Trigger Studies for the Emerging Jets Analysis and Machine Learning for Tracker DQM at CMS Experiment

by

Guillermo A. Fidalgo Rodríguez

A thesis presented for the degree of

MASTER OF SCIENCE
in Physics
UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2024

Approved by:

Sudhir Malik, Ph.D. President, Graduate Committee	Date
Pablo Marrero, Ph.D. Member, Graduate Committee	Date
Samuel Santana Colón, Ph.D. Member, Graduate Committee Chairperson of the Department	Date

Abstract

The need of new physics has brought many exotic searches in hopes of answering the questions that the Standard Model has yet to address.

Keywords: [Emerging Jets, Dark matter, Quantum Chromodynamics, Machine Learning, Data Quality Monitoring]

Resumen

La necesidad de nueva física ha llevado a muchas búsquedas exóticas en esperanzas de contestar las preguntas que el Modelo Estándar de física de partículas no ha logrado responder.

Palabras claves: [Emerging Jets, Dark matter, Quantum Chromodynamics, Machine Learning, Data Quality Monitoring]

Acknowledgments

I wish to thank CERN, the CMS collaboration and the University of Puerto Rico at Mayagüez (UPRM), with special thanks to my advisor, Professor Sudhir Malik, and Dr. Scarlet Norberg for their support, guidance and patience. They have been invaluable to my education since my undergrad years at UPRM and my words here cannot express enough gratitude towards them.

Thanks to the Emerging Jets Analysis group: Sarah Eno, Long Wang, Yi-Mu Chen, Kevin Pedro, Kevin Stenson, Bill Ford, Keith Ulmer, Claire Savard. Special thanks to Dr. Alexx Perloff for his guidance and mentoring. Thanks to scientists and staff at Fermi National Lab in particular Dr. Gabriele Benelli for his phenomenal patience, leadership and guidance throughout the years.

A special thanks to all the wonderful people I have met in my early career as an aspiring physicist in HEP. Among these are those I met through the HEP Software Foundation (HSF) and IRIS-HEP whom have been influential for my profesional and academic training in analysis software. In particular, Kilian Lieret, Michel Villanueva, Wouter Deconick, Jim Pivarski, Henry Schreiner, Alexander Moreno and many more that I cannot possibly fit into to these pages.

I would like to thank National Science Foundation (NSF) and the Departemnt of Energy for funding this work.

Contents

List of Figures

Chapter 1

Introduction

The work for this thesis was performed with resources from European Organization for Nuclear Research (CERN)[1], the CMS Experiment[2], and the LHC Physics Center (LPC) at Fermi National Lab (FNAL). It was founded in 1954 and is located at the Franco-Swiss border near Geneva. At CERN, physicists and engineers are probing the fundamental structure of the universe. They use the world's largest and most complex scientific instruments to study the basic constituents of matter – the fundamental particles. The instruments used at CERN are purpose-built particle accelerators and detectors. Accelerators boost beams of particles to high energies before the beams are made to collide with each other or with stationary targets. Detectors observe and record the results of these collisions. The accelerator at CERN is called the Large Hadron Collider (LHC), the largest machine ever built by humans and it collides particles (mostly protons) at just 3 m/s under the speed of light. The process gives physicists clues about how the particles interact and provides insights into the fundamental laws of nature. Seven experiments at the LHC use detectors to analyze particles produced by proton-proton collisions. The biggest of these experiments, ATLAS and CMS, are general-purpose detectors designed to study the fundamental nature of matter and fundamental forces and to look for new physics or

[1] https://home.cern/about

[2] http://cms.web.cern.ch/news/what-cms or https://cms.cern/detector

evidence of particles that are beyond the Standard Model[3]. Having two independently designed detectors is vital for cross-confirmation of any new discoveries made with minimal bias. The other two major detectors ALICE and LHCb, respectively, study a state of matter that was present just moments after the Big Bang and a preponderance of matter than antimatter. Each experiment does important research that is key to understanding the universe that surrounds and makes us. Chapter 2 presents a basic description of the Large Hadron Collider and the CMS Detector. Chapter 3 presents a description and background on the Emerging Jets theory and analysis. Chapter 4 develops the technical aspects and usage of the HLT system in CMS. Chapter 5 gives a brief description of what is Data Quality Monitoring (DQM) and its importance for CMS, as well as describe the Machine Learning tasks developed for it. Chapter 6 summarizes the results and ongoing of this project.

[3] https://home.cern/about/physics/standard-model

Chapter 2

The CMS Experiment

The Compact Muon Solenoid (CMS) is a general-purpose particle detector designed to investigate various physical phenomena concerning the SM and beyond it, such as Supersymmetry, Extra Dimensions and Dark Matter. As its name implies, the detector is a solenoid that is constructed around a superconducting magnet capable of producing a magnetic field of 3.8 T. The magnetic coil is 13m long with an inner diameter of 6m, making it the largest superconducting magnet ever constructed. The CMS detector itself is 21m long with a diameter of 15m and it has a weight of approximately 14,000 tons. The CMS experiment is one of the largest scientific collaborations in the history of mankind with over 4,000 participants from 42 countries and 182 institutions. CMS is located at one of these points and it essentially acts as a giant super highspeed camera that makes 3D images of the collisions that are produced at a rate of 40 MHz (40 million times per second). The detector has an onion-like structure to capture all the particles that are produced in these high energy collisions. These particles are unstable and later decay to stable particles that are detected. The CMS detector was designed with the following features (as shown in Figure 2.1) :

1. A **magnet** with large bending power and high performance muon detector for good muon identification and momentum resolution over a wide range of momenta and

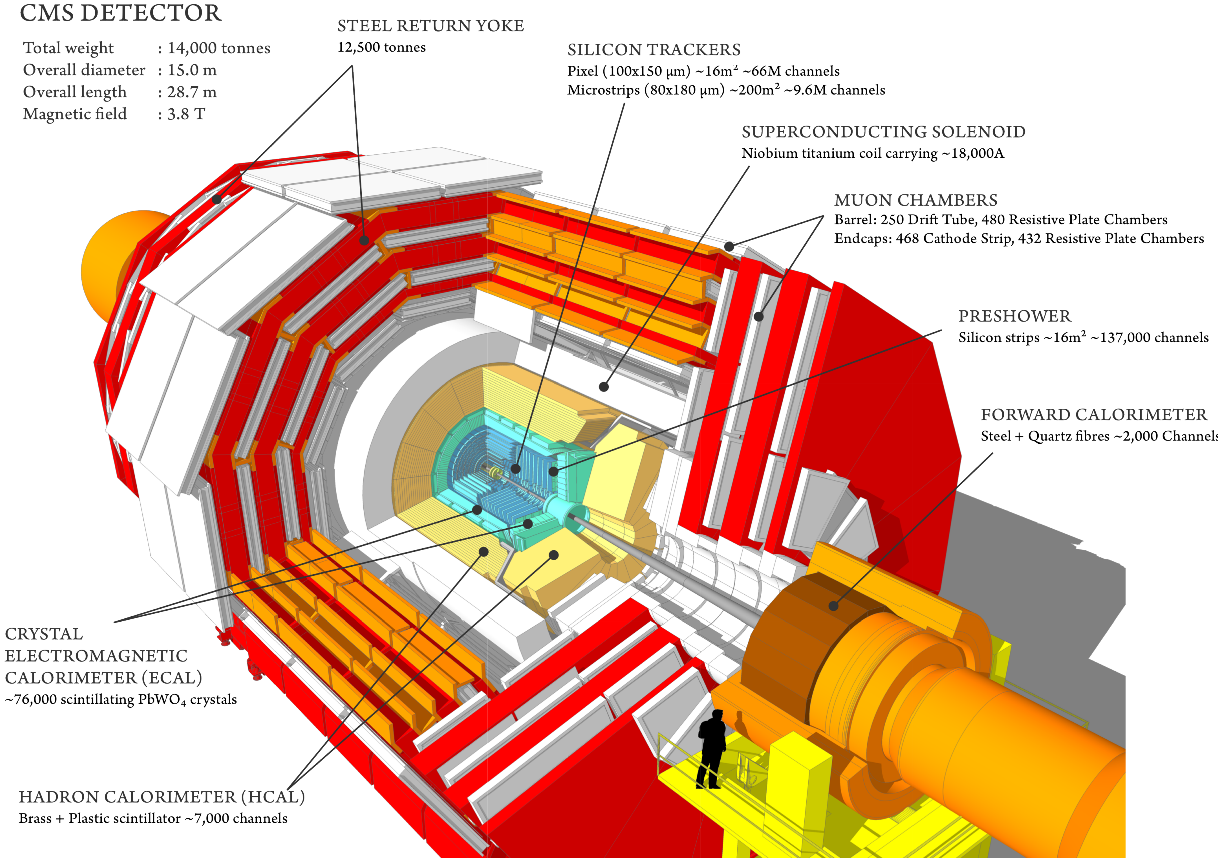

Figure 2.1: The CMS Detector

angles.

2. An **inner tracking system** capable of high reconstruction efficiency and momentum resolution requiring **pixel detectors** close to the interaction region.

3. An **electromagnetic calorimeter** able to provide good electromagnetic energy resolution and a high isolation efficiency for photons and leptons.

4. A **hadron calorimeter** capable of providing precise missing-transverse-energy and dijet-mass resolution.

A property from these particles that is exploited is their charge. Normally, particles produced in collisions travel in a straight line, but in the presence of a magnetic field, their paths are curved. Except the muon system, the rest of the sub-detectors lie inside a 3.8 Tesla magnetic field . Due to the magnetic field the trajectory of charged particle

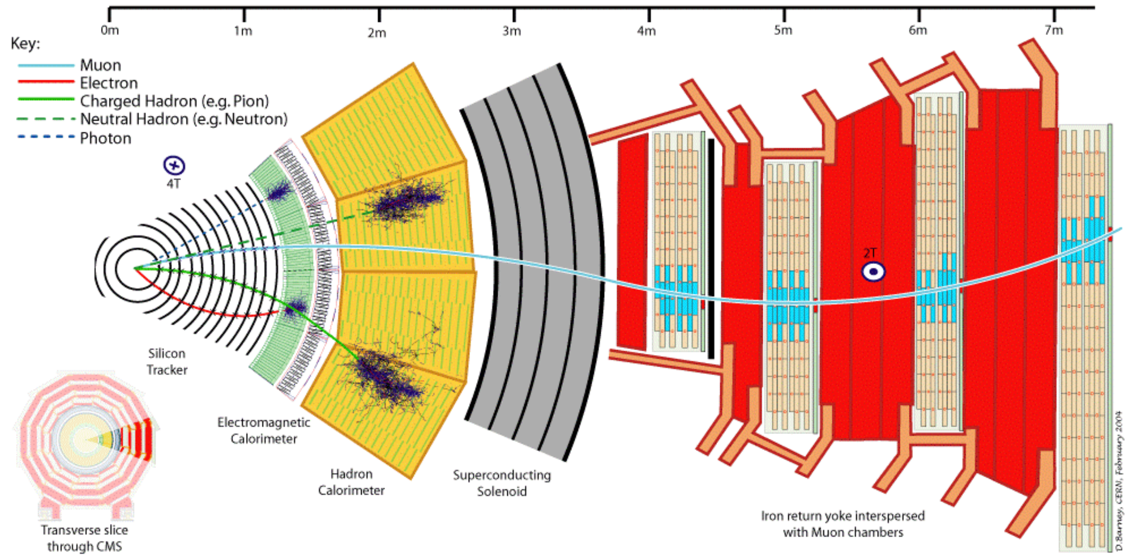

Figure 2.2: The trajectory of a particle traveling through the layers of the detector leaving behind it's signature footprint

produced in the collisions gets curved (as shown in Figure 2.2) and one can calculate the particle's momentum and know the type of charge on the particle. The Tracking devices are responsible for drawing the trajectory of the particles by using a computer program that reconstructs the path by using electrical signals that are left by the particle as they move. The Calorimeters measure the energy of particles that pass through them by absorbing their energy with the intent of stopping them. The particle identification detectors work by detecting radiation emitted by charged particles and using this information they can measure the speed, momentum, and mass of a particle. After the information is put together to make the "snapshot" of the collision one looks for results that do not fit the current theories or models in order to look for new physics. [1]

The project focusses specifically on data collected from one of the Calorimeters, - the Hadron Calorimeter (HCAL). The HCAL, as its name indicates, is designed to detect and measure the energy of hadrons or, particles that are composed of quarks and gluons, like protons and neutrons. Additionally, it provides an indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos (missing energy) . Measuring these particles is important as they can tell us if new particles such as the Higgs boson or supersymmetric particles (much heavier versions of the standard particles we know) have been formed. The layers of the HCAL are structured in a staggered fashion to prevent any gaps that a particle might pass through undetected. There are two main parts: the barrel and the end caps. There are 36 barrel wedges that form the last layer of the detector inside the magnet coil, there is another layer outside this, and on the endcaps, there are another 36 wedges to detect particles that come out at shallow angles with respect to the beam line.

Chapter 3

Emerging Jets (EJs)

3.1 Background information on EJs

The Emerging Jets concept arises from the paper by P. Schwaller [2] where it was proposed to search for the Emerging Jets signature in the Run 1 dataset of the LHC Experiments to set limits on a combination of parameter ranges.

Chapter 4

High-Level Trigger (HLT)

4.1 Developing the Algorithm

Chapter 5

Data Quality Monitoring (DQM)

Chapter 6

Conclusion

There are three major topics of research that were discussed in this dissertation: The simulation studies involving the counting of L1-stubs for the HL-LHC CMS Inner Tracker upgrade (stubs), the overall 2016 search for SUSY in the all-hadronic channel using a customized top-tagger (AnalysisChap) and the improvements made for the estimation of the Z$\rightarrow$ $\nu\bar{\nu}$+ jets background using an additional control region from γ+ jets events (estimation). These studies were explained in detail in their respective chapters and their individual results are provided. A summary of the most important results from each study is provided in this chapter.

6.1 L1 Stub Counting for the HL-LHC CMS Tracker Upgrade

Results from this study (detailed in stubs) reflect the overall effects that were expected beforehand. The removal of discs from the standard pixel geometry (consisting of 8 small and 4 large discs) results in a noticeable reduction of stub hits in the upgraded CMS Outer Tracker. This effect is specially apparent if the disc that is removed is closer to the interaction point, due to the much larger volume of particles that are present in this region.

Therefore, the reduction in stubs is more pronounced when a small disc is removed (as in the case of the $7s4l$ geometry) than if a large disc is removed (as in the $8s3l$ pixel geometry). The reason for this effect stems from the fact that as particles travel through the various layers of the Inner Tracker material, some of them are bound to interact with it, producing particles that did not originate from the initial proton-proton collision. The stubs produced via such processes are considered to be "fake" stubs. To confirm these findings, an additional study was conducted using a sample that was virtually indistinguishable from the standard pixel geometry, but with the second disc on the positive side "turned off" or "dead". The results from this study confirm the initial findings and shows that there is indeed a correlation between the average number of stubs detected in the Outer Tracker and the total amount of material present in the upgraded Inner Tracker. An important factor that needs to be taken into account when interpreting these results is the re-optimization of the disc positions after removing a disc in the different pixel geometries considered. This feature could provide a possible explanation as to why the $6s3l$ geometry, which has two less small discs than the standard geometry (and one less large one), was found to have less of an effect on the average number of stubs than the $7s4l$ geometry.

6.2 Search for SUSY in the All-Hadronic Channel

The analysis presented in AnalysisChap shows the results of a search for SUSY in the 0-lepton final state using a customized top-tagger. The data was obtained from proton-proton collisions at the CMS detector during 2016 with a total integrated luminosity of 35.9 fb^{-1} at a center-of-mass energy of 13 TeV. The search was conducted by specifying 84 non-overlapping regions of phase space with varying requirements on the N_{b}, N_{t}, p_{T}^{miss}, H_{T} and m_{T2} variables (SearchBinDef). Several dominant and non-dominant backgrounds were identified and estimated to account for all the majority of the processes that were seen in the collected data. The estimation procedures and their respective systematic and statistical uncertainties are discussed in backgrounds. The total background prediction vs. data for all 84 search bins (SearchBinResults) shows no statistically significant deviation from the predicted SM background. The biggest sources background were shown to be the $t\bar{t}$ and W+jets processes, followed by Z($\nu\bar{\nu}$)+jets, which were seen to be dominant in regions with a high p_T threshold. Meanwhile, the contributions from the QCD multijet and rare backgrounds are found to be nearly negligible in all of the 84 search bins. Exclusion limits were calculated from these results for each of the signal models used, by applying a binned likelihood fit on the data. The likelihood function was obtained for each of the 84 search regions as well as for each of the background data control samples from the product of the Poisson probability density function. Exclusion limits were placed on the top squark, gluino and LSP production cross-sections with a 95% confidence level (CL), calculated using a modified frequentist approach with the CL_s criterion and asymptotic results for the test statistic. The 95% CL exclusion limits obtained for the T2tt model, which consists of direct top squark production, excludes top squark masses up to 1020 GeV and LSP masses up to 430 GeV. For the T1tttt model, gluino masses of up to 2040 GeV and LSP masses up to 1150 GeV are excluded, with corresponding limits of 2020 and 1150 GeV for the T1ttbb model, 2020 and 1150 GeV for the T5tttt model, and 1810 and 1100 GeV for the T5ttcc model.

6.3 Estimation of the Z $\rightarrow \nu\bar{\nu}$+ jets Background with a Hybrid Method

estimation presents a different method (than the one briefly described in AnalysisChap) for estimating the total amount of Z$\rightarrow$ $\nu\bar{\nu}$+ jets events per search bin. The new method makes use of an additional "loose γ+ jets" control sample for the estimation of the $S_\gamma(N_j)$ shape correction factor, in addition to the "tight Z$\rightarrow$ $\mu\mu$+ jets" sample (used in the 2016 analysis) to calculate the R_{norm} correction factor (ZnunuSection). The γ+ jets sample was chosen to substitute the previously used "loose Z$\rightarrow$ $\mu\mu$+ jets" control region, due to its much higher cross-section and kinematic similarity to the Z$\rightarrow$ $\nu\bar{\nu}$+ jets process at high p_{T}. The implementation of this method sought to refine the results obtained in 2016 by reducing the statistical uncertainties that stem from the low branching fraction of Z$\rightarrow$ $\mu\mu$+ jets events. The γ+ jets control region was found to have a high purity and a low fake-rate in the regime of high p_{T} that was being studied. The N_j-dependent S_{γ}^{i} was obtained for each bin i in N_j from a comparison between data and the γ+ jets MC, after subtracting the other backgrounds from data and normalizing both samples to 1. This resulted in the S_{γ} scale factor plot depicted in NjetsCR, on the right. As can be seen from this plot, the statistical uncertainty of the various correction factors is small due to the large number of events available. Using these values, the R_{norm} normalization correction factor was obtained from the tight $\mu\mu$ control sample as $R_{norm} = 1.070 \pm 0.085$. Both of these scale factors were then applied to the final estimation of the Z$\rightarrow$ $\nu\bar{\nu}$+ jets background (results).

Bibliography

[1] CERN, *CMS Detector*, CMS, (2023) https://cms.cern/detector.

[2] P. Schwaller, D. Stolarski, and A. Weiler, “Emerging Jets”, JHEP **05**, 059 (2015).