IL NUOVO CIMENTO **46 C** (2023) 92
DOI 10.1393/ncc/i2023-23092-y

COMMUNICATIONS: SIF Congress 2022

Study of b- and c-jets identification for Higgs coupling measurements at the Muon Collider

G. DA MOLIN(*)

LIP, Laboratory of Instrumentation and Experimental Particle Physics - Lisbon, Portugal

received 28 January 2023

Summary. — The Muon Collider is a possible option for the next generation of high-energy collider machines. It would permit to achieve very high center-of-mass energy using leptons without occurring in significative synchrotron radiation losses as in positron-electron rings. Due to the muon decay, the detector has to sustain a high level of background: beam decay products and subsequent particles from secondary interactions with the machine elements can reach the interaction point, thus jeopardizing the physical performance of the detector. Nevertheless, this machine offers the possibility of producing huge Higgs boson samples with respect to the backgrounds, allowing the precise determination of the b- and c-quark couplings. In this contribution, a study of the identification of b- and c-jets in the Muon Collider environment is presented, and an estimation of the statistical precision on the H $\rightarrow b\bar{b}$ cross-section is evaluated, accounting for the presence of the beam and the physical backgrounds. A preliminary study on the use of machine learning techniques to improve the result and to study the H $\rightarrow c\bar{c}$ will also be reported.

1. – Introduction

The Higgs sector of the Standard Model (SM) can offer great insights into possible New Physics (NP). In fact, new particles may acquire mass through the Higgs field and so couple with the Higgs boson. This behaviour could be manifested in a deviation of the Higgs boson couplings from the values predicted by the SM. As such, to detect these deviations, it is imperative that the new generation of colliders provide accurate measurements of the Higgs parameters.

In this article, the expected precision of a Muon Collider at $\sqrt{s} = 3$ TeV on the H $\rightarrow b\bar{b}$ cross section is shown with the use of a detailed simulation, accounting for both the physical and the beam-induced background (BIB). The initial study on jet identification techniques to identify the H $\rightarrow c\bar{c}$ is also presented. In fact, a Muon Collider can be considered a Higgs factory since, thanks to the vector boson fusion mechanism,

(*) E-mail: damolingia@lip.pt

around half a million Higgs bosons are expected to be produced at $\sqrt{s} = 3\,\text{TeV}$ with an integrated luminosity of 1 ab^{-1}, corresponding to 5 years of data-taking. However, muons are not stable and may decay while travelling along the beam-line. Their decay products may interact with parts of the accelerator producing showers of secondary and tertiary particles, forming the so-called BIB, which may enter the detector and produce a challenging background. The simulation and mitigation of BIB will be described in the next chapter.

2. – Monte Carlo samples

The signal particles are generated either with PYTHIA8 [1] or WHIZARD [2]. The response of the detector is instead simulated by using the ILCSoft package [3], which handles the geometry of the detector, the interactions of particles passing through the material and the event reconstruction algorithm. The BIB particles are instead produced with MARS15, which computes all the BIB particles emerging from the beam interactions within 25 meters from the Interaction Point (IP), except for the Bethe-Heitler muons which are simulated from within 100 meters from the IP. The hits in the detector left by a signal and a BIB event are then superimposed before the digitisation of hits. In this study, the BIB events used are the ones corresponding to the center-of-mass energy of $\sqrt{s} = 1.5\,\text{TeV}$. This is a conservative assumption: the BIB level decreases as $\sqrt{s}$ increases, since at higher $\sqrt{s}$, the muon lifetime is longer. The BIB causes an extensive background in the detector and it is necessary to look for mitigation's strategies. The tungsten shielding in the machine detector interface reduces the BIB yield by a factor of 500 with respect to the non-absorber case. Most of the remaining BIB particles have a soft p_T spectra, arrive late with respect to the bunch crossing time and their direction is not fully consistent with particles coming from the interaction point.

These characteristics will be exploited by designing a dedicated detector to remove a sizable component of this background. The current detector model is a cylinder and, starting from the innermost components, is made up of: a silicon tracking system, composed by a Vertex Detector (VD), an Inner Tracker (IT) and an Outer Tracker (OT); an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL); the solenoid producing a magnetic field $B = 3.57\,\text{T}$ directed along the axis of the cylinder and the Muon Detector. Timing requirements on each of these subdetectors are used to remove a sizable component of hits left by the BIB particles: in the tracking system, defining $t = 0$ as the bunch crossing time, hits outside the time windows $[-0.18\,\text{ns}, 0.24\,\text{ns}]$ for the VD and $[-0.36\,\text{ns}, 0.48\,\text{ns}]$ for the trackers are removed. The same is done for the calorimeters and muon detector for hits outside $[-0.25\,\text{ns}, 0.25\,\text{ns}]$. Additionally, the BIB leaves an underlying energy deposit in the calorimeter and so an (η, ϕ) dependent energy subtraction is performed before looking for signal jets.

Requirement on the impact parameters, χ^2 and number of hits are employed to remove fake tracks built with lingering hits from BIB particles. These tracks and clusters in the calorimeters are combined to build particle objects with the PandoraPFA algorithm [4]. Jets are built clustering particles using the k_T algorithm [5] with $R = 0.4$.

3. – Estimation of the $\text{H} \to b\bar{b}$ cross section sensitivity

At a $\sqrt{s} = 3\,\text{TeV}$ Muon Collider, the Higgs boson is produced mainly via WW fusion. The sample used in this work comprises 2000 $\text{H} \to b\bar{b}$ events with the BIB superimposed and it was generated with PYTHIA8 ($\sigma_{Hbb} = 324\,\text{fb}$). Jets originated from b-quarks are selected looking for a secondary vertex (SV), produced by the decay of the long lived b-hadron, inside a jet. The tracks to be used in the vertex reconstruction algorithms are required to have $p_T > 0.8\,\mathrm{GeV}$ and to be displaced from the primary vertex both in the transverse and longitudinal plane. With these requirements, the b-tagging efficiency (ratio of selected b-jets over total b-jets), c-mistag and light-jets mistag (ratio of selected c- or lights jets over total number of c- or light jets, respectively), estimated on dijets samples, were found to be around 60%, 30% and < 2% respectively. Signal events were selected by requiring 2 b-tagged jets with $p_T > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. The invariant mass built from the two b-tagged jets is required to be within [0,300] GeV.

Since the light jets mistag is found to be small, the background processes with two light jets in the final state can be neglected in first approximation. The dominant background is instead given by a Z produced via WW fusion and decaying into a pair of b- or c-quarks. The contribution of other processes with two heavy flavour quarks are found to be negligible. The effect of Z $\to b\bar{b}$ and Z $\to c\bar{c}$ have been studied using a sample of 2000 events generated with WHIZARD Monte Carlo, to which the BIB was then superimposed. The cross section of this background is 610.7 fb.

The relative statistical uncertainty on the H $\to b\bar{b}$ cross section $\frac{\Delta\sigma_{H\to b\bar{b}}}{\sigma_{H\to b\bar{b}}}$ has been estimated with a toy Monte Carlo study based on ROOFIT [6]. Gaussian fits were performed on the dijet invariant mass of signal and background. The yields of both signal and background were left free in the fit and by analysing pseudo-experiments (fig. 1), assuming an integrated luminosity of 1 ab^{-1}, $\frac{\Delta\sigma_{H\to b\bar{b}}}{\sigma_{H\to b\bar{b}}}$ is found to be 0.8%. This is similar compared to what found by CLIC, an e^+e^- collider, which expects a relative statistical uncertainty of 0.3% with 2 ab^{-1} [7]. However, the present result can be further improved and only serves as a starting benchmark. Different algorithms can be used to reject lingering BIB hits in the tracker and calorimeters and improve the tracking reconstruction efficiency and jet energy resolution. Also, machine learning can be applied to jet flavor identification and improve the b-tagging efficiency while decreasing the c-jet mistag contamination.

4. – Preliminary study of Machine Learning for Jet flavour identification

A better separation of b-jets and c-jets (and an improved light-jets rejection) would not only significantly improve the precision available in the $H \to b\bar{b}$, but also allows studying the still unmeasured $H \to c\bar{c}$. For this reason, a feed-forward Deep Neural

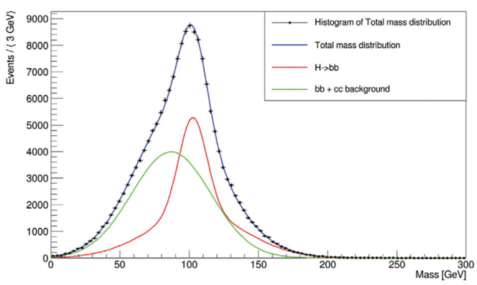

Fig. 1. – Result of one of the pseudo-experiments obtained with ROOFIT. The total mass distribution is shown in blue, the Higgs mass distribution is depicted in red and the background distribution is reported in green.

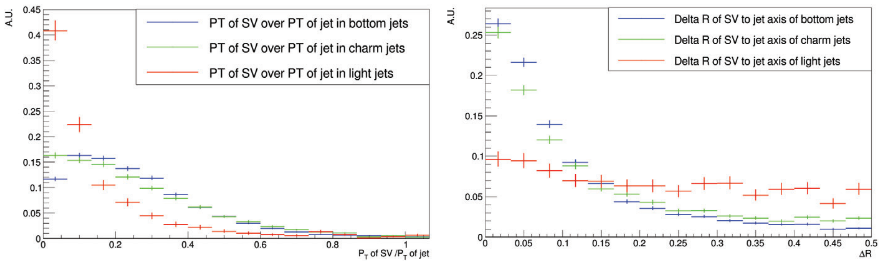

Fig. 2. – Distribution of ratio of P_T of an SV w.r.t. P_T of the jet (a) and $\Delta\text{R} = \sqrt{\eta^2 + \phi^2}$ between SV and jet axis, where ϕ is the polar angle. Even when a SV is found in a light jet (red), it usually has a relatively low p_T and a random position inside the jet cone, unlike what happens in b-jets (blue) and c-jets (green).

Network (DNN) is explored, aimed at classifying the input jet as a b-jet, c-jet or light jet. Inputs to the DNN are the most significant features of b-, c-jets and light jets, which are studied from dijets samples produced with PYTHIA8 and with the BIB superimposed. The observables were chosen to be process independent, as well as not to change from the translation or rotation of the studied jet. These variables (shown in fig. 2) come from global features of the jet, properties of the secondary vertex (if present) and the kinematics and positions (relative to the jet, to maintain process and translational invariance) of the particles making up the substructure of the jet.

In this test, some separation is achieved: an area under the Receiving Operating Characteristic (ROC) of 0.776 is found for the light jets and an area under the ROC of 0.714 is obtained by testing just the b-jet *vs.* c-jets. However, as only around 10k jets for each flavour were available between training and testing, higher statistics is needed to strengthen these results.

5. – Conclusions

The relative statistical uncertainty on $\sigma_{H\to b\bar{b}}$ with 1 ab^{-1} is found to be 0.8% using a simple SV-tagging algorithm to identify b-jets. Improvements are possible by enhancing the reconstruction and BIB rejection algorithms and refine the jet flavour identification.

REFERENCES

[1] Sjöstrand T., Mrenna S. and Skands P., *Comput. Phys. Commun.*, **178** (2008) 852.
[2] Kilian W., Ohl T. and Reuter J., *Eur. Phys. J. C*, **71** (2011) 1742.
[3] Gaede F. *et al.*, iLCSoft Github repository, `https://github.com/iLCSoft/ilcsoftDoc`.
[4] Marshall J. S. and Thomson M. A., *The Pandora Particle Flow Algorithm*, `https://doi.org/10.48550/arXiv.1308.4537`.
[5] Catani S., Dokshitzer Yu. L., Seymour M. H. and Webber B. R., *Nucl. Phys. B*, **406** (1993) 187.
[6] Hageboeck S., *PoS*, **ICHEP2020** (2021) 910.
[7] Abramowicz H., Abusleme A., Afanaciev K. *et al.*, *Eur. Phys. J. C*, **77** (2017) 475.