PROCEEDINGS OF SCIENCE

# Enhancing Prompt Lepton Identification: Development and Optimization of the PLIT Tagger

**Ema Maričić**[a,b,*] **on behalf of the ATLAS Collaboration**

[a] *Institute of Physics, University of Belgrade,*
*Pregrevica 118, Belgrade, Serbia*

[b] *IRFU, CEA, Université Paris-Saclay,*
*Bat. 141, Gif-sur-Yvette, France*

*E-mail:* ema.maricic@cern.ch

Within the ATLAS experiment the Prompt Lepton Isolation Tagger (PLIT) has been developed to serve as an essential tool to distinguish between prompt muons originating from the decays of the Higgs, W or Z bosons and non-prompt muons generated in the semi-leptonic decays of $b$- and $c$-hadrons. Its central role is to effectively mitigate the presence of fake and non-prompt leptons in various multi-lepton final state analyses. This contribution presents the ongoing efforts in developing and optimizing this tagger for Run-3 data analyses.

ATL-MUON-PROC-2024-016
02 October 2024

*42nd International Conference on High Energy Physics (ICHEP2024)*
*18-24 July 2024*
*Prague, Czech Republic*

[*] Speaker

## 1. Introduction

The Prompt Lepton Isolation Tagger (PLIT) is an isolation algorithm using a transformer neural network architecture to separate prompt electrons and muons, arising in decays of electroweak bosons, from the non-prompt leptons produced in semi-leptonic decays of $b$- and $c$-hadrons. Preliminary results on its performance using simulated $t\bar{t}$ events produced with setups corresponding to the Run 2, as well as the partial Run 3 (2022 and 2023) configurations of ATLAS [1] are presented. The PLIT is based on the same neural network architecture used in ATLAS to identify jets originating from $b$- and $c$-quarks [2, 3]. It uses features from the lepton, features from tracks within a cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.4$ of the lepton, and features which are computed with respect to the variable-radius track jet [4] closest in $\Delta R$ with $\Delta R < 0.4$ to the lepton. Discrimination is based on the final PLIT disriminant score. It is defined as

$$D_{\mathrm{PLIT}} = \log \frac{p_{\mathrm{prompt}}}{p_{\mathrm{non\text{-}prompt}}}, \tag{1}$$

where $p_{\mathrm{prompt}}$ and $p_{\mathrm{non\text{-}prompt}}$ are the prompt and non-prompt muon probabilities, respectively.

The training is performed on 40 million muons from a mixture of simulated $t\bar{t}$ samples with centre of mass energies of 13 and 13.6 TeV. The muons in the training dataset are required to pass the Medium identification working point [5], satisfy $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$, and have additional criteria regarding the distance between primary vertex and each track.

The supervised training is based on preserved generator-level information to match the track of each reconstructed lepton candidate to a suitable stable particle produced by the generator or the Geant4 [6] simulation. Muons in this dataset are labelled as prompt if they are matched to a generator-level muon originating from the decay of a $W$ boson. They are labelled non-prompt otherwise. A resampling procedure to ensure a proportion of 3:1 and equal $p_{\mathrm{T}}$ and $\eta$ distributions between prompt muons and non-prompt muons is applied.

## 2. Architecture of the PLIT

First, input features associated with the lepton are concatenated with features of tracks within a cone of $\Delta R < 0.4$ of the lepton. Next, an initial embedding to a representation space is performed with a Deep Sets [7] network without using the aggregation over the output track representations. The track representations are fed into a transformer encoder [8]. The output representation of each track is then combined to form a global representation of the lepton to be used for classification. This global representation is formed by a weighted sum over the track representations, where the attention weights for the sum are learned during training. In addition to the primary lepton classification task, PLIT also has auxiliary training objectives to classify the track origin and grouping tracks into common vertices. Each of the tasks consists of three hidden layers containing 128, 64 and 32 neurons respectively. These auxiliary tasks enhance the sensitivity to leptons from semi-leptonic $B$-decays which result in the presence of secondary vertices.

## 3. Results and conclusions

As shown in Figures 1 and 2, this novel algorithm shows impressive ability to separate prompt from non-prompt muons. It was observed that the improvement of the PLIT over recommended working points at a fixed value of prompt muon efficiency is of order 2.2 - 2.7 in terms of non-prompt lepton rejection. In both cases the statistical uncertainties are calculated using binomial uncertainties and are indicated as colored bands. The next steps in its development will be the optimization of its performance on electrons and its calibration using data.

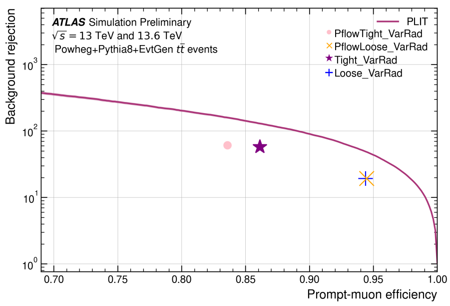


**Figure 1:** The non-prompt muon rejection factor as a function of the prompt muon efficiency for PLIT as well as for some cut-based isolation working points (PFlow Loose VarRad, PFlow Tight VarRad, Loose VarRad and Tight VarRad), as described in Ref. [5].

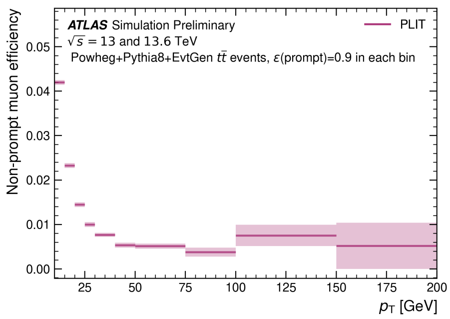


**Figure 2:** The non-prompt muon efficiency as function of the muon $p_T$ for PLIT, for a fixed prompt muon efficiency $\epsilon_{\text{prompt}}$ of 90%, obtained on a sample of 3.5 million muons from $t\bar{t}$ events simulated at centre of mass energies of 13 and 13.6 TeV.

## References


[1] ATLAS Collaboration, 2008 JINST 3 S08003

[2] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027, 2022, URL: https://cds.cern.ch/record/2811135

[3] ATLAS Collaboration, *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*, ATL-PHYS-PUB-2023-021, 2023, URL: https://cds.cern.ch/record/2866601

[4] ATLAS Collaboration, *Variable Radius, Exclusive-$k_t$, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017, URL: https://cds.cern.ch/record/2268678

[5] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex]

[6] S. Agostinelli et al., `Geant4` *- a simulation toolkit*, Nucl. Instrum. Meth. A 506 (2003) 250

[7] M. Zaheer et al., *DeepSets*, 2017, URL: https://arxiv.org/abs/1703.06114

[8] S. Shleifer, J. Watson and M. Ott, *NormFormer: Improved Transformer Pretraining with Extra Normalization*, 2021, URL: https://arxiv.org/abs/2110.09456