# Study of the efficiency of the first CMS High Level Trigger using tagged protons

Pedro Batista[1,a]

[1] *Faculdade de Ciências da Universidade de Lisboa, Portugal*

Project supervisors: Andrea Bellora, Gustavo Gil da Silveira *September 8, 2024*

**Abstract.** The Precision Proton Spectrometer (PPS) is a subdetector of the CMS experiment and uses silicon pixel tracking detectors to measure protons scattered at very small angles. A High Level Trigger (HLT) consists of an array of processors that execute an accelerated version of the event reconstruction algorithm, and is responsible for reducing the event rate to around 1 kHz before data storage. This study focuses on the determination of the efficiency of the first CMS jet matching HLT using data from LHC run 3, before TS1, involving three eras. The efficiency of the HLT is determined by comparison with a looser HLT. Differences between the three eras are spotted in some properties of the jets. Some differences are also identified between events failing the HLT and events passing HLT.

Keywords: LHC, CMS-PPS, NANOAOD, HLT

## 1 Introduction

### 1.1 The CMS-PPS detector

The Compact Muon Solenoid (CMS) detector [2], displayed in Figure 1, is located at interaction point 5 (IP5), and its main distinguishing feature is a superconducting solenoid of 6 m diameter, capable of providing a magnetic field of 3.8 T. A silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter — all consisting of a barrel and two endcap sections — are housed within the solenoid volume [7]. The barrel and endcap detectors' pseudorapidity coverage is expanded by forward calorimeters. Muons are measured using gas-ionization detectors that are inserted outside of the solenoid in the steel flux-return yoke. In Figure 2, it's represented the interaction of different particles in the detector. Depending if the particles are charged, they leave a curved trajectory. In the most inner layer, there is the silicon tracker, where no particles are stopped and their trajectories are recorded. Next there is the electromagnetic and hadronic calorimeters, measuring the energy of particles that interact primarily via the electromagnetic interaction and via the strong nuclear force, respectively. The main difference between the calorimeters is the type of material used, like elements with high atomic number for the electromagnetic calorimeter and high density elements for the hadronic calorimeter. Photons and electrons are completely stopped by the electromagnetic calorimeter, and hadrons by the hadronic calorimeter. Muons, being very low interacting particles, are not stopped by any of the subsystems.

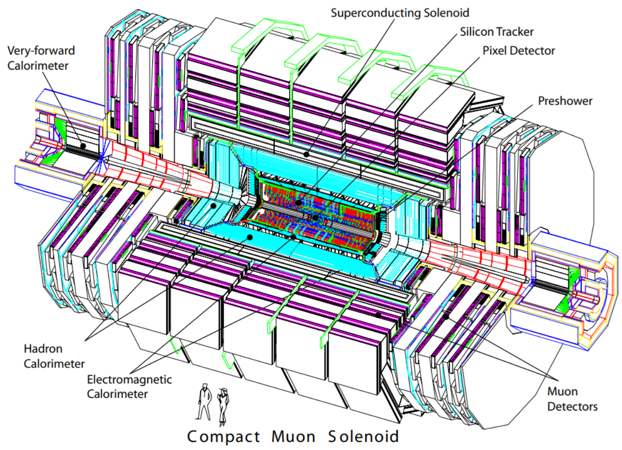


**Figure 1.** A perspective view of the CMS detector [2].

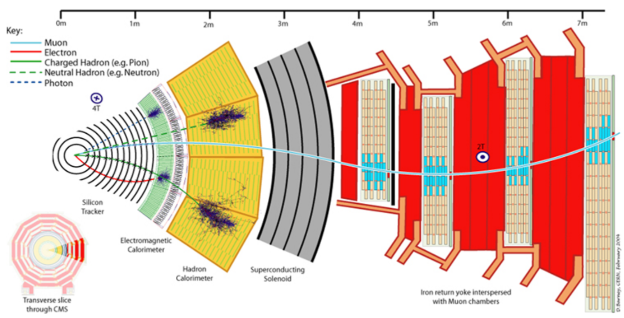


**Figure 2.** Transverse view of different particles interacting with the CMS detector.

The Precision Proton Spectrometer (PPS) detector system [4] is a subdetector of the CMS experiment and measures protons scattered at very small angles at high instantaneous luminosity. The PPS system consists of groups of silicon-pixel tracking detectors and timing detectors, called Roman Pots (RPs), placed inside movable beam

[a]e-mail: cbatista.pedro@gmail.com

pipe insertions and approaching the beam within a few mm, that are capable of measuring the scattered protons that remain inside the beam pipe.

The RPs are located on each side of the CMS detector, at around 200-220 m from IP5, along the beam line in the LHC sector between the interaction points 4 and 5, referred as sector 45, and between interaction points 5 and 6, referred as sector 56 and represented in Figure 3. Some RPs are installed vertically and approach the beam from the top and bottom, while some are installed horizontally. The vertically installed RPs are only used in special configurations of the LHC and in low luminosity proton-proton fills for the purpose of calibration and alignment of the detectors. The horizontally installed RPs are used during standard machine operation since, in this case, protons undergo a large displacement in the horizontal direction and only a small vertical displacement at the RP positions. Each RP system is also defined as a detector arm, where the tracking RP closer to IP5 is referred as "near", and the other as "far". Each detector arm consists of two RPs instrumented with silicon-pixel tracking detectors that measure the transverse displacement of the protons with respect to the beam, and one RP with timing detectors, which are composed of single- and double-sided single crystal chemical vapor deposition diamond sensor planes [6], to measure their time-of-flight. Since 2018, and also in run 3, each RP hosts six 3D pixel sensor planes [5]. The timing detectors are currently not being used.

### 1.2 Coordinate system and kinematic variables

The coordinate system adopted by CMS has the origin centered at the nominal collision point, the $y$-axis pointing vertically upwards, and the $x$-axis pointing radially inward toward the center of the LHC, which automatically defines the direction of the $z$-axis that sits along the beam line, with positive sense pointing to the Jura. The azimuthal angle $\phi$ is measured from the $x$-axis in the $x-y$ plane and the radial coordinate in this plane is denoted by $r$. The polar angle $\theta$ is measured from the $z$-axis. Pseudorapidity is defined as $\eta = -\ln(\tan\frac{\theta}{2})$. Therefore, from the $x$ and $y$ components, the momentum and energy transverse to the beam direction, represented by $P_T$ and $E_T$, respectively, are calculated. $E_T^{miss}$ represents the energy imbalance as measured in the transverse plane. The fraction of momentum lost by the protons $\xi$ is determined as

$$\xi_{1,2} = 1 - \frac{P_{1,2}^{final}}{P_{1,2}^{initial}}, \quad (1)$$

and is used to define the central system mass $M_X = \sqrt{s \cdot \xi_1 \cdot \xi_2}$ and rapidity $Y_X = \frac{1}{2}\ln\left(\frac{\xi_1}{\xi_2}\right)$.

### 1.3 High level trigger

A two-tiered trigger mechanism is used to choose events of interest. Using data from the muon detectors and calorimeters, the Level 1 trigger (L1), made up of specialized hardware processors, selects events at a set latency of around 4 $\mu$s and at a rate of about 100 kHz [1]. At the second level, referred to as the high-level trigger (HLT), an array of processors executing an accelerated version of the entire event reconstruction software reduces the event rate to roughly 1 kHz prior to data storage [3].

In this case, the study will be targeting the efficiency of a jet matching HLT. The main physics motivation behind this HLT is that the PPS physics program aims specifically at the study of Central Exclusive Production (CEP) events, like $pp \to pWWp$ and $pp \to pZZp$, which didn't have a dedicated HLT. Standard Model (SM) CEP events are generally in the low-mass (low $\xi$, low jet $P_T$) region. It was observed, in preliminary studies for $WW$, that jet $P_T$ thresholds generally create a very low trigger efficiency, or no efficiency, in the low-mass region. As a result, it was proposed to create an HLT that required a matching between the multi-jet system mass/rapidity/$\xi$ and the protons, and also with a lower jet $P_T$ requirement, which aims at increasing the sensitivity of the SM processes. Besides SM CEP measurements, proton matching also provides very clean final states, where, potentially, all the energy is reconstructed, providing a strong background rejection. This trigger also provides further sensitivity to all Beyond the Standard Model (BSM) processes with hadronic final states in the low $\xi$ region, like $ZZ$, $HH$, $H^+H^-$.

The current implementation of this HLT requires a number of jets greater than four, a di-jet average $P_T$ greater than 180 GeV, a maximum number of two tracks per RP and a $\xi$ matching between the multi-jet system and protons smaller than 0.3 for the $\xi$ matching, first the $\xi$ of the multi-jet system is determined as

$$\xi_{jets}^{\pm} = \frac{1}{\sqrt{s}} \sum_{i=0}^{N_{jets}} [E(j_i) \pm P_Z(j_i)], \quad (2)$$

where the $\pm$ refers to which detector arm the jet belongs, being + for sector 45 and − for sector 56, and then the $\xi$ matching is done as

$$\xi_{matching} = \frac{[\min](|\xi_{jets}^{\pm} - \xi_{protons}^{\pm}|)}{\xi_{jets}^{\pm}}. \quad (3)$$

## 2 HLT efficiency

### 2.1 Data utilized

The analysis was performed on top of the JetMET dataset, from LHC Run 3 and before TS1, using runs from era C, D and E, all at the NANOAOD level[1]. A summary of the data utilized is present in Table 1. In this study no type of data filtering based on the detector operational status was done for the sake of simplicity. This entails that, in some small samples, the detector could have been in a non-working status. However, the impact of this choice on the results is expected to be small and the implementation of the selection mechanism is planned for the future.

The data samples were selected using CMS Online Monitoring System (OMS), were it was filtered by total

[1] https://gitlab.cern.ch/cms-pro/validation/run3/multijet-trigger/efficiency

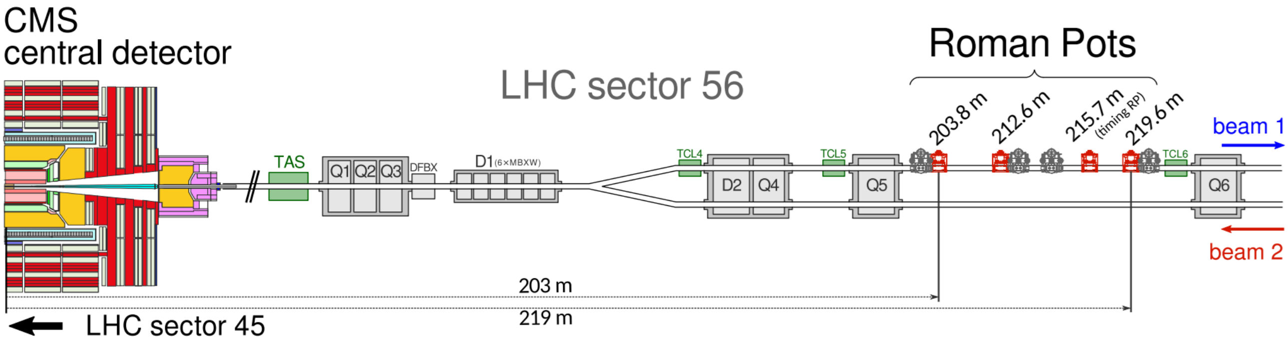


**Figure 3.** Side view of the PPS detector in LHC sector 56. The accelerator magnets are indicated in grey and the collimator system elements in green. The horizontal RPs, which constitute PPS, are marked in red. The vertical RPs are indicated in dark grey, and are part of the TOTEM experiment.

delivered luminosity, and its quality was verified through CMS Data Quality Monitoring (DQM) service. A query was then performed to CMS Data Aggregation Service (DAS) to retrieve the file names corresponding to that specific run and dataset.

| Dataset | Level | Era | Fill Number |
|---|---|---|---|
| JETMET | NANOAOD | C | 9579 - 9518 |
| | | D | 9653 - 9585 |
| | | E | 9708 - 9662 |

**Table 1.** Summary of the data used for the study of the HLT. For each era, the runs with the highest delivered luminosity were selected, up until 0.1 $fb^{-1}$. The total delivered luminosity was 7.7 $fb^{-1}$, 8.9 $fb^{-1}$ and 12.2 $fb^{-1}$, for era C, D and E, respectively. The total number of events was 120,631,863.

## 2.2 Data processing

One of the ways to determine the efficiency of the HLT is to compare it with a looser HLT and define the efficiency relatively to it. For this comparison, the HLT being studied will be referred to as tight HLT. The loose HLT only selects events with di-jet system average $P_T$ higher than 140 GeV. This approach works on the basis that all the events that pass the loose HLT and all the requirements of the tight HLT, applied offline, should be selected by the tight HLT, allowing the estimation of the relative efficiency of the tight HLT by performing the ratio between events that are correctly select and events that satisfy all HLT conditions but have not been selected.

Before performing this analysis for all eras, it was first done a study of the tight HLT efficiency in each era. The purpose of this study was to identify the presence of any differences between eras. A cutflow table for each era is registered in Table 4, 5 and 6, where there is also present the percentage of events that survive each cut with respect to the previous step (cut efficiency) and with respect to the initial number of events (cumulative efficiency).

From the results shown in the cutflow tables, it is clear that the efficiency of the tight HLT with respect to the loose HLT is not constant between eras. It can also be seen that the efficiency of the cuts in the number of jets and on the $\xi_{matching}$ was also not constant between eras. The fact that the efficiency of these cuts is not constant for all eras, tells that there is either something different happening with the protons or with the jets. Therefore, the properties of both the protons and the jets were studied and compared between each era, in an effort to identify what was dragging down the relative efficiency of the tight HLT in era E. From these properties, the only difference between eras was found in the $\eta$ of the jets, which is represented in figure 7.

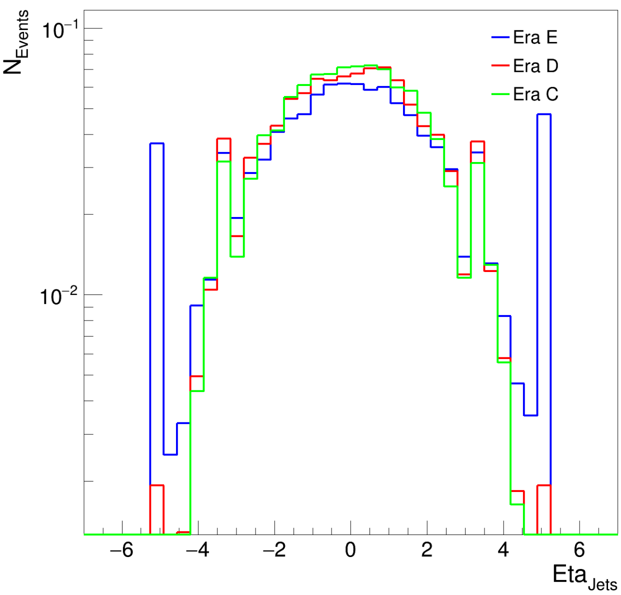


**Figure 7.** Normalized histogram of all the events that pass the loose HLT and the offline cuts for the tight HLT, but don't pass the tight HLT selection. Era E shows two peaks at $\eta = -5$ and $\eta = 5$, that are not present in the other eras. Era C and D have a similar $\eta$ distribution.

From the jet $\eta$ distribution it is visible that, in era E, there are 2 big peaks around $\eta = -5$ and $\eta = 5$. Further investigation for why these 2 peaks appear in era E is planned, but it is known that some settings of the hadronic forward (HF) calorimeters have been changed during this

| Triggers | Tight HLT | | Loose HLT | |
|---|---|---|---|---|
| Status | Pass | Fail | Pass | Fail |
| Initial number of events | 5801924 | 59656570 | 447464 | 65011030 |
| Number of jets >= 4 | 4214966 (72.65% \| 72.65%) | 42881643 (71.88% \| 71.88%) | 322877 (72.16% \| 72.16%) | 46773732 (71.95% \| 71.95%) |
| Di-jet system average $P_T$ >= 180 | 4084247 (96.90% \| 70.39%) | 24512654 (57.16% \| 41.09%) | 137162 (42.48% \| 30.65%) | 28459739 (60.85% \| 43.78%) |
| Number of tracks per RP <= 2 | 4084247 (99.99% \| 70.39%) | 14496968 (59.14% \| 24.30%) | 83217 (60.67% \| 18.60%) | 18497998 (65.00% \| 26.71%) |
| $\xi_{matching}$ <= 0.3 | 2122334 (51.96% \| 36.58%) | 179832 (1.24% \| 0.30%) | 2686 (3.23% \| 0.60%) | 2299480 (12.43% \| 3.32%) |
| Passing tight HLT | 100% | 0% | 1369 (50.97% \| 0.31%) | 2120965 (92.24% \| 3.06%) |

**Figure 4.** Cutflow table for all the runs from era C. In every step it is visible the number of events after each cut and inside the parenthesis is the cut efficiency and the cumulative efficiency, respectively. For era C, a relative efficiency of 50.97% was determined for the tight HLT.

| Triggers | Tight HLT | | Loose HLT | |
|---|---|---|---|---|
| Status | Pass | Fail | Pass | Fail |
| Initial number of events | 6440003 | 69297086 | 531352 | 75205737 |
| Number of jets >= 4 | 4925272 (76.48% \| 76.48%) | 51975842 (75.00% \| 75.00%) | 398248 (74.95% \| 74.95%) | 56502866 (75.13% \| 75.13%) |
| Di-jet system average $P_T$ >= 180 | 4814598 (97.75% \| 74.76%) | 30384819 (58.46% \| 43.85%) | 172935 (43.42% \| 32.55%) | 35026482 (61.99% \| 46.57%) |
| Number of tracks per RP <= 2 | 4814515 (99.99% \| 74.76%) | 16925070 (55.70% \| 24.42%) | 99096 (57.30% \| 18.65%) | 21640489 (61.78% \| 27.09%) |
| $\xi_{matching}$ <= 0.3 | 2957611 (61.43% \| 45.93%) | 245329 (1.45% \| 0.35%) | 3946 (3.98% \| 0.74%) | 3198994 (14.78% \| 4.00%) |
| Passing tight HLT | 100% | 0% | 2063 (52.28% \| 0.39%) | 2955548 (92.39% \| 3.70%) |

**Figure 5.** Cutflow table for all the runs from era D. In every step it is visible the number of events after each cut and inside the parenthesis is the cut efficiency and the cumulative efficiency, respectively. For era D, a relative efficiency of 52.28% was determined for the tight HLT.

era, which allowed for the detection of high $\eta$ jets. This change in the settings of the HF calorimeters leaves the question of why there are no peaks at $\eta = -4$ and $\eta = 4$ for era E. This difference between era E and era C, and D, might be an explanation for the low relative efficiency of the tight HLT determined in era E.
It was also observed that era C and D have a similar $\eta$ distribution, which is a possible explanation for the similar relative efficiency of the tight HLT between these eras.

Despite this difference, it was decided to merge all the data to improve the statistics and be able to investigate why from the events that passed the loose HLT and also satisfied the offline conditions of the tight HLT, only some were correctly selected by the tight HLT. The cutflow table with all the data from the 3 eras merged is present in Table 8.

With all the eras merged, it was determined an average relative efficiency for the tight HLT of 49.81%. A comprehensive look into the distributions of the events that were correctly selected by the tight HLT and those that were not

| Triggers | Tight HLT | | Loose HLT | |
|---|---|---|---|---|
| Status | Pass | Fail | Pass | Fail |
| Initial number of events | 7180718 | 113451145 | 733687 | 119898176 |
| Number of jets >= 4 | 6274245 (87.38% \| 87.38%) | 96438644 (85.00% \| 85.00%) | 622324 (84.82% \| 84.82%) | 50839236 (85.15% \| 85.15%) |
| Di-jet system average $P_T$ >= 180 | 6156553 (98.12% \| 85.74%) | 44958253 (46.62% \| 39.63%) | 275570 (44.28% \| 37.56%) | 50839236 (49.80% \| 42.40%) |
| Number of tracks per RP <= 2 | 6156506 (99.99% \| 85.74%) | 24036472 (53.46% \| 21.19%) | 150895 (54.76% \| 20.57%) | 30042083 (59.09% \| 23.42%) |
| $\xi_{matching}$ <= 0.3 | 3567045 (57.94% \| 49.68%) | 347452 (1.45% \| 0.31%) | 5160 (3.42% \| 0.70%) | 3909337 (13.01% \| 3.05%) |
| Passing tight HLT | 100% | 0% | 2442 (47.33% \| 0.33%) | 3564603 (91.18% \| 2.78%) |

**Figure 6.** Cutflow table for all the runs from era E. In every step it is visible the number of events after each cut and inside the parenthesis is the cut efficiency and the cumulative efficiency, respectively. For era C, a relative efficiency of 47.33% was determined for the tight HLT.

| Triggers | Tight HLT | | Loose HLT | |
|---|---|---|---|---|
| Status | Pass | Fail | Pass | Fail |
| Initial number of events | 19422645 | 242404801 | 1712503 | 260114943 |
| Number of jets >= 4 | 15414483 (79.36% \| 79.36%) | 191296129 (78.92% \| 78.92%) | 1343449 (78.45% \| 78.45%) | 205367163 (78.95% \| 78.95%) |
| Di-jet system average $P_T$ >= 180 | 15055398 (97.67% \| 77.51%) | 99855726 (52.20% \| 41.19%) | 585667 (43.59% \| 34.20%) | 114325457 (55.67% \| 43.95%) |
| Number of tracks per RP <= 2 | 15055268 (99.99% \| 77.51%) | 55458510 (55.54% \| 22.88%) | 333208 (59.89% \| 19.46%) | 70180570 (61.39% \| 25.29%) |
| $\xi_{matching}$ <= 0.3 | 8646990 (57.43% \| 44.52%) | 772613 (1.39% \| 0.32%) | 11792 (3.54% \| 0.69%) | 9407811 (13.41% \| 3.39%) |
| Passing tight HLT | 100% | 0% | 5874 (49.81% \| 0.34%) | 8641116 (91.85% \| 3.11%) |

**Figure 8.** Cutflow table for all the runs of all three eras used in the study. In every step it is visible the number of events after each cut and inside the parenthesis is the cut efficiency and the cumulative efficiency, respectively. Between all eras, an average relative efficiency of 49.81% was determined for the tight HLT.

selected but should have was performed. It was identified that the property with the biggest difference between these events was the di-jet system average $P_T$, which is represented in figure 9.

It is clear that the events that are not passing the tight HLT have a lower di-jet system average $P_T$ than those that are correctly selected. The average $P_T$ was computed and the values of 223 GeV and 252 GeV were obtained for the events that fail the tight HLT and for events that pass the tight HLT, respectively. This leads to the idea that the events failing the tight HLT have a lower di-jet system average $P_T$ than what is required at HLT level, which makes the HLT reject these events, but that the reconstruction increases the di-jet system average $P_T$ of these events, making them pass the offline cut.

## 3 Conclusions

In this study the relative efficiency of the tight HLT, averaged across three eras, was determined to be 49.81%. Some differences between the eras were spotted, mainly

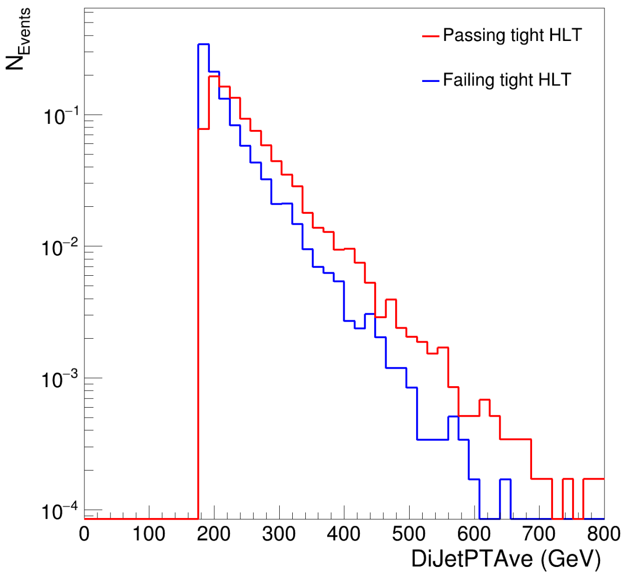


**Figure 9.** Normalized histogram of all the events that pass the loose HLT and the offline cuts for the tight HLT. Events that fail the tight HLT have a distribution that is shifted to lower values of average $P_T$, when compared to the distribution of the events that pass the tight HLT.

that the relative efficiency of the tight HLT with respect to the loose HLT was not constant between them. Also the efficiency of the cuts in the number of jets and on the $\xi_{matching}$ changed between eras.
The tight HLT condition for events with $\xi_{matching} < 0.3$ was observed to be the most restrictive, which pointed in the direction of something different with either the protons or the jets, prompting the search and discovery of the difference in the $\eta$ distribution of the jets, that is partially explained by the change of some settings in the HF calorimeters.

Trying to understand why some events that satisfied the tight HLT were not selected by it, it was found that the di-jet system average $P_T$ distribution of the events failing the tight HLT was shifted to lower values of $P_T$, with its average being 223 GeV, which is lower than the average of 252 GeV for the events passing the tight HLT. This suggests that these events have a lower di-jet system average transverse momentum at HLT level, making them get rejected by the tight HLT, but after the reconstruction this value increases and these events pass the offline cut.

## Acknowledgments

I would like to thank my supervisors Andrea Bellora and Gustavo Gil da Silveira for the unconditional support and the way they received and taught me in such a warm and passionate way. I would also like to thank the researchers from the CMS-PPS group for being so friendly and always ready to help. Last but not least, I want to thank the CERN Summer Student Programme organizers who made this wonderful learning experience possible.

## References

[1] The CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s}$=13 TeV”. In: *Journal of Instrumentation* (2020). DOI: 10.1088/1748-0221/15/10/P10017.

[2] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* (2008). DOI: 10.1088/1748-0221/3/08/S08004.

[3] The CMS Collaboration. “The CMS trigger system”. In: *Journal of Instrumentation* (2017). DOI: 10.1088/1748-0221/12/01/P01020.

[4] The CMS Collaboration and The TOTEM Collaboration. “Proton reconstruction with the CMS-TOTEM Precision Proton Spectrometer”. In: *Journal of Instrumentation* (2023). DOI: 10.1088/1748-0221/18/09/P09009.

[5] The CMS-TOTEM Collaboration. *CMS-TOTEM Precision Proton Spectrometer*. 2014. URL: https://cds.cern.ch/record/1753795.

[6] The TOTEM Collaboration. “Diamond detectors for the TOTEM timing upgrade”. In: *Journal of Instrumentation* (2017). DOI: 10.1088/1748-0221/12/03/P03007.

[7] *Technical proposal*. LHC technical proposal. Cover title : CMS, the Compact Muon Solenoid : technical proposal. Geneva: CERN, 1994. URL: https://cds.cern.ch/record/290969.