geq 0} \sum_{ij} f_{ij}\frac{_{ij}}{R} + \left|\sum_i E_i - \sum_j E'_j\right|;$$
$$\sum_j f_{ij} \leq E_{ij}; \sum_i f_{ij} \leq E'_{ij}; \sum_{ij} f_{ij} = E_{\min} \tag{1}$$

Where i is an index for particles in event E and j is an index for particles in event E', f_{ij} corresponds to a movement of energy from particle i in E to particle j in E', $_{ij}$ is the angular distance between the two particles indexed by i and j, R controls the importance of the first term relative to the second, and finally, E_i and E_j are the energies of the particles indexed by i and j. is an angular weighting parameter from Ref. [8] which results in the modification to equation 1 in equation 2:

$$EMD(\mathsf{E};\mathsf{E}') = \min_{f_{ij}\geq 0} \sum_{ij} f_{ij}\left(\frac{_{ij}}{R}\right) + \left|\sum_i E_i - \sum_j E'_j\right|;$$
$$\sum_j f_{ij} \leq E_{ij}; \sum_i f_{ij} \leq E'_{ij}; \sum_{ij} f_{ij} = E_{\min} \tag{2}$$

In Ref. [9], EMD is used to define event isotropy, an event shape observable to measure how closely an event resembles a uniform radiation distribution. Mathematically Ref. [9] defines the event isotropy as:

$$\mathsf{I}_n^{geo}(\mathsf{E}) = EMD_{geo}(\mathsf{U}_n^{geo};\mathsf{E}) \tag{3}$$

In equation 3, geo specifies the type of reference geometry (in this project, we consider the ring and cylinder reference geometries denoted $Ring$ and Cyl), n specifies the number of points in the reference geometry, E is an event energy distribution, and U_n^{geo} is the uniform reference with n points. [9]

2 Analysis

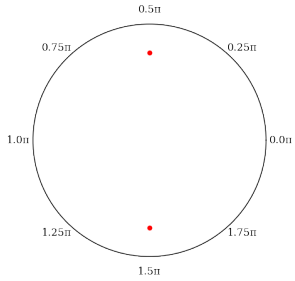
In this work, we use EMD as a measure of jet isotropy rather than event isotropy. Whereas the event isotropy from Ref. [9] is defined as the EMD between an event radiation pattern and a uniform radiation pattern (such as a ring or cylinder with n points), the jet isotropy is the EMD between a single jet radiation pattern and a uniform radiation pattern. Specifically, we use the isotropy with values of 1, 2, and 4. Pictured in Figures 1 and 2 are the uniform radiation patterns used for comparison with the jet radiation patterns.

In the same format as the uniform radiation patterns in Figures 1 and 2, Figures 3 and 4 are examples of jets displayed over the ring and cylinder reference frames.

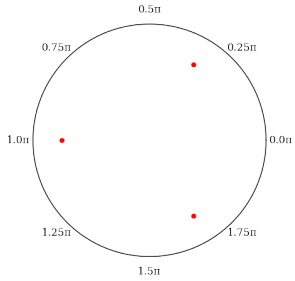
Now, for example, consider Jet A from Figures 3 and 4. To measure how close it is to the ring-like uniform radiation pattern in Figure 1a, we compute the isotropy defined by the Energy Mover's Distance:

$$\mathsf{I}_2^{Ring}(\mathsf{E}_A) = EMD(\mathsf{U}_2^{Ring};\mathsf{E}_A)$$

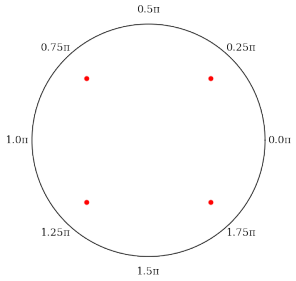
In section 2.1 we test the isotropy jet substructure observables on a quark and gluon jet data set. In section 2.2, we test the isotropy jet substructure observables on a large transverse momentum top jet tagging data set.



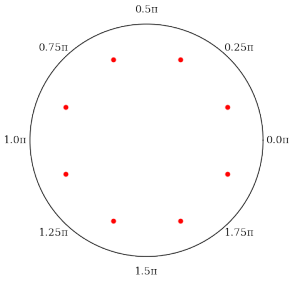
(a) Isotropic Ring reference with $N = 2$



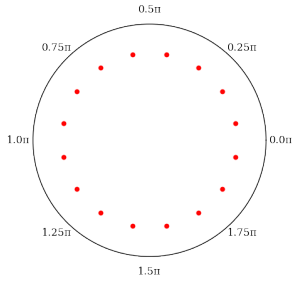
(b) Isotropic ring reference with $N = 3$



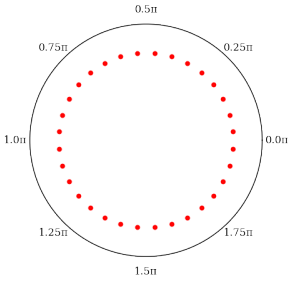
(c) Isotropic ring reference with $N = 4$



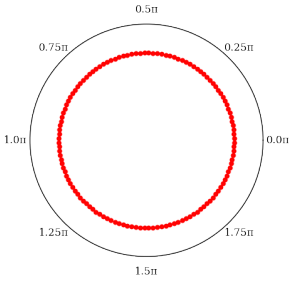
(d) Isotropic ring reference with $N = 8$



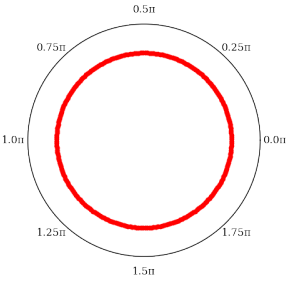
(e) Isotropic ring reference with $N = 16$



(f) Isotropic ring reference with $N = 32$

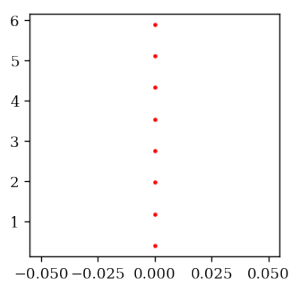


(g) Isotropic ring reference with $N = 128$

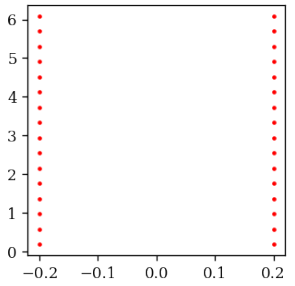


(h) Isotropic ring reference with $N = 256$

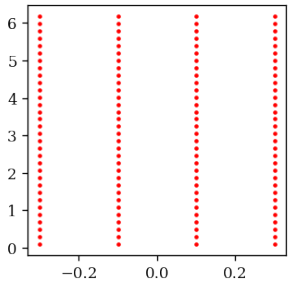
Figure 1: Ring-like uniform radiation patterns with N points



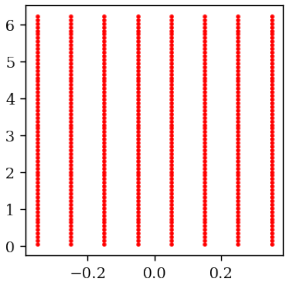
(a) Isotropic cylinder reference with $N = 8$



(b) Isotropic cylinder reference with $N = 16$

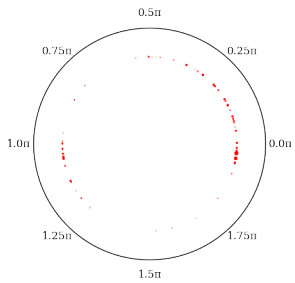


(c) Isotropic cylinder reference with $N = 32$

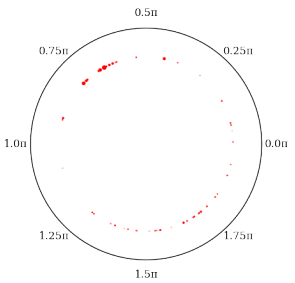


(d) Isotropic cylinder reference with $N = 64$

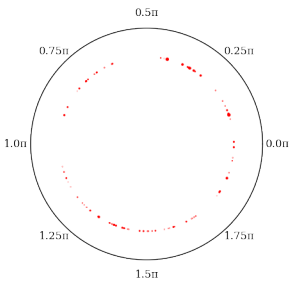
Figure 2: Cylinder-like uniform radiation patterns with N points



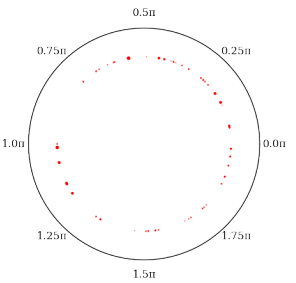
(a) Jet A



(b) Jet B

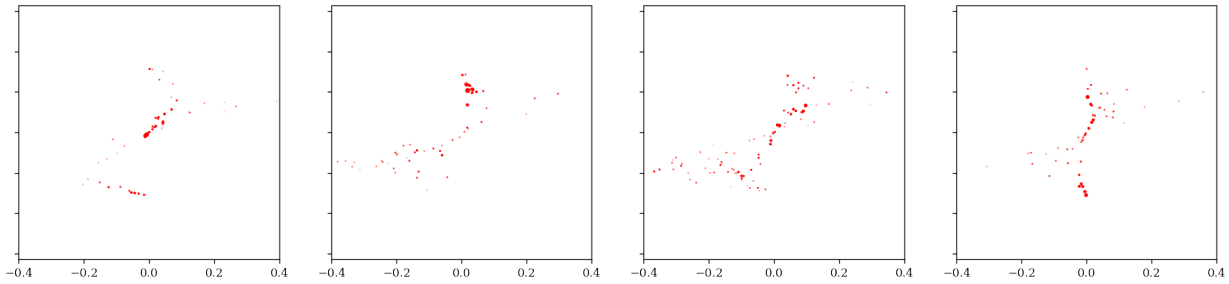


(c) Jet C



(d) Jet D

Figure 3: Sample jets displayed over the ring reference frame



(a) Jet A (b) Jet B (c) Jet C (d) Jet D

Figure 4: Sample jets displayed over the cylinder reference frame

2.1 Quark and Gluon Jet Dataset

The dataset used in this section contains 10 000 *uds* (Up, Down, Strange) quark (q) jets and 10 000 gluon (g) jets. It is a subset of the publicly available data from the EnergyFlow Python package and the data is simulated at particle level (without detector simulation) [10, 11]. The jets all have radius $R = 0.4$ and range from 500 GeV to 550 GeV. We computed the jet isotropies (labelled as costs in the figures) using EnergyFlow's emd_wasserstein function for each jet (E_i) in the dataset with each uniform reference (U_n^{ring} and U_n^{Cyl}) in figures 1 and 2. In sections 2.1.1 and 2.1.2 we compare the classification ability of the isotropy observables with the classification ability of the constituent multiplicities by plotting Receiver Operating Characteristic (ROC) curves.

2.1.1 Ring-Like Geometry

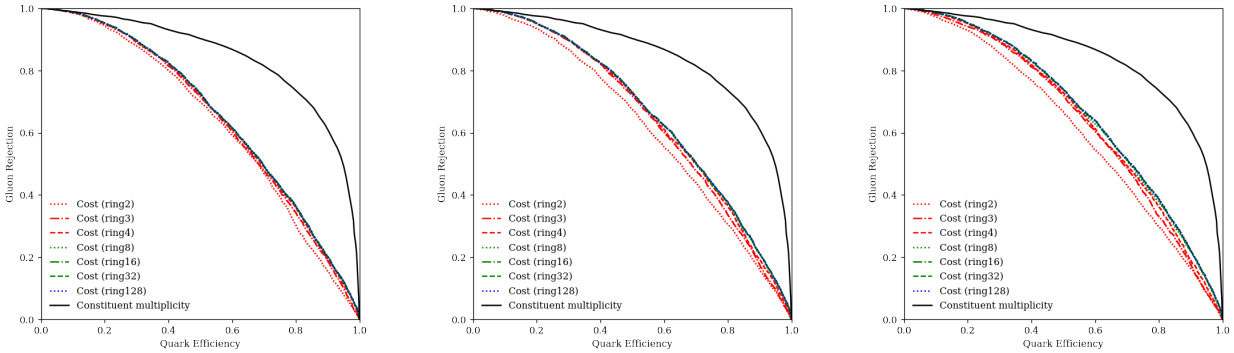
For each $\beta = 1, 2, 4$, there is a ROC curve in figures 5a, 5b, and 5c respectively which corresponds with the costs (isotropies) computed with the q/g jet dataset and the uniform ring references: U_2^{ring}, U_3^{ring}, U_4^{ring}, U_8^{ring}, U_{16}^{ring}, U_{32}^{ring}, U_{128}^{ring}, and U_{256}^{ring}. These are the uniform references displayed in figure 1. The constituent multiplicity outperforms the isotropies as a classifier for all three values of β. As beta increases, it appears that the isotropy for the dipole ($N = 2$) ring geometry, $\mathsf{I}_2^{Ring}(\mathsf{E})$, performs increasingly worse.

2.1.2 Cylinder-Like Geometry

In the case of the cylinder-like reference geometries, U_8^{Cyl}, U_{16}^{Cyl}, U_{32}^{Cyl}, and U_{64}^{Cyl}, the ROC curves for the three β values (1,2, and 4), are displayed in figure 6 with figure 6a corresponding to $\beta = 1$, figure 6b with $\beta = 2$, and figure 6c with $\beta = 4$. The isotropies with the cylinder-like geometries ($\mathsf{I}_n^{Cyl}(\mathsf{E})$) perform worse as classifiers than the isotropies with the ring-like isotropies ($\mathsf{I}_n^{Ring}(\mathsf{E})$) as seen through the comparison of figures 5 and 6.

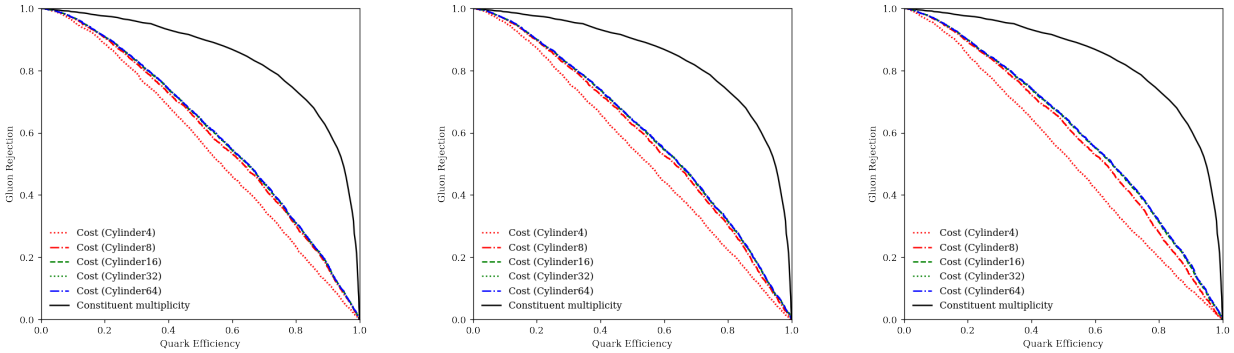
2.2 Boosted Top Quark Tagging Dataset

To study the isotropy observables' ability to classify large transverse momentum (boosted) top quark jets, We used a subset containing 10 000 jets from a public ATLAS boosted top quark tagging detector-level dataset with a 1:1 ratio of top jets to background jets (those



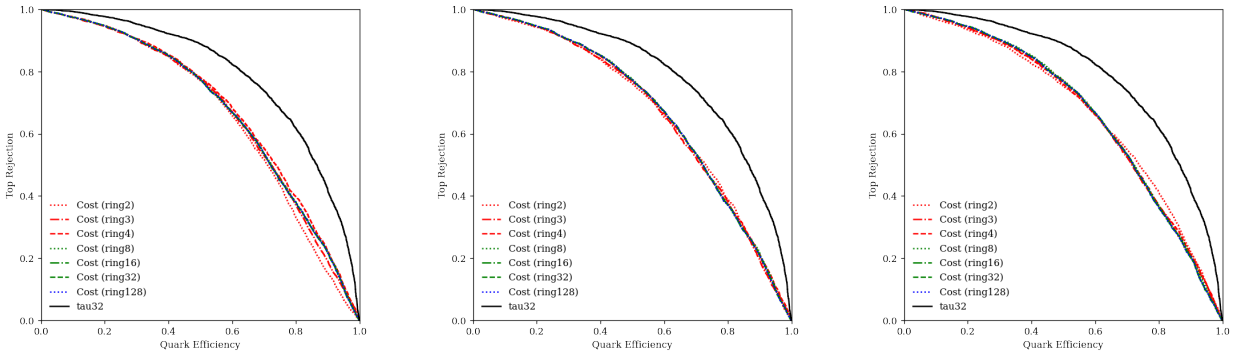
(a) = 1 (b) = 2 (c) = 4

Figure 5: ROC curves for the isotropies with the ring-like uniform references and the constituent multiplicities using the quark/gluon tagging dataset.



(a) = 1 (b) = 2 (c) = 4

Figure 6: ROC curves for the isotropies with the cylinder-like uniform references and the constituent multiplicities using the quark/gluon tagging dataset.



(a) = 1 (b) = 2 (c) = 4

Figure 7: ROC curves for the isotropies with the ring-like uniform references and the N-Subjettiness ratio $_3/\ _2$ using the high transverse momentum top quark tagging dataset.

initiated by other quarks and gluons). The dataset is available in Ref. [12] and documented in Ref. [13]. In this case, we used jets with radius $R = 1.0$ and with energies above 350 GeV that were groomed using the soft-drop grooming algorithm. See Ref. [5] for an introduction to the soft-drop grooming algorithm.

In the detector-level dataset (where the ATLAS detector is simulated in GEANT4), jets are represented as signals from the simulated ATLAS calorimeter and inner tracking detector. The signals are reconstructed as Unified Flow Objects (UFOs) which are jet input objects defined in Ref. [14]. Unified Flow Objects are designed to work well for a large range of kinematics by using information from the inner tracker at low p_T, but also making use of information from the hadronic callorimeter at high p_T to account for the decrease in resolution of the inner tracker with high p_T [13].

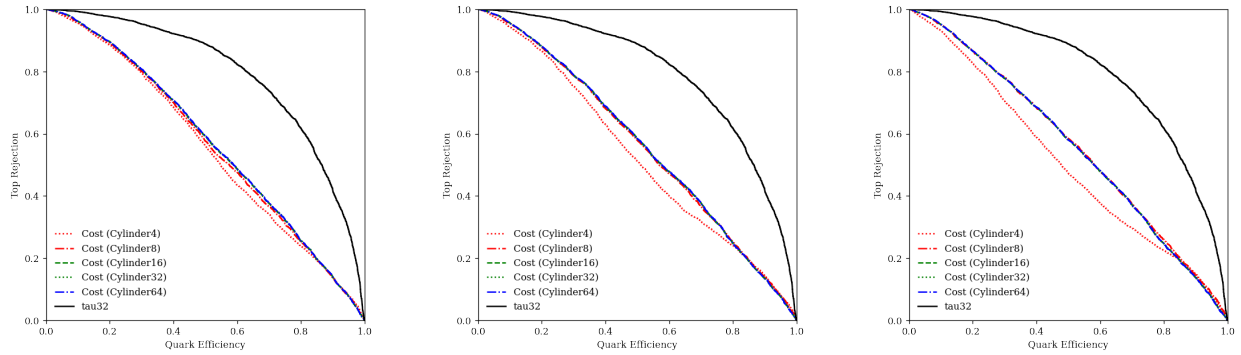
We compare, in sections 2.2.1 and 2.2.2, the ability of the isotropy observables as classifiers with that of the jet shape observable known as N-subjettiness which was introduced in Ref. [6]. Specifically, we compare the isotropy observables with the N-Subjettiness ratio $_3/\ _2$ that was shown to be effective in identifying boosted top jets [6].

2.2.1 Ring-Like Geometry

As was the case with the q/g dataset, for each $= 1; 2; 4$, there is a ROC curve in figures 7a, 7b, and 7c respectively which corresponds with the costs (isotropies) computed with the q/g jet dataset and the uniform ring references: U_2^{ring}, U_3^{ring}, U_4^{ring}, U_8^{ring}, U_{16}^{ring}, U_{32}^{ring}, U_{128}^{ring}, and U_{256}^{ring} (See again, figure 1). The N-Subjettiness ratio $_3/\ _2$ outperforms the isotropies as a classifier for all three values of .

2.2.2 Cylinder-Like Geometry

We also considered the cylinder-like reference geometries, U_8^{Cyl}, U_{16}^{Cyl}, U_{32}^{Cyl}, and U_{64}^{Cyl}, for the boosted top quark tagging dataset. The ROC curves for the three values (1,2, and 4), are displayed in figure 8 with figure 8a corresponding to $= 1$, figure 8b with $= 2$, and figure 8c with $= 4$. For this dataset, we see again that the isotropies with the cylinder-like geometries ($I_n^{Cyl}(E)$) perform worse as classifiers than the isotropies with the ring-like



(a) = 1 (b) = 2 (c) = 4

Figure 8: ROC curves for the isotropies with the cylinder-like uniform references and the N-Subjettiness ratio $_3/_2$ using the high transverse momentum top quark tagging dataset.

isotropies. ($I_n^{Ring}(E)$) as seen through the comparison of figures 7 and 8.

3 Conclusion

For quark and gluon tagging, the constiuent multiplicity performs better as a jet shape observable than the isotropy jet shape observables defined using the Energy Mover's Distance (EMD) for ring-like and cylinder-like uniform reference geometries. For high transverse momentum top quark tagging, the N-subjettiness ratio $_3/_2$ [6] performs better than the isotropy jet shape observables. In both datasets (q/g tagging and high p_T top tagging), the jet-shape observables which use the ring-like geometries tend to do better than those which use the cylinder-like geometries.

4 Science Outreach and EDII in STEM

This experience is helpful in promoting physics in Newfoundland and Labrador by providing an inspiring story about this exciting opportunity for physics students. I am the first Indigenous student in the province to have had the privilege to participate in the CERN/IPP program, and the people of Newfoundland and Labrador are interested in that experience. The CERN/IPP program demonstrates that undergraduate students, even if they come from small, remote communities, can still participate in "Big Science" and have successful careers in STEM. While continuing with my physics degree, I am working part-time for a province-wide science outreach program for underrepresented youth in Newfoundland and Labrador, especially Indigenous students, women, and youth from remote and isolated communities, which is engaging thousands of students per year. The research experience at CERN has helped me to become a role model for youth in the province, so I participate in panels (IYG 2022 Program – Indigenous Youth Initiatives (wisenl.ca) [15], give interviews, and deliver public talks and webinars (Events (mun.ca)) [16], recruiting the next generation of researchers in subatomic physics. The experience also helped build connections between CERN and physics in NL that will make future collaborations possible.

5 Acknowledgements

Many thanks to my supervisors, Matt and Max, for their help and guidance. I appreciate that this work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (USRA) and the travel/living costs for the duration of the work by the Institute of Particle Physics (Canada) (IPP). Thanks to the IPP for organizing the IPP Summer Student Fellowship Program and providing the opportunity to travel for this research experience. Thanks to the organizers of the CERN Summer Student Program and the IPP Summer Student Fellowship Program for this excellent opportunity. Thanks to Dr. Barkanova for feedback and the initial encouragement to participate in this program.

References

[1] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets*. Springer International Publishing, 2019.

[2] G. Rauco, "Distinguishing Quark and Gluon Jets at the LHC," in *Parton Radiation and Fragmentation from LHC to FCC-ee*, pp. 73–78, 2 2017.

[3] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k_t Jet Clustering Algorithm," *Journal of High Energy Physics*, vol. 2008, pp. 063–063, apr 2008.

[4] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *The European Physical Journal C*, vol. 72, mar 2012.

[5] ATLAS Collaboration, "Measurement of Soft-Drop Jet Observables in pp Collisions with the ATLAS Detector at $\sqrt{s}$ =13 TeV," *Phys. Rev. D*, vol. 101, no. 5, p. 052007, 2020.

[6] J. Thaler and K. V. Tilburg, "Identifying Boosted Objects with N-subjettiness," *Journal of High Energy Physics*, vol. 2011, mar 2011.

[7] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Metric Space of Collider Events," *Physical Review Letters*, vol. 123, jul 2019.

[8] P. T. Komiske, E. M. Metodiev, and J. Thaler, "The Hidden Geometry of Particle Collisions," *Journal of High Energy Physics*, vol. 2020, jul 2020.

[9] C. Cesarotti and J. Thaler, "A Robust Measure of Event Isotropy at Colliders," *Journal of High Energy Physics*, vol. 2020, Aug 2020.

[10] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy Flow Networks: Deep Sets for Particle Jets," *Journal of High Energy Physics*, vol. 2019, jan 2019.

[11] P. Komiske, E. Metodiev, and J. Thaler, "Pythia8 Quark and Gluon Jets for Energy Flow." https://doi.org/10.5281/zenodo.3164691, May 2019. (Version v1) [Data set]. Zenodo.

[12] ATLAS Collaboration, "ATLAS Top Tagging Open Data Set." CERN Open Data Portal. DOI: 10.7483/OPENDATA.ATLAS.FG5F.96GA, 2022. Äccessed:2022-10-10.

[13] ATLAS Collaboration, "Constituent-Based Top-Quark Tagging with the ATLAS Detector," tech. rep., CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-039.

[14] ATLAS Collaboration, "Optimisation of Large-Radius Jet Reconstruction for the ATLAS Detector in 13 TeV Proton-Proton Collisions," *Eur. Phys. J. C*, vol. 81, no. 4, p. 334, 2021.

[15] Indigenous Youth Initiatives, WISENL, "IYG 2022 Program - Indigenous Youth Initiatives." https://iyi.wisenl.ca/students-schools/iyg/program/. Accessed 2022-10-17.

[16] Grenfell Campus, Memorial University of Newfoundland and Labrador, "Grenfell Campus Observatory Events." https://www.grenfell.mun.ca/campus-services/Pages/community/observatory/Events.aspx. Accessed 2022-10-17.