

UNIVERSITY OF CALIFORNIA

Los Angeles

NO TITLE!?!?

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Physics

by

Tyler Christopher Lam

2023

© Copyright by
Tyler Christopher Lam
2023

The dissertation of Tyler Christopher Lam is approved.

Michalis Bachtis, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

1	Introduction	1
2	The Standard Model and Searches for Displaced Photons	2
2.1	The Standard Model	2
2.2	Motivations for Long Lived Scalar Bosons	2
3	Experimental Apparatus	3
3.1	The Large Hadron Collider	3
3.1.1	The High Luminosity LHC	6
3.2	The Compact Muon Solenoid	7
3.2.1	CMS Coordinate System	8
3.2.2	Charged Track Momentum Resolution	9
3.2.3	Inner Tracker	10
3.2.4	Electromagnetic Calorimeter	12
3.2.5	Hadronic Calorimeter	14
3.2.6	Muon Detectors	16
3.2.7	Trigger System	21
3.2.8	Event Reconstruction	25
4	Real Time Muon Reconstruction at the Compact Muon Solenoid	29
4.1	Introduction	29
4.2	The Kalman Filter Algorithm	30
4.3	The Kalman Barrel Muon Track Finder	31

4.3.1	Muon Trajectory Propagation	32
4.3.2	Covariant Matrices and the Kalman Gain	36
4.3.3	Firmware Implementation of the KBMTF	38
4.4	Performance of the Kalman Barrel Muon Track Finder	39
4.4.1	Trigger Efficiencies and Rate	40
4.4.2	Firmware and Emulator Agreement	43
4.5	The Phase-2 KBMTF	43
5	Search for a Long Lived Scalar Boson	47
5.1	Introduction	47
5.2	Data and Simulated Samples	47
5.2.1	Data Samples	47
5.3	Event Selection	48
References		49

LIST OF FIGURES

3.1	LHC luminosity report. Left: breakdown of the CMS integrated luminosity by year from 2010-2022. Right: peak luminosity from 2016-2018 data taking [1]	5
3.2	Diagram of the CERN accelerator complex during 2018 data taking. Protons begin as hydrogen atoms at Linac2 and collide at various detectors along the LHC at $\sqrt{s} = 13$ TeV	5
3.3	Timeline of runs and shutdowns leading to the HL-LHC, showing the improvements to \sqrt{s} and \mathcal{L} [2]. The extended year-end technical stops (EYETS) are shorter shutdowns generally for routine maintenance, although upgrades such as the CMS Pixel Detector upgrade can occur during this time as well.	6
3.4	Cutaway diagram of CMS showcasing major detector components [3].	7
3.5	Diagram showing the sagitta for a charged particle track	9
3.6	r - z slice of the CMS pixel detector [4]	12
3.7	r - z slice of the CMS silicon detector, detailing the four main components of the strip detector [5]	13
3.8	Geometry of the CMS ECAL showing the barrel, endcap, and preshower, adapted from [6]	15
3.9	r - z slice of the CMS HCAL showing the HB, HE, HF, and HO [7]	17
3.10	Diagram of the CMS Muon system [8]	18
3.11	Diagram of the CMS Drift Tubes [9]	19
3.12	Cut-away diagram of a single CSC module [10]	20
3.13	Diagram of the double RPC at CMS [11]	21
3.14	CMS Level-1 Trigger architecture during Run 2 [12]	23

3.15	Designed trigger architecture of the HL-LHC. The addition of the track primitives allows the computation of more advanced reconstruction algorithms at the L1 level [13].	24
4.1	The iterative process of propagation and updating a muon track through the KBMTF algorithm. The track properties at the innermost station are stored in order to trigger on muons not originating from the beamline [14]	33
4.2	Diagram showing a muon trajectory between two stations with the origin set at the outer station. The x-axis is oriented using the detector vertex and the outer station.	35
4.3	Calculation of phi propagation coefficients. Units are converted to their digitized form used in firmware calculations.	36
4.4	σ of vertical slice fits of the 2-dimensional histogram in figure 4.3a used to calculate uncertainty coefficients.	37
4.5	Comparison of efficiency vs cosmic track d_{xy} using cosmic data from the 2018 data taking run. The prompt KBMTF takes the reconstructed p_T using the vertex constrained measurement, while the displaced KBMTF uses the veretx unconstrained measurement. The displaced KBMTF shows substantial improvements to the previous BMTF [15]	41
4.6	Rate comparison calculated using 2017 zero bias data of the Phase I BMTF, vertex constrained (prompt) KBMTF, vertex unconstrained (displaced) KBMTF, and displaced KBMTF for displaced tracks.	42
4.7	Comparison of the track p_T between the firmware and emulator using data from a 2017 Z tag-and-probe sample. hwPt represents the integer value of p_T used in hardware. The emulator and firmware output show 100% agreement across 500000 events.	44

4.8 Performance of the L1 Barrel Muon algorithms using simulated samples with HL-LHC run conditions. The Track + KBMTF algorithm matches tracks to loose KBMTF muons, while the L1 Track + Muon Stubs algorithm matches tracks to muon stubs.	46
---	----

LIST OF TABLES

4.1 Comparison of FPGA resource utilization [14]. The KBMTF utilizes DSP cores for arithmetic operations to propagate and update tracks and LUTs for the propagation coefficients and approximate Kalman Gain, while the BMTF uses LUTs to assign momentum.	39
5.1 Table of integrated luminosity per year [16–18].	48

CHAPTER 1

Introduction

CHAPTER 2

The Standard Model and Searches for Displaced Photons

2.1 The Standard Model

2.2 Motivations for Long Lived Scalar Bosons

CHAPTER 3

Experimental Apparatus

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular collider spanning the border between France and Switzerland, based at the European Organization for Nuclear Research (CERN). The central features of the LHC are the superconducting rings, located about 100 m underground with a circumference of 27 km, designed to collide counter-rotating beams of protons or heavy ions at highly relativistic energies. Along the rings lie four major experiments: ATLAS, CMS, ALICE, and LHCb. Both ATLAS and CMS are general purpose detectors, designed to probe a wide range of physics including the Higgs boson, precision measurements of fundamental constants, and physics beyond the standard model (BSM). The remaining two experiments are more specialized; ALICE measures quark gluon plasma produced in heavy ion collisions and LHCb focuses on b -quark physics and CP violation.

Two prominent aspects of the LHC are the high center of mass energy of the proton beams, referred to using the Mandelstam variable \sqrt{s} , and high instantaneous luminosity \mathcal{L} , often referred to as just luminosity. High \sqrt{s} allows for the production of more massive particles, giving more access to possible BSM physics, while high luminosity is essential for measuring rare processes and precision measurements. A process with cross section σ will have a rate R given by

$$R = \mathcal{L}\sigma \tag{3.1}$$

The cross section σ is a measure of how probable a process is to occur, and is measured in

units of area. A frequent used unit for cross sections is a barn (b), corresponding to 100 fm^2 . Conversely, luminosity uses units of Hz/b . In cases where the relevant quantity is the total number of events, the integrated luminosity can be defined as $\mathcal{L}_{\text{int}} = \int \mathcal{L} dt$ to give

$$N_{\text{events}} = \mathcal{L}_{\text{int}} \sigma \quad (3.2)$$

The luminosity depends on the characteristics of the proton beam and can be written in terms of the operational parameters of the detector given by

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (3.3)$$

where N_b is the number of particles per bunch, n_b is the number of bunches per ring, f_{rev} is the LHC revolution frequency, γ_r is the Lorentz factor for the proton, ϵ_n is the transverse normalized beam emittance, β^* is the amplitude function at the collision point, and F is a geometric reduction factor based on the crossing angle of the two beams. The nominal design parameters of the LHC were intended to support a peak luminosity of 12 Hz/nb [19], but were exceeded by nearly twice that value of 20.7 Hz/nb during 2017 data taking and again with 21.4 Hz/nb in 2018. The LHC produced an integrated luminosity of 41.6 fb^{-1} in 2016, 49.8 fb^{-1} in 2017, and 67.9 fb^{-1} in 2018 for a total of 164 fb^{-1} during run 2 data taking. A more detailed breakdown of the LHC luminosity records can be seen in Fig. 3.1.

The protons used in collisions are extracted from hydrogen atoms by using ionizing electric fields to strip them of their electrons. Protons are first accelerated to an energy of 50 MeV through a linear accelerator Linac2 before entering the Proton Synchrotron Booster (PSB), where they reach a kinetic energy of 1.4 GeV . Next, the protons are accelerated by the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), where they are accelerated to 26 GeV and 450 GeV respectively. Finally, the beams are injected into the LHC where they undergo acceleration to 6.5 TeV , producing the desired center of mass energy of $\sqrt{s} = 13 \text{ TeV}$.

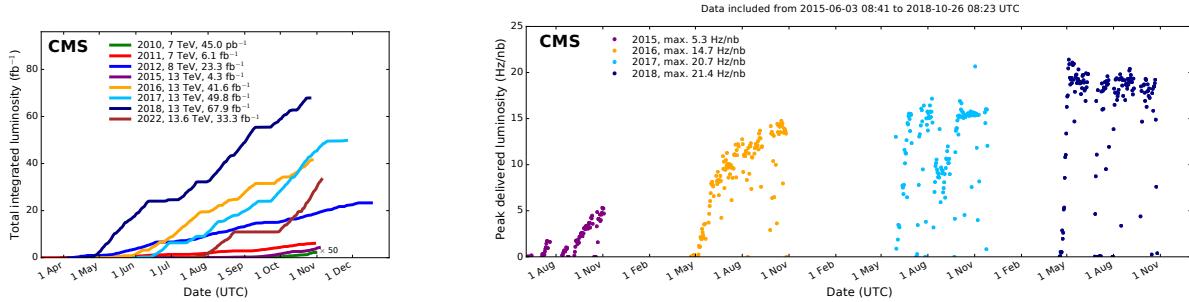


Figure 3.1: LHC luminosity report. Left: breakdown of the CMS integrated luminosity by year from 2010-2022. Right: peak instantaneous luminosity from 2016-2018 data taking [1].

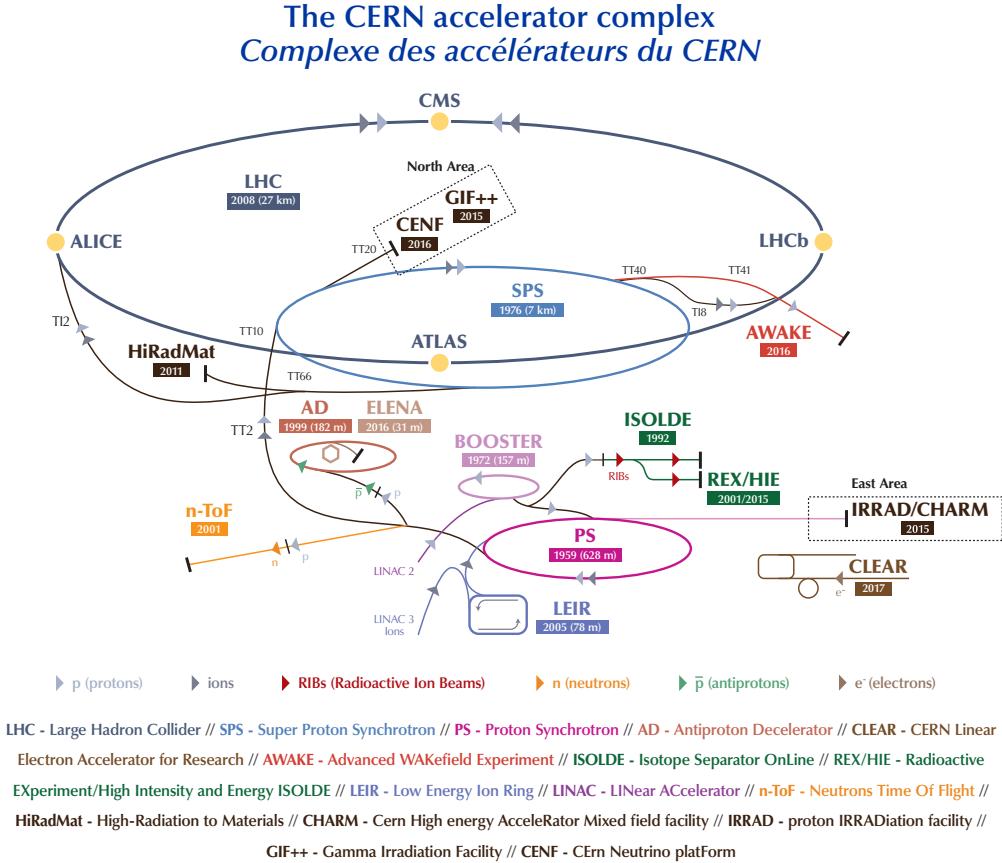


Figure 3.2: Diagram of the CERN accelerator complex during 2018 data taking [20]. Protons begin as hydrogen atoms at Linac2 and are accelerated in several stages to reach 6.5 TeV.

3.1.1 The High Luminosity LHC

The LHC is designed to go through periods of near continuous data taking, followed by long shut downs during which the detectors undergo various upgrades. As of the time of this writing, the LHC is currently in its third data taking run (Run 3), which is planned to operate from 2022-2025. The end of Run 3 marks the end of Phase-1 and the start of the third long shutdown (LS3), during which the LHC will receive major upgrades to increase the peak instantaneous luminosity up to 75 Hz/nb [14], over six times the designed peak luminosity and three times the peak luminosity achieved as of Run 2. The upgraded high luminosity LHC (HL-LHC) will gather substantially more data, which better allows experiments to probe rare processes and measure precise values predicted from the standard model.



Figure 3.3: Timeline of runs and shutdowns leading to the HL-LHC, showing the improvements to \sqrt{s} and \mathcal{L} [2]. The extended year-end technical stops (EYETS) are shorter shutdowns generally for routine maintenance, although upgrades such as the CMS Pixel Detector upgrade can occur during this time as well.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of the two general purpose detectors at the LHC, designed to reconstruct physics events from proton-proton collisions. It is a cylindrical apparatus with its major axis aligned with the proton beams from the LHC, consisting of several concentric layers of specialized detectors. The innermost layer consists of a silicon tracker, followed by an electromagnetic and hadronic calorimeter (ECAL and HCAL, respectively). Surrounding those is the superconducting solenoid, which provides a 3.8 T magnetic field to the trajectory of charged particles. Lastly, layers of muon chambers are interspersed with an iron return yolk, designed to contain the magnetic field lines near the muon chambers.

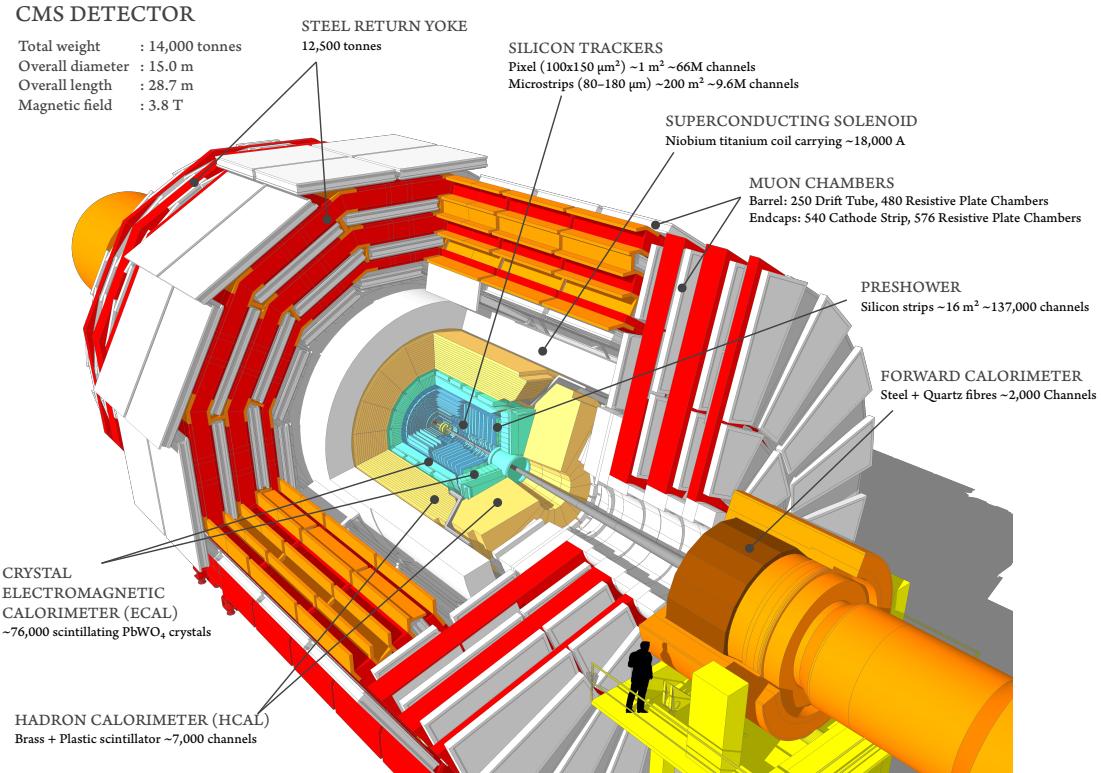


Figure 3.4: Cutaway diagram of CMS showcasing major detector components [3].

3.2.1 CMS Coordinate System

The origin of the CMS coordinate system is located at the center of the detector, at the nominal collision point of the proton beams. The $+\hat{x}$ axis points radially towards the center of the LHC, while the $+\hat{y}$ axis points vertically upward. This sets the $+\hat{z}$ direction along the beamline, counterclockwise along the LHC. Due to the cylindrical symmetry of CMS, coordinates in the transverse $\hat{x}\text{-}\hat{y}$ plane are commonly replaced with the radius r and the azimuthal angle ϕ , which is measured from the positive \hat{x} -axis. The polar angle θ is measured from the $+\hat{z}$ -axis, though in practice this variable is rarely used. It is more common to define the polar angle in terms of the pseudorapidity η , which is defined as

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) = -\ln \tan \left(\frac{\theta}{2} \right) \quad (3.4)$$

The motivation for using η instead of θ stems from a fundamental property of hadron colliders: the center of mass frame for particle production rarely coincides with the lab frame. Thus, when measuring the separation between particles, it is useful to define quantities that remain invariant under Lorentz transformations in the \hat{z} direction. The rapidity y (not to be confused with the Cartesian coordinate \hat{y}) is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (3.5)$$

It is trivial to show that differences in y remain invariant under Lorentz boosts along the \hat{z} -axis. In the highly relativistic limit, which is valid for most particles produced at the LHC, rapidity approaches pseudorapidity as $E \approx |\mathbf{p}|$. The angular separation between two particles can then be expressed as an invariant quantity under Lorentz boosts as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. One advantage of pseudorapidity is that η can be calculated using only geometric quantities of the detector, whereas y requires calculating both the energy and momenta of a particle, making pseudorapidity the natural choice for defining the polar

angle. η can range from $(-\infty, \infty)$, where $\eta = \pm\infty$ points directly along the $\pm\hat{z}$ axis. Higher values of $|\eta|$ are commonly referred to as "forward".

3.2.2 Charged Track Momentum Resolution

The 3.8 T magnet curves the trajectory of charged particles, which allows us to precisely determine their momentum. Using the geometry described in section 3.2.1, a particle with charge q and transverse momentum p_T travels in a circular trajectory with radius R when viewed in the $r\text{-}\phi$ plane. Per the Lorentz Force Law, these can be related by $p_T = qBR$. In particle physics, when working with an object of elementary charge, this is commonly rewritten as

$$p_T = 0.3BR \text{ [GeV/c]} \quad (3.6)$$

The track followed by a charged particle traveling over a length L can be described by the sagitta s shown in figure 3.5, defined by

$$s = R - \sqrt{R^2 - \frac{L^2}{4}} \quad (3.7)$$

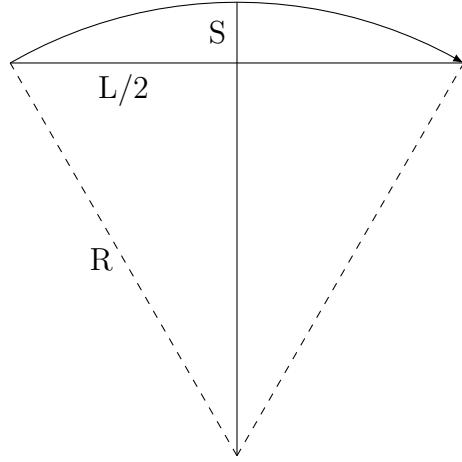


Figure 3.5: Diagram showing the sagitta for a charged particle track

For highly energetic particles, the length L is substantially smaller than the radius R , so

the sagitta can be approximated as

$$s \approx \frac{L^2}{8R} = \frac{0.3BL^2}{8p_T} \quad (3.8)$$

The relative uncertainty of the momentum is proportional to the uncertainty of the sagitta, the track length L, and the magnetic field B as

$$\frac{\delta p_T}{p_T} \propto \frac{p_T}{BL^2} \delta_s \quad (3.9)$$

The uncertainty on the sagitta δs is analogous to the hit resolution of a given detector. Therefore, a long lever arm and high magnetic field strength are crucial to precisely determine the momentum of energetic particles.

3.2.3 Inner Tracker

The CMS inner tracker is the first detector surrounding the primary interaction point (IP). Its purpose is to measure the tracks from charged particles as they curve in the strong magnetic field within CMS in order to calculate their momentum and reconstruct secondary vertices. As the closest detector to the primary IP, the tracker experiences the highest particle flux within CMS, and must have high enough granularity to distinguish the multitude of tracks. This high granularity also lends the inner tracker the best momentum resolution for charged tracks out of all the subdetectors comprising CMS, with a momentum resolution of 2.8% for a 100 GeV muon [21]. The high flux also makes the detector susceptible to radiation damage, so the tracker must be robust in order to maintain efficiency over the long operational period of the LHC.

The inner tracker utilizes silicon tracking modules composed of P-N type junctions to detect the location of charged particles. When an external voltage is applied to a module, particles passing through will deposit energy and create electron-hole pairs, which drift to

their respective electrodes and generate an electrical signal. The applied voltage is tuned such that only a small amount of energy is required to produce electron-hole pairs, which reduces the total amount of material required in order to minimize the energy loss of the charged particles. The inner layer of the tracker is composed of higher granularity pixel detectors, while the outer layer is composed of coarser silicon strips.

3.2.3.1 Silicon Pixel Detector

The pixel detector is comprised of 124 million silicon pixel sensors, each with an area of $100 \times 150 \mu\text{m}^2$ and a thickness of 300 μm . The barrel (BPIX) consists of four cylindrical layers of pixel sensors spanning from $z = -54 \text{ cm}$ to $z = 54 \text{ cm}$ at radii of $r = 2.9, 6.8, 10.9,$ and 16 cm. The endcap (FPIX) consists of three layers located at $z = \pm 29.1, \pm 39.6,$ and $\pm 51.6 \text{ cm}$, which when combined with the barrel nets a total sensitive area of 1.85 m^2 with coverage up to $|\eta| < 2.5$. With regards to detector performance, the pixel detector provides spatial resolution of $9.5 \mu\text{m}$ in the $r\text{-}\phi$ direction and $22.2 \mu\text{m}$ in the z direction, as well as a hit efficiency of $> 99\%$ in each layer at nominal LHC luminosity, with the innermost BPIX layer dropping to 97.5% at peak run 2 efficiency [4].

3.2.3.2 Silicon Strip Detector

The silicon strip detector surrounds the pixel detector, and is segmented into a tracker inner barrel (TIB), tracker outer barrel (TOB), tracker inner disks (TID), and tracker endcap (TEC). The TIB consists of four layers covering $|z| < 65 \text{ cm}$ and radius $25.5 \text{ cm} < r < 49.8 \text{ cm}$. The endcaps of the TIB are covered by the TID, consisting of three disks covering $90 < |z| < 90 \text{ cm}$. Surrounding the TIB is the TOB, with six layers spanning $|z| < 188 \text{ cm}$ and $60.8 < |r| < 108 \text{ cm}$. Lastly, both the TOB and TID are closed by the TEC, which covers radii ranging from $22 < r < 113.5 \text{ cm}$ and stretches from $124 < |z| < 280 \text{ cm}$.

The strip detector modules function similarly to the pixel modules, but utilize much

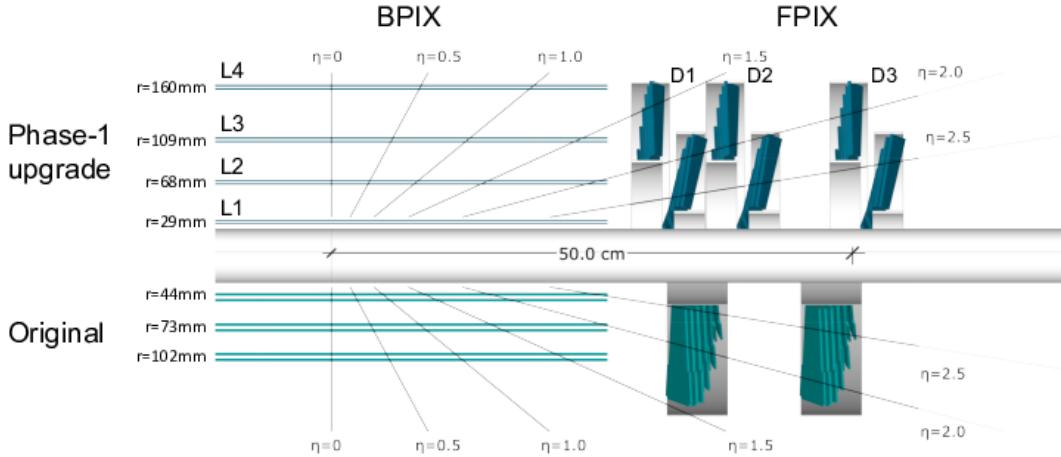


Figure 3.6: r - z slice of the CMS pixel detector comparing original design (bottom, teal) with the Phase-I upgrade implemented in 2016/17, which added one layer to both the FPIX and BPIX (top, blue) [4].

larger silicon strips due to the expected reduced particle flux compared to the pixel detector. Each strip detector is composed of $300\ \mu\text{m}$ thick micro-strip sensors with pitch varying from $80 - 205\ \mu\text{m}$ to form $7 - 12.5\ \text{cm}$ long strip modules. Overall, the strip detector uses 15,148 modules covering an area of $200\ \text{m}^2$ up to $|\eta| < 2.5$.

3.2.4 Electromagnetic Calorimeter

The CMS Electromagnetic Calorimeter (ECAL) is designed to stop and capture the energy of photons and electrons. When a high energy photon or electron enters the ECAL, it primarily interacts with the material in the detector through e^+e^- pair production or bremsstrahlung, respectively. These processes create more photons and electrons that result in an electromagnetic cascade until the particles lose enough energy that ionization and compton scattering processes begin to dominate. This process is known as an electromagnetic shower. The light from these showers is measured using photodetectors to determine the initial energy of the incident particle with resolution given by

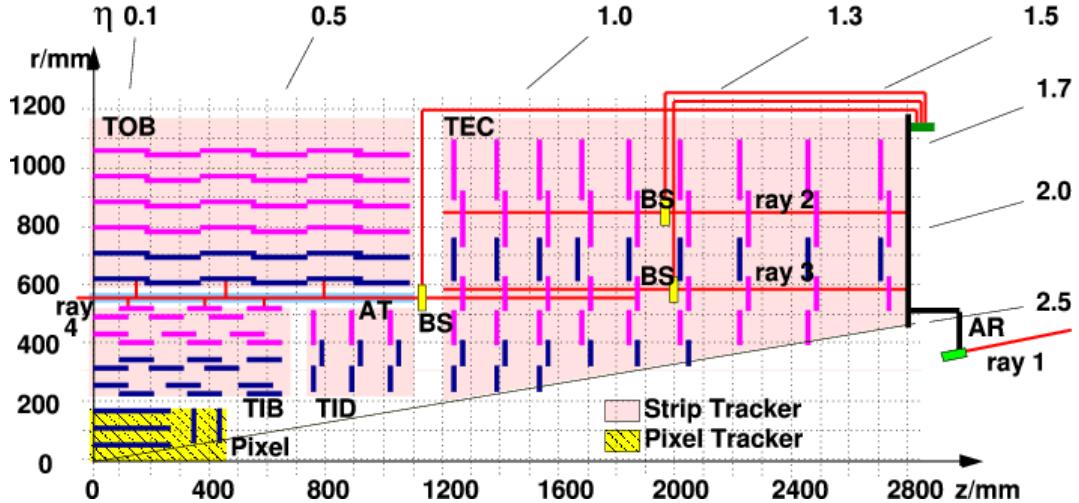


Figure 3.7: r - z slice of the CMS silicon detector, detailing the four main components (TIB, TOB, TID, and TEC) of the strip detector. Pixel detector is shown in its Run-I configuration [5].

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E/\text{GeV}}}\right)^2 + \left(\frac{12\%}{E/\text{GeV}}\right)^2 + (0.3\%)^2 \quad (3.10)$$

The first term arises from statistical fluctuations in the electromagnetic shower and light production, which is a poisson process and therefore proportional to $1/\sqrt{E}$. The second term is due to the noise in the detector electronics. The last constant term is due to imperfections in the detector such as non-uniformity, radiation damage, and calibration uncertainty. The material used for an electromagnetic calorimeter can be characterized by the radiation length (X_0) and molière radius (R_M). The radiation length is the average distance an electron will travel before its energy is reduced by a factor of $1/e$, and is roughly proportional to A/Z^2 , where A is the atomic mass and Z is the atomic number. The molière radius is directly proportional to the radiation length and measures the spread of an electromagnetic shower in the transverse direction. Both are generally measured in units of $[\text{g cm}^{-2}]$, but can be divided by the density of the material to obtain a distance in $[\text{cm}]$. A low radiation length and Molière radius ensure all of the energy is captured in the detector and is contained to a

small area to precisely determine the position of an incident particle. It is for this reason that the ECAL is composed of lead-tungstate (PbWO_4): a dense, high Z scintillator crystal with $X_0 = 7.39 \text{ g/cm}^2$ (or 0.89 cm after dividing by the density of PbWO_4) and $R_M = 2.2 \text{ cm}$ [22].

The ECAL is constructed from $61200 + 15000$ lead-tungstate crystals in the barrel + endcap region. Barrel crystals have a cross sectional area of $2.2 \times 2.2 \text{ cm}^2$ to match the molière radius and a depth of 23 cm (or $25.8 X_0$), and are arranged in rings along the ϕ direction, tilted to align with lines of constant η . The barrel has an inner radius of 129 cm and covers an eta region up to $|\eta| < 1.479$. Endcap crystals have a slightly higher cross sectional area of $2.6 \times 2.6 \text{ cm}^2$ and a slightly lower depth of 22 cm , and are arranged in an x - y grid. The endcap provides the remaining coverage from $1.479 < |\eta| < 3.0$. A diagram showing the geometry of the ECAL can be seen in figure 3.8. The size and alignment of the crystals is designed to contain showers from incident photons/electrons within a 2×2 grid of crystals.

One common process in CMS is the production of π^0 s, which decay to two photons. When the two photons deposit their energy into the ECAL, they can fake a signal from a single high energy photon. In order to identify these background events, a higher granularity detector known as a preshower is placed before the ECAL endcap. It consists of a layer of lead absorber to initiate the shower, followed by silicon strip sensors to measure the tracks within the shower. The lead has a thickness of 1.57 cm (or $2.8X_0$), which is thick enough to reliably cause a shower while only causing energy loss of a few percent before the shower can reach the ECAL. The silicon strips are similar to those in the inner tracker, with a pitch of 1.9 mm [23].

3.2.5 Hadronic Calorimeter

The CMS Hadronic Calorimeter (HCAL) is a sampling calorimeter surrounding the ECAL. Unlike the ECAL, which is a homogeneous calorimeter where the crystals both induce showers and scintillate, the HCAL is composed of alternating layers of dense absorber and scintilla-

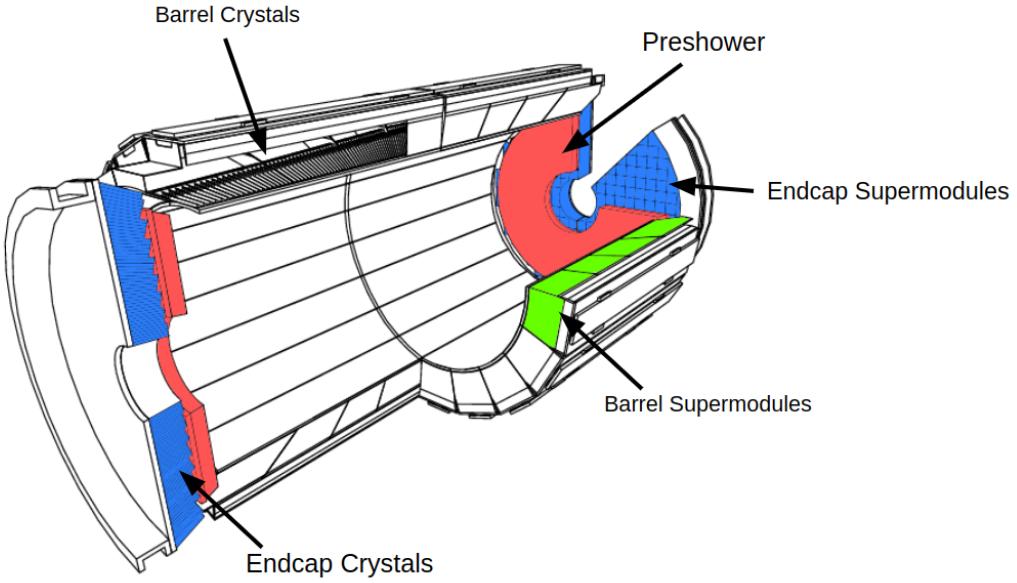


Figure 3.8: Geometry of the CMS ECAL showing the barrel, endcap, and preshower, adapted from [6].

tor. The absorber induces hadronic showers - cascades of hadrons resulting from inelastic scattering off a target nuclei. Particles in hadronic showers then pass through the scintillator, which produces light that gets transmitted through wavelength shifting fibers and read out by photodetectors, before repeating this process with the next layer of absorber/scintillator. The measured energy can be summed in sequential layers of detectors (referred to as "towers") to calculate the total energy of the incident hadron. As an additional complication, several hadrons decay to electrons or photons, which will create electromagnetic showers, requiring the scintillator to have good response to electromagnetic interactions. The HCAL must be hermetic, as a complete picture of the energy from proton collisions is required to infer the production of particles like neutrinos, which are otherwise invisible to the detector.

The relevant property for a hadronic absorber is the nuclear interaction length λ_I , which is the mean distance a hadron will travel before undergoing an inelastic nuclear interaction. Like the radiation length, λ_I is measured in $[g\text{ cm}^{-2}]$ and can be divided by the density to

obtain a distance in [cm]. It is proportional to the density of nuclear matter, which goes as $A^{1/3}$. The HCAL uses a combination of brass and stainless steel, which have an interaction length of ~ 16.5 cm [24].

The HCAL consists of four subsystems: an inner barrel (HB), an outer barrel (HO), an endcap (HE), and a forward calorimeter (HF). The HB covers $|\eta| < 1.3$ with 36 azimuthal wedges, each containing 17 layers of alternating brass absorber and scintillator. These wedges begin at a radius of 1.806 m and are 89 cm thick, resulting in a minimum of $5.3 \lambda_I$ at $\eta = 0$, increasing to $10 \lambda_I$ and $\eta = 1.2$. The HE consists of two disks, also formed of 36 wedges with 19 layers of brass absorber and scintillator, totaling $10 \lambda_I$ in order to contain highly boosted, forward showers resulting from quarks and gluons.. The HE extends the coverage of the HCAL up to $|\eta| < 3.0$. The HO is located outside the magnetic coil to catch high energy (> 100 GeV) hadrons that escape the HB. It consists of five rings in eta, covering up to $|\eta| < 1.2$, and relies on the iron solenoid to induce hadronic showers. The HF covers the most forward region of $3 < |\eta| < 5$. This zone has the highest exposure of radiation, and so the detector must be sensitive only to the highest energy particles and robust against damage from heavy radiation. It uses steel as the absorber and quartz fibers to scintillate through Cherenkov radiation. The four subsystems provides a minimum of 10λ to capture incident hadrons.

3.2.6 Muon Detectors

The muon chambers lie outside the magnetic solenoid as the outermost subsystem in CMS. Due to the distance and amount of material between them and the primary IP, they are the lowest occupancy detectors, with few particles other than muons able to reach them before decaying or stopping in the inner subsystems. Although they lie outside the solenoid, the iron yolk pulls the magnetic field back to curve the trajectory of muons, which is required calculate their momenta. The detector subsystem is composed of three types of gaseous detection chambers: drift tubes (DTs), resistive plate chambers (RPCs), and cathode strip

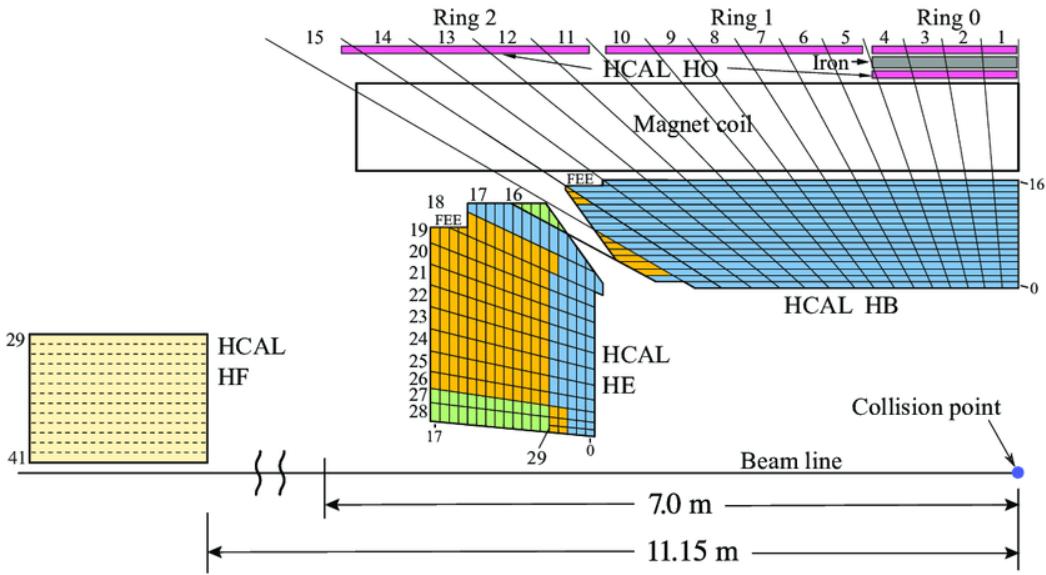


Figure 3.9: r - z slice of the CMS HCAL showing the HB, HE, HF, and HO [7].

chambers (CSCs). A diagram showing the configuration of these detection chambers can be seen in figure 3.10.

The operating physics principles of all three chambers are similar: muons pass through a gas mixture and ionize the gas, knocking loose electrons. Anodes and cathodes create an electric field inside the gas chamber, causing the electrons to drift to the anode where they can be measured and read out by electronics. Electrons that are excited but not ionized from the incident muon will emit photons as they transition back to a lower energy level, thus a quenching gas is used to absorb these photons to prevent further cascades. The specifics of the three detectors are described in sections 3.2.6.1-3.2.6.3.

3.2.6.1 Drift Tubes

Drift tubes are used in the barrel of the muon system, covering $|\eta| < 1.2$. They consist of long chambers filled with an mixture of Ar/CO₂, with an anode wire running lengthwise through the center and cathode strips along the edges. Electrode strips along the outer edge of the chamber shape the electric field lines to linearize the drift velocity of electrons

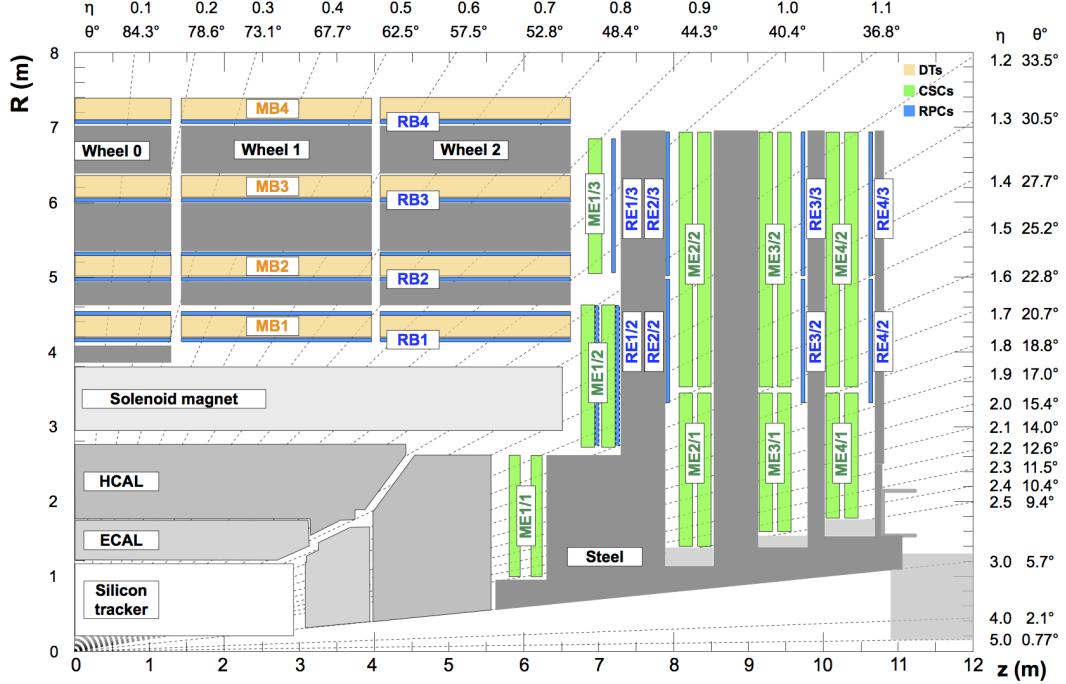


Figure 3.10: r - z cross section of the CMS muon detector subsystem, configured for the 2016–2018 data taking run. The barrel consists of five wheels of DTs and RPCs in η , each wheel consisting of four layers in r and 12 sectors in ϕ , covering up to $|\eta| < 1.2$. The endcap consists of four layers of DTs and RPCs in z [8].

throughout the chamber. When a muon passes through the chamber, ionized electrons follow the electric field lines towards the anode wire. Due to the known field inside the chamber, the drift time can be used to calculate the transverse distance from the wire to the incident muon with a resolution of $250 \mu\text{m}$. The DTs are arranged in superlayers, which consist of 4 layers of DTs, with subsequent superlayers alternating in the r - z and r - ϕ direction in order to obtain the muon trajectory/curvature in both the η and ϕ direction. Figure 3.11 shows a diagram of a DT superlayer and a single DT cell.

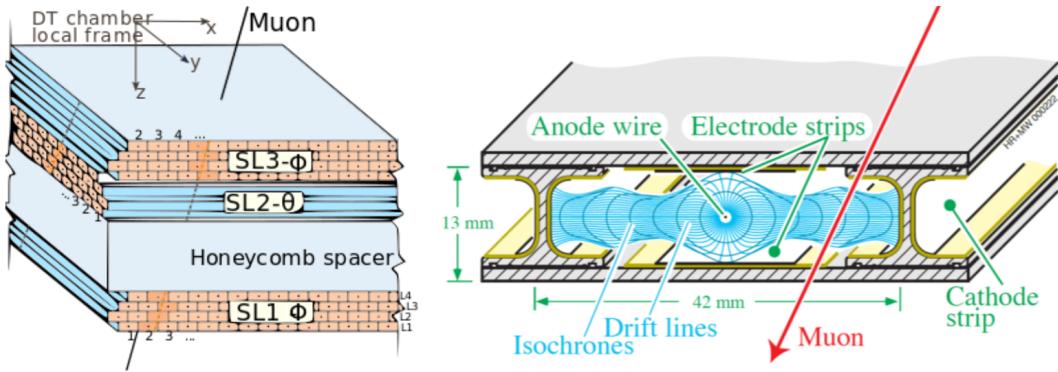


Figure 3.11: Drift tube superlayers are aligned in alternating orientations in order to provide a complete picture of muon trajectories (left). A cross section of a single drift tube chamber (right). Electrons ionized from the muon drift to the anode wire to produce a measurable current [9].

3.2.6.2 Cathode Strip Chambers

Cathode strip chambers are used in the endcaps of the muon system to provide coverage from $0.9 < |\eta| < 2.4$. CSC modules are trapezoidal chambers filled with a mixture of Ar/CO₂/CF₄ and arranged in disks to cover the muon endcap. Each module is composed of six layers of radial cathode strips, which have high granularity in ϕ , and transverse anode wires which provide granularity in r . Similar to the DTs, muons passing through CSC chambers ionize the gas molecules, knocking loose electrons which drift to the anode wires. The charge from these electrons can be measured and provides the position of the muon in r . Simultaneously,

the electrons induce a charge in the cathode strips, which provides the position of the muon in ϕ . One chamber provides a spatial resolution of between $40 - 50 \mu\text{m}$ with a time resolution of 3 ns.

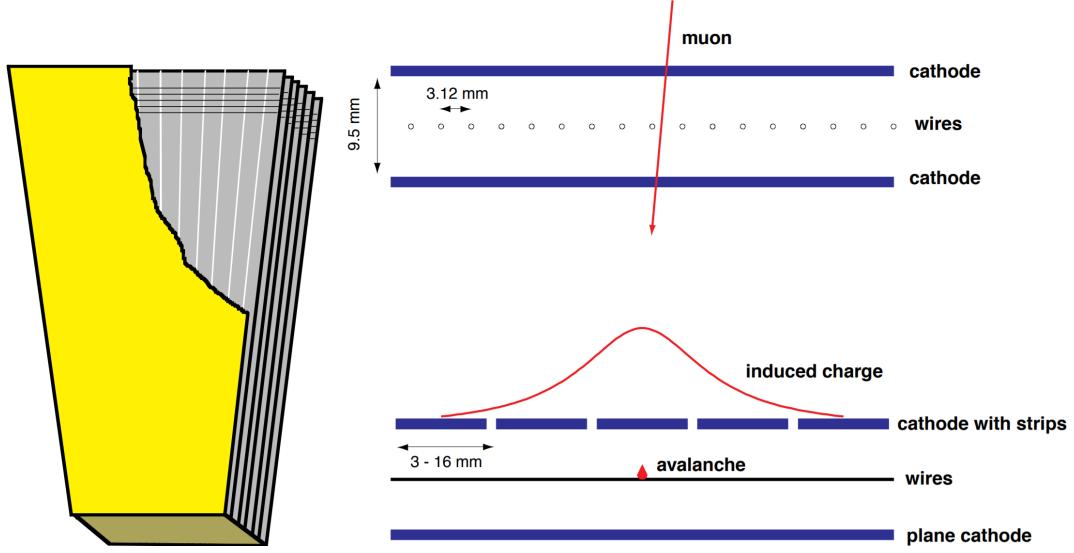


Figure 3.12: Cut-away diagram of a single CSC module showing the six layers of radial cathode strips and transverse anode wires (left). Muons ionize the gas chamber, causing an avalanche of electrons on the anode wires and inducing a charge on the cathode strips (right) [10].

3.2.6.3 Resistive Plate Chambers

RPC detectors are used in conjunction with the DTs and CSCs in both the barrel and endcap. One RPC consists of two high resistivity plastic plates: one positively charged anode and a negatively charged cathode which create a strong, uniform electric field between the two. The space between the two plates is filled with a mixture of 95% $\text{C}_2\text{H}_2\text{F}_4$ and 5% iso- C_4H_{10} . Incident muons ionize the gas between the two plates, which cause a cascade of electrons towards the anode plate due to the strong electric field. External metal strips lie outside the anode in order to measure the avalanche of electrons. The coarse nature of the strips means the RPC spatial resolution is lower than that of the CSCs and DTs at around

1 cm, but the time resolution is much faster on the order of 1 ns. For this reason they are used to supplement the CSC and DTs to aid with triggering and accurate timing.

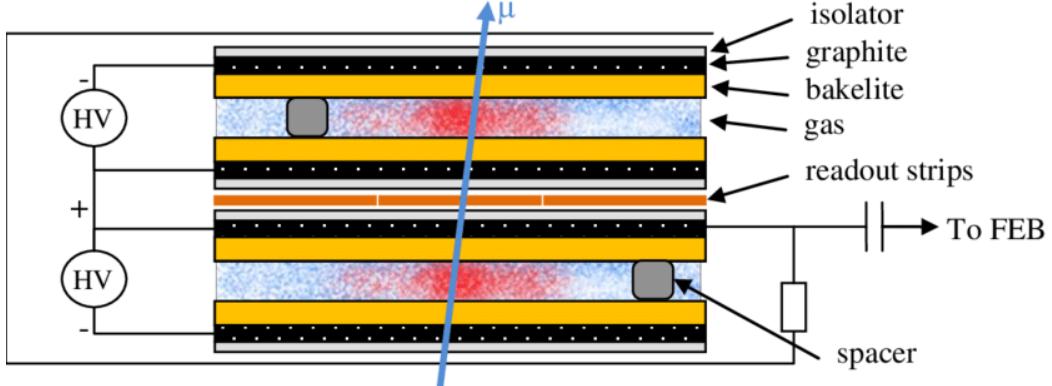


Figure 3.13: Diagram of the double RPC at CMS. Incident muons ionize the gas, causing a cascade of electrons towards the readout strips [11].

3.2.7 Trigger System

The high luminosity of the LHC yields an incredibly high rate of information to store from the various detectors. Each event contains approximately 1 MB of data to store, which would take bandwidth and storage space beyond current capacities to save given the 40 MHz collision rate. Additionally, many events are soft collisions that do not contain interesting physics worth storing for further analysis. In order to reduce the rate of events and select only the most interesting collisions, CMS employs a two level trigger system. The first level is known as the Level-1 (L1) trigger and uses hardware to identify physics objects in real time. The second is known as the High-Level trigger (HLT), consisting of a conventional CPU farm which performs more advanced calculations.

3.2.7.1 Level-1 Trigger

The L1 trigger utilizes special processors such as Field Programmable Gate Arrays (FPGAs) in order to process coarse information from the calorimeters and muon chambers in real time. The coarse detector information, known as trigger primitives, is subjected to several algorithms to create physics objects (e.g. muons, electrons/photons, etc). Objects passing a predetermined set of criteria are called trigger "seeds", with events that satisfy at least one seed getting passed to the HLT. A diagram of the L1 trigger architecture can be seen in figure 3.14. This entire decision making process must take place in $3.2 \mu\text{s}$ after a bunch crossing. As this is greater than the time between bunch crossings, the L1 trigger must be pipelined in order to operate continuously, while the full data is kept in a rolling buffer and read out only if the event passes a trigger seed. The L1 trigger was designed to accept events at a maximum rate of about 100 kHz, meaning only one in every four hundred events is passed to the HLT.

The L1 trigger will undergo several changes during LS3 as a result of the HL-LHC upgrade discussed in section 3.1.1. The main consequence of higher luminosity is an increase in additional proton collisions along the beamline per bunch crossing, known as pileup. Interesting physics events generally stem from a single hard scattering interaction, known as the primary vertex (PV), with lower energy pileup collisions creating background that can obscure the physics from the PV. The L1 trigger is well equipped for the average 40 pileup for Run 3, but will require upgrades in order to accommodate the expected 200 pileup for the HL-LHC. As part of the L1 trigger update, the maximum bandwidth will be increased from 100 kHz to 750 kHz and the trigger decision latency will be increased from $3.2 \mu\text{s}$ to $12.5 \mu\text{s}$. Additionally, new tracker primitives will be available as inputs to the calorimeters and muon triggers. These primitives allow the precise p_T , η , and ϕ information from tracks to be associated with calorimeter clusters and muon primitives, greatly increasing the precision and efficiency of L1 reconstruction. The proposed trigger architecture for the HL-LHC can be seen in figure 3.15.

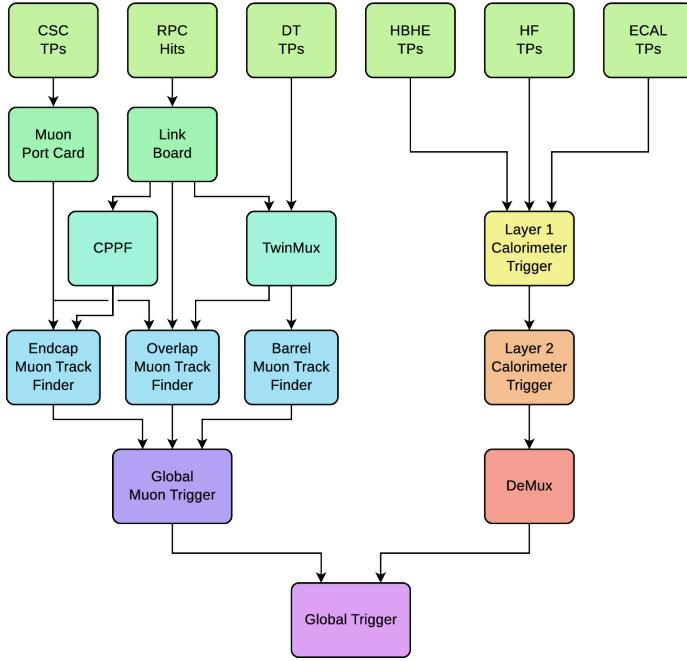


Figure 3.14: Block diagram of the CMS Level-1 Trigger architecture during Run 2 [12]. Muon objects found in the barrel, overlap, and endcap regions are ranked and sorted using their momentum and quality criteria before being passed to the Global Muon Trigger (GMT), which cleans duplicate tracks and selects the top eight muons. The muons are then passed to the global trigger, which combines the muons with objects created from the calorimeters and makes the final decision to pass the event to the HLT.

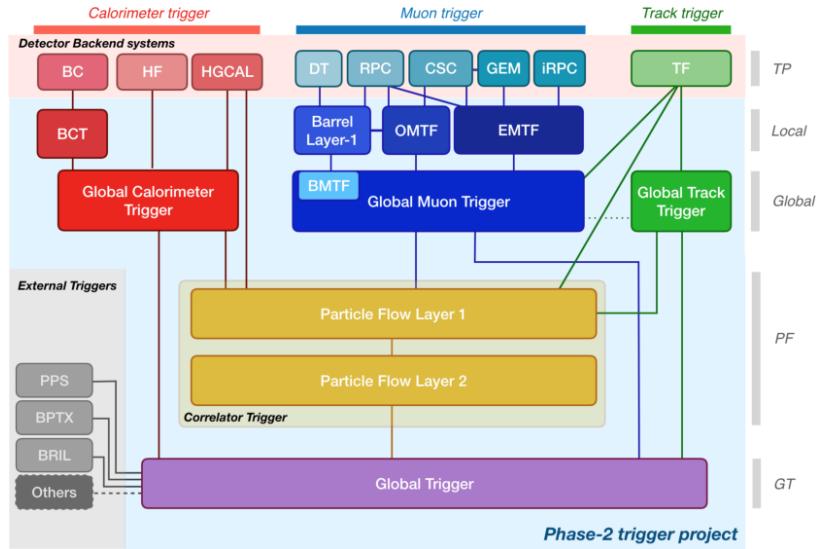


Figure 3.15: Designed trigger architecture of the HL-LHC. The addition of the track primitives allows the computation of more advanced reconstruction algorithms at the L1 level [13].

3.2.7.2 High Level Trigger

The HLT receives the full precision detector data for events passing the L1 trigger, and is designed to reduce the final event rate to approximately 100 Hz, five orders of magnitude less than the LHC bunch crossing rate. Information from the tracker allows the HLT algorithms to reconstruct a more complete picture of an event, as tracker information can be matched to both calorimeter hits and muon tracks to provide more detail to reconstructed physics objects. As with the L1 trigger, the HLT processes data through a series of trigger paths, which sequentially reconstruct and filter physics objects. In order to optimize computation time, the HLT uses L1 trigger seed to inform which trigger paths to use, and structures the paths in order of increasing complexity. For example, electrons are first created using only calorimeter data, then matched using coarse hits in the tracker, before being reconstructed using full tracker information. Events passing the HLT are then recorded permanently for offline reconstruction and analysis.

3.2.8 Event Reconstruction

Events that pass the HLT then undergo a complete event reconstruction, where data from all sub-detectors is combined to identify each physics object created in a collision. This section provides an description of the relevant objects for this dissertation, namely muons, electrons, and photons, as well as a brief description of the particle flow (PF) algorithm used for global event reconstruction.

3.2.8.1 Tracks and Calorimeter Clusters

Raw detector information taken from the inner tracker and calorimeter are used to create tracks and clusters, which are the building blocks used to create a global event reconstruction. Tracks are built using a Kalman Filter algorithm to fit particle trajectories to hits in the pixel and strip detectors. This process begins with a track seed which is built from hits in the pixel detector compatible with a charged particle track. The initial seed track is propagated outward to collect nearby hits in successive layers of the tracker. Once the seed and hits are collected, the track is iteratively fit to determine the final properties of the charged particle, namely the origin, p_T , and direction. Various quality cuts are applied based on the number of hits in order to reject fake or misidentified tracks.

Clusters are seeded from individual cells in the calorimeters with an energy above a preset threshold. Adjacent cells are added to the cluster if they have an energy above twice the expected noise level to form superclusters. The average noise level per cell in the electromagnetic calorimeter is roughly 40 and 150 MeV in the barrel and endcap respectively, while the noise per tower in the hadronic calorimeter is around 200 MeV [25]. Lastly, various algorithms calculate the substructure of these topological clusters to separate contributions from individual sources into smaller clusters.

3.2.8.2 Muon Reconstruction

Muon reconstruction relies on tracks reconstructed in the inner tracker (tracker tracks) and tracks reconstructed in the muon systems. There are three categories of muon objects based on the method used to reconstruct tracks.

Standalone-muon tracks utilize a Kalman-filter algorithm to construct tracks using only information from the muon system. Tracks seeds are formed from CSC or DT track segments and propagated to nearby hits in the muon chambers.

Global muon tracks begin with a standalone track, which is propagated to the silicon tracker and matched to compatible tracker tracks. Two tracks are matched if the standalone track and tracker track can be propagated to a common surface. The track is then refit using a Kalman filter by combining the information from the tracker and standalone track.

Tracker muon tracks recapture muons in detector gaps and lower p_T muons that leave hits in the muon chambers but do not fully penetrate the muon system. In these cases, it is common for a muon to leave a segment of hits in consecutive DT/CSC layers that are compatible with a track but not produce a standalone track. Tracker muons are formed by propagating all tracker tracks with $p_T > 0.5$ GeV and $p > 2.5$ GeV into the muon chambers, and matching to nearby muon segments.

This reconstruction has an efficiency of 99% for muons produced within the geometric acceptance of the muon system [8]. Muon objects are then fed into the PF algorithm, which performs selection criteria based on the quality of the track. As the purity of each track varies based on type (e.g. hadrons that "punch through" the HCAL are more probable to be reconstructed as track muon tracks), the cuts vary for each category.

3.2.8.3 Electron and Photon Reconstruction

In theory, electrons should leave a smoothly curving track in the silicon tracker and deposit nearly all of their energy within a few crystals in the ECAL. However in practice, electrons

frequently interact with the material in the tracker, which induces electromagnetic showers before the particle can reach the calorimeter, resulting in tracks which can have sudden changes in curvature and ECAL clusters that are smeared in ϕ .

The e/γ reconstruction algorithm accounts for these factors when track fitting and forming clusters for electrons. Electron track seeds are formed when the momentum of a track and the energy of an associated ECAL cluster are compatible with unity [25]. These tracks are refit using a Gaussian Sum Filter (GSF), which uses a sum of gaussian distributions to approximate the probability density function for energy loss due to bremsstrahlung [26].

Bremsstrahlung photons are emitted tangent to the trajectory of the electron, which bends in ϕ . To ensure this energy is accounted for in the reconstruction, clusters located tangent to the GSF track are associated to the track and supercluster. Additionally, the total cluster energy is calibrated as a function of energy and η to account for energy loss. Photons, which are seeded by ECAL clusters with no linked GSF track, can undergo the same process of pair production to induce electromagnetic showers. As such, photon candidate clusters are subject to the same energy calibration.

Several cuts are applied to the candidate particles to reject background from hadronic processes. Electrons and photons are expected to have a low ratio of linked HCAL cluster energy to their ECAL cluster energy. Electron tracks have several variables that are fed to a boosted decision tree, which decides whether or not to accept an electron candidate. All tracks, GSR tracks, clusters, and superclusters associated with reconstructed electrons or photons are passed to the PF algorithm.

3.2.8.4 Particle Flow

Particle flow (PF) is a global reconstruction algorithm that takes advantage of the unique signatures of various physics objects as they pass through each detector. It takes information passed from the muon, electron, and photon reconstruction as well as the remaining tracks

and clusters as input to form "pf blocks" in a process known as linking, which topologically associates objects from one or more detector. The particle flow algorithm then takes these blocks and reconstructs the individual particles. In order, the algorithm reconstructs muons, electrons and isolated photons, charged hadrons, and neutral particles. After each step the corresponding tracks and clusters are removed from the pf blocks to prevent double counting. Once all objects are reconstructed, a final post processing step is done to further reduce misreconstructed particles. The PF candidates are used to reconstruct higher level objects such as hadronized quarks and gluons, decay products from τ mesons, or b -quark decays, and calculate the imbalance of transverse momentum created from particles that are invisible to the detector (called missing E_T , or MET).

CHAPTER 4

Real Time Muon Reconstruction at the Compact Muon Solenoid

4.1 Introduction

Section 3.2.7.1 provides an overview of the L1 trigger and its importance to the CMS trigger system. This section will detail the design and performance of a Kalman Filter algorithm used in the L1 trigger to identify muon tracks in the barrel region ($|\eta| < 0.83$) of the CMS detector. This algorithm, known as the Kalman Barrel Muon Track Finder (KBMTF), was fully implemented for 2018 data taking and received improvements for continued use in Run-3.

The L1 barrel muon trigger receives inputs from the TwinMux system, which creates trigger primitives by combining information from the DT and RPC detectors to determine the position, trajectory, and timing of hits from incident muons at each station [27]. These hits, referred to as "stubs", are fed to the L1 muon trigger electrons where they are used to reconstruct muon tracks. The previous track finding algorithm, known as the Phase-I Barrel Muon Track Finder (BMTF), used a maximum of two stubs to reconstruct muon trajectories and was designed under the assumption that muons originate only from the beam line. As a part of upgrading the L1 Trigger, the KBMTF algorithm was developed to utilize stubs from all four muon stations as well as reconstruct displaced muon tracks resulting from decays of exotic long lived particles.

4.2 The Kalman Filter Algorithm

A Kalman Filter is a tracking algorithm that performs a recursive, iterative chi-square-like fit. Qualitatively, it propagates a system from its current state to the next and combines the predicted value with measurements to "update" the state of the system. This updated system is then iteratively propagated and updated with new measured values. This section focuses on a discrete linear Kalman Filter, which is applicable when a system can be described with a vector of variables whose evolution can be modeled with a linear transformation plus random uncertainties [28].

Abstractly, the state of a system at a given step $n - 1$ can be described by a vector labeled x_{n-1} with covariance matrix P_{n-1} . Let the matrix F_n represent the linear transformation that propagates x_{n-1} to the next state x_n , and the matrix Q_n be the covariance matrix due to additional uncertainty. The predicted state of the system at step n can be given by

$$x_n = F_n x_{n-1} \quad \text{and} \quad P_n = F_n P_{n-1} F_n^T + Q_n \quad (4.1)$$

where $F_n P_{n-1} F_n^T$ represents the propagation of the initial covariance matrix.

Now define a set of measurements taken at state n as z_n with covariance matrix R_n . The measured variables are not restricted to the same set of variables defining x_n as long as there exist a "change of basis" matrix H that relates the sets. It should be noted that mathematically H is not a strict change of basis matrix, as the measured values can have smaller dimensionality than the propagated ones. The predicted state and covariance matrix of the system written in the same variables as the measured quantities are defined as

$$\mu_n = H x_n \quad \text{and} \quad \Sigma_n = H P_n H^T \quad (4.2)$$

Updating the system relies on a matrix known as the Kalman Gain, which acts as a weight based on Σ_n and R_n and determines if the updated system should skew more towards

the measured or predicted values. The Kalman gain is defined as

$$K := \Sigma_n (\Sigma_n + R_n)^{-1} \quad (4.3)$$

which is then used to calculate the updated system as follows

$$\mu'_n = \mu_n + K(z_n - \mu_n) \quad \text{and} \quad \Sigma'_n = \Sigma_n - K\Sigma_n \quad (4.4)$$

The term $z_n - \mu_n$ is frequently referred to as the residual of the prediction and measured value. A Kalman gain equal to the identity I would set the updated coordinates to the measured values, while a Kalman gain of 0 would effectively ignore the measured values. Finally, we substitute x_n and P_n into equation 4.4 using equation 4.2 and simplify to give

$$x'_n = x_n + K'(z_n - Hx_n) \quad \text{and} \quad P'_n = P_n - K'HP_n \quad (4.5)$$

where the Kalman Gain K has been redefined to

$$K' = P_n H^T (H P_n H^T + R_n)^{-1} \quad (4.6)$$

The system x'_n and P'_n can now be propagated to step $n + 1$ where the Kalman Algorithm can be iterated.

4.3 The Kalman Barrel Muon Track Finder

A rough outline KBMTF algorithm for track finding is as follows:

1. A track seed is chosen from a stub in the muon station, from which a preliminary track is built. This seed cannot be chosen from the innermost muon station.
2. The track is propagated inward to the next station and matched to the closest stub.

If there is a matching stub, update the track with the stub information. Repeat until the track is at the innermost station.

3. The track is propagated from the innermost station to the beamline. The track properties at this point are stored as the "vertex unconstrained" measurement. These properties include the p_T , ϕ , η , and d_{xy} , which is defined as the closest distance from the propagated track to the beamline.
4. The vertex propagated track is updated with the constraint that the track originated from the beamline, and the track properties are stored as the "vertex constrained" measurement.

Muon track finding begins at the outer stations in order to get both the vertex unconstrained measurement and the vertex constrained measurement with only one iteration of the KBMTF algorithm. Starting from the inner station and propagating outward would require an additional propagation back inward in order to get the vertex constrained measurement, which would cause the algorithmic latency to exceed timing restrictions. The outer stations are also the lowest occupancy, meaning stubs are less likely to be from background. Tracks are then overlap cleaned, which ensures that stubs are not shared among multiple tracks, and cut and selected based on various goodness of fit criteria. A diagram showing the KBMTF propagation and update procedure can be shown in figure 4.1.

4.3.1 Muon Trajectory Propagation

In order to implement a Kalman Filter, the propagation of muon tracks between stations must be expressed as a linear function of the track variables. From equation 3.6, a charged particle will travel in a circular orbit in the $\hat{r} - \hat{\phi}$ plane. The high center of mass energy of the LHC results in highly energetic particles, whose trajectories have radii substantially larger than the size of the CMS detector. The large radius of curvature allows us to approximate these trajectories as parabolas. Assume a Cartesian coordinate system with the origin placed

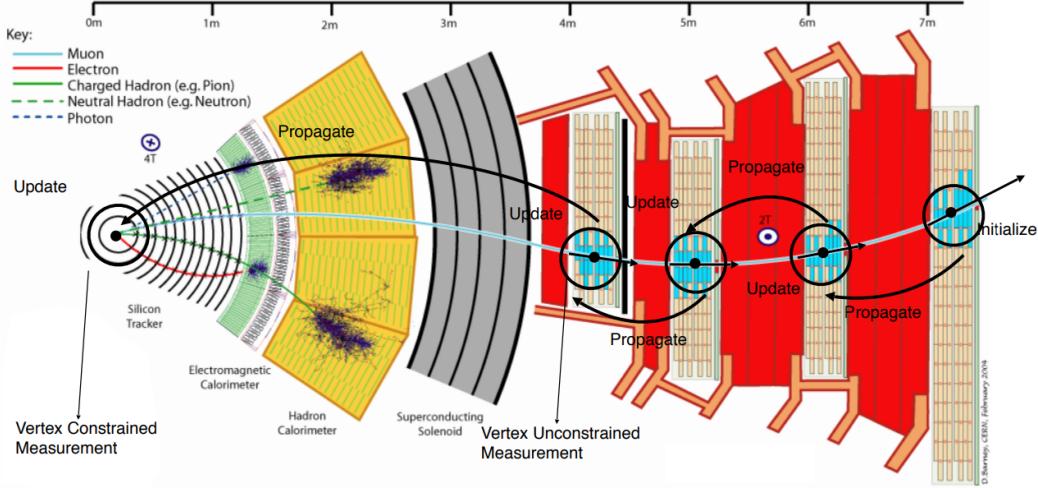


Figure 4.1: The iterative process of propagation and updating a muon track through the KBMTF algorithm. The track properties at the innermost station are stored in order to trigger on muons not originating from the beamline [14].

along a particle's trajectory that has radius R . This trajectory can then be approximated as

$$y(x) = \frac{x^2}{2R} + bx \quad (4.7)$$

where b is a coefficient depending on the orientation of the coordinate system. If the initial trajectory is defined as $\phi_{b,0}$, taking the derivative and evaluating at the origin yields

$$y'(0) = \tan(\phi_{b,0}) = b \quad (4.8)$$

The curvature $k = q/p_T$, where the muon charge $q = \pm 1$, is the preferred variable to work with when propagating the trajectory of a muon, as the propagation is linear in k . Substituting values from equations 3.6 and 4.8 into equation 4.7 yields

$$y(x) = akx^2 + \tan(\phi_{b,0})x \quad (4.9)$$

where $a = \frac{0.3B}{2}$. Muon hits, or "stubs", provide information on the position ϕ and bending angle ϕ_b . Let the radius of two sequential muon stations be r_1 and r_2 , with the outer radius given by r_2 and $\Delta r = r_2 - r_1$. Assuming a muon has curvature k at the outer station and ignoring energy loss, equation 4.9 can be used to propagate stubs from the outer station towards the inner station.

$$y(-\Delta r) = (a\Delta r^2) k - \Delta r \tan(\phi_{b,0}) \quad (4.10)$$

$$y'(-\Delta r) = -(2a\Delta r) k + \tan(\phi_{b,0}) \quad (4.11)$$

Converting these to the quantities to angles measured by the detector yields

$$\tan(\Delta\phi) = \frac{y(-\Delta r)}{r_1} \quad (4.12)$$

$$\phi_b = \Delta\phi + \tan^{-1}[y'(-\Delta r)] \quad (4.13)$$

The $\Delta\phi$ term is added to equation 4.13 because the bending angle at each station is measured relative to an axis system oriented towards the center of the detector. Due to the large radius of curvature of the muon trajectories, the small angle approximation $\tan(x) \sim \tan^{-1}(x) \sim x$ can be applied, giving the equations

$$\Delta\phi = \frac{a\Delta r^2}{r_1} k - \frac{\Delta r}{r_1} \phi_{b,0} \quad (4.14)$$

$$\phi_b = a\Delta r \left(\frac{\Delta r}{r_1} - 2 \right) k + \left(\frac{r_2}{r_1} \right) \phi_{b,0} \quad (4.15)$$

A diagram showing this propagation can be seen in figure 4.2.

Equations 4.12-4.13 show that the track propagation fits the criteria for a discrete linear Kalman Filter discussed in section 4.2. The matrices for propagation can now be constructed

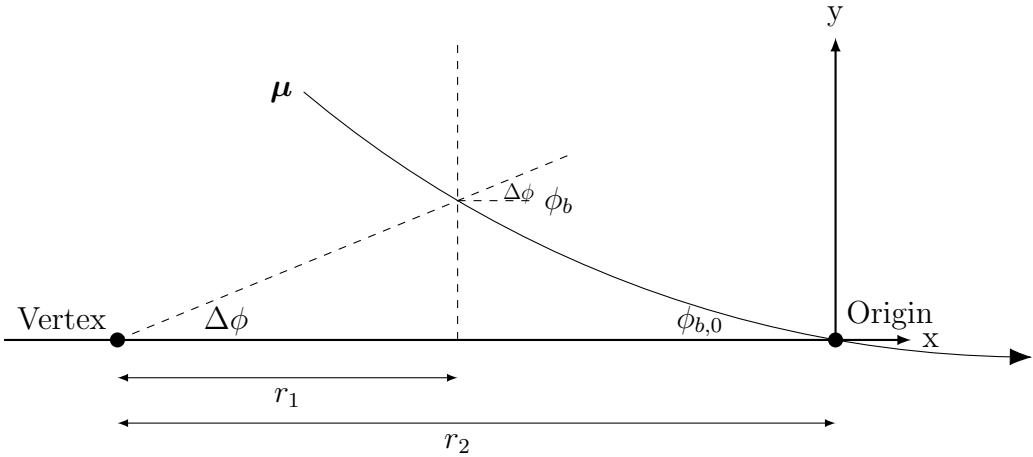


Figure 4.2: Diagram showing a muon trajectory between two stations with the origin set at the outer station. The x-axis is oriented using the detector vertex and the outer station.

as follows. The transfer matrix F_n from equation 4.1 can be expressed as

$$x_n = \begin{pmatrix} k \\ \phi \\ \phi_b \end{pmatrix}_n = \begin{pmatrix} 1 & 0 & 0 \\ \alpha_n & 1 & -\frac{\Delta r}{r_n} \\ \beta_n & 0 & \frac{r_{n-1}}{r_n} \end{pmatrix} \begin{pmatrix} k \\ \phi \\ \phi_b \end{pmatrix}_{n-1} = F_n x_{n-1} \quad (4.16)$$

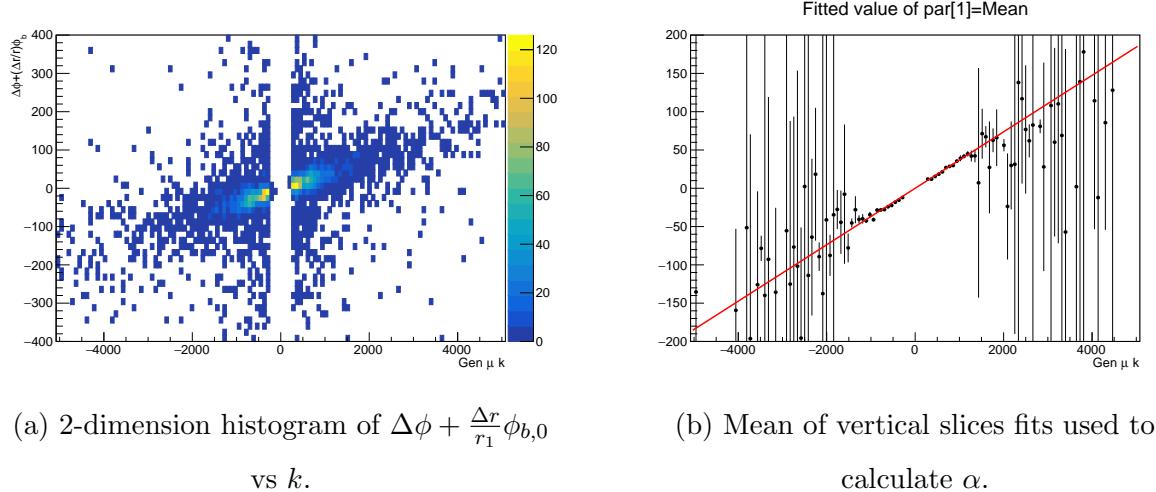
where the coefficients

$$\alpha_n = a \frac{\Delta r^2}{r_n} \quad \text{and} \quad \beta_n = a \Delta r \left(\frac{\Delta r}{r_n} - 2 \right) \quad (4.17)$$

with r_{n-1} being the radius of the previous (outer) muon station and r_n the radius of the sequential (inner) muon station.

While the radii of the muon stations are determined by detector geometry, the constants α_n and β_n are measured using single muon monte-carlo samples where stubs are matched to the incident muons using generator level information. To determine α_n , stubs in stations n and $n - 1$ are first associated if they resulted from the same incident muon. For matching stubs, the quantity $\Delta\phi + \frac{\Delta r}{r_1}\phi_{b,0}$ is plotted versus the muon k in a 2-dimensional histogram.

Slices along the y-axis are fit using a Gaussian distribution, and the mean value for each slice is used to calculate a linear fit as a function of k . β_n is calculated similarly by plotting $\phi_b - (r_{n-1}/r_n) \phi_{b,0}$ vs k . This process can be seen in figure 4.3.



(a) 2-dimension histogram of $\Delta\phi + \frac{\Delta r}{r_1} \phi_{b,0}$ vs k .
(b) Mean of vertical slices fits used to calculate α .

Figure 4.3: Calculation of phi propagation coefficients. Units are converted to their digitized form used in firmware calculations.

4.3.2 Covariant Matrices and the Kalman Gain

There are two sources of uncertainty that affect the covariance matrices. The first is multiple coulomb scattering, which occurs when muons scatter elastically with material in the detector. This can be thought of as a small deflection in momentum that alters the trajectory the muon. Lower momentum muons will be deflected at greater angles, so this uncertainty is proportional to k . Additionally, the probability for multiple scattering is proportional to the amount of material traversed by the muon, so this uncertainty must be calculated for each step in propagation. This is an uncertainty resulting from the propagation, and therefore plays the role of Q_n from equation 4.1. The second source is from the intrinsic detector resolution, which is a fixed value for each detector. However, the angular resolution is dependent on the radial distance from the center of the detector, so these must be calculated for each

station as well. Since this is an uncertainty on the measurements, the resolution plays the role of R_n in equation 4.3. The total uncertainty resulting from both of these factors is given by

$$\sigma(k) = \sqrt{\sigma_{ms}^2 k^2 + \sigma_{res}^2} \quad (4.18)$$

The coefficients can be calculated by using the Gaussian slice fits described when deriving α and β . For each propagation, the sigma of the Gaussian fits is plotted versus k and fit to the form in equation 4.18. This fitting process for ϕ propagation coefficients is shown in figure 4.4, and is functionally the same for ϕ_b .

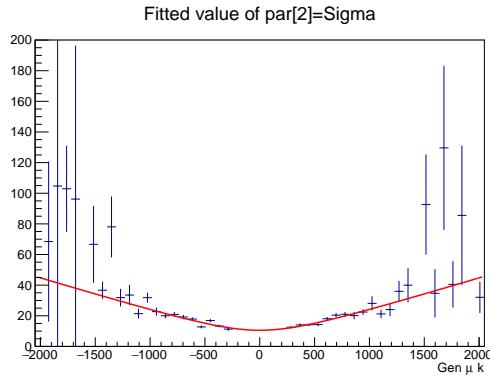


Figure 4.4: σ of vertical slice fits of the 2-dimensional histogram in figure 4.3a used to calculate uncertainty coefficients.

Each stub has a measured value representing the "quality" of the hits, with higher values representing a more precise reconstruction of the bending angle. For this reason, the ϕ_b position resolutions are calculated separately for high and low quality stubs, as the uncertainty can differ by over a factor of 10 between the two.

Since the muon detectors only measure ϕ and ϕ_b , the matrix H described in equation 4.2 must be a 2×3 matrix that takes the form

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.19)$$

From this, the Kalman Gain K' defined in equation 4.6 must be a 3x2 matrix with indices K_{ij} , and the update step can be explicitly written as

$$\begin{aligned} k_n^{\text{upd}} &= k_{n-1} + K_{00}\Delta\phi + K_{01}\Delta\phi_b \\ \phi_n^{\text{upd}} &= \phi_n^{\text{prop}} + K_{10}\Delta\phi + K_{11}\Delta\phi_b \\ \phi_{b,n}^{\text{upd}} &= \phi_{b,n}^{\text{prop}} + K_{20}\Delta\phi + K_{21}\Delta\phi_b \end{aligned} \quad (4.20)$$

where $\Delta\phi$ and $\Delta\phi_b$ are the residuals for ϕ and ϕ_b .

4.3.3 Firmware Implementation of the KBMTF

Two minor approximations are made in order to reduce algorithmic latency and memory usage for firmware implementation. The position resolution of ϕ is substantially smaller than the multiple scattering resolution, so the track ϕ is set automatically to the measured stub ϕ . This is equivalent to fixing $K_{10} = 1$ and $K_{11} = 0$ in equation 4.20. Secondly, due to the high uncertainty associated with both the measured and propagated ϕ_b , tracks using three or more stubs set K_{i1} equal to zero and only update using the residual from ϕ .

Although the Kalman Gain can be exactly calculated exactly given the covariance matrices using equation 4.3, matrix inversion is a resource and latency expensive process for the hardware based trigger. Implementing division into hardware is significantly more resource intensive than multiplication, so the calculation of the determinant for matrix inversions would cause the algorithm to exceed the latency requirements for real time track finding. Instead, the values of K_{ij} are stored in look-up tables (LUTS) as a function of curvature and the combination of stations that were used to update the track, referred to as the track pattern. For tracks with only two stubs, the LUT contains the exact values of the gain that would have been calculated using equation 4.6. For tracks with more than two stubs, this is an approximation that loses some information from updates at previous stations but produces nearly identical performance.

As outlined in section 4.3.2, the ϕ_b resolution coefficients are different for high and low quality stubs, thus the values for K_{ij} would have to be provided for every possible combination of high and low quality stubs for each track pattern. This would exponentially increase the number of gains stored on LUTs beyond the memory capacity of the hardware. However, from previous approximations, residuals for ϕ_b are not used for tracks containing three or more stubs, which fixes K_{01} and K_{21} at 0 for these track patterns. Only gains for track patterns containing two stubs must account for the quality of the stubs. With four stations, this yields six track patterns with four combinations of stub quality that must be stored in LUTs, which is within the capacity of the current hardware.

The firmware is written using Vivado High Level Synthesis (HLS) which compiles firmware written in C to HDL, optimizing it for a given FPGA and clock frequency [29]. When synthesized in a Virtex 7 690T FPGA with a clock frequency of 160 MHz, the KBMTF algorithm can process a single event in 9 bunch crossings, and is pipelined to allow for continuous processing of events. Table 4.1 shows the resource utilization of the KBMTF compared to the previous BMTF algorithm, which was implemented using an identical FPGA and clock frequency.

Algorithm	LUT	FF	BRAM	DSP
Phase I BMTF	43%	23%	35%	0%
KBMTF	16%	11%	15%	25%

Table 4.1: Comparison of FPGA resource utilization [14]. The KBMTF utilizes DSP cores for arithmetic operations to propagate and update tracks and LUTs for the propagation coefficients and approximate Kalman Gain, while the BMTF uses LUTs to assign momentum.

4.4 Performance of the Kalman Barrel Muon Track Finder

The performance of an trigger algorithm is primarily determined by the rate and efficiency. The efficiency is the fraction of signal events that successfully have a track reconstructed

by the algorithm, while the rate is the frequency that the algorithm will trigger. A high efficiency is desirable to ensure any interesting events are stored for future analysis, while a low rate is crucial to reject background events and provide a reasonable bandwidth to store events.

4.4.1 Trigger Efficiencies and Rate

The previous L1 barrel muon trigger (BMTF) functioned by using pairs of stubs to form tracks, then using a LUT to assign track parameters based on the ϕ and ϕ_b of the stubs. This LUT takes only the ϕ and ϕ_b of the two stubs and was calculated with the assumption that the muon originated from the beamspot. In the case of displaced muons, which arise from secondary decays of long lived, beyond standard model particles, the assumption is no longer valid and results in large inefficiencies as the muon becomes more displaced from the beamspot.

The general efficiency of a trigger is defined as the fraction of signal muons that have a matching reconstructed track, and can be expressed abstractly as

$$\epsilon = \frac{N_{\text{signal}}(\text{Passing trigger})}{N_{\text{signal}}} \quad (4.21)$$

For muon triggers, the denominator of signal muons consist of events that have known information about the muons. This can be from generator information in simulated monte-carlo samples, tracker information using cosmic muon data, or probe muons using the tag-and-probe method on data. These signal muons are cut using the known muon information to produce a reasonable sample for track reconstruction. The numerator consists of signal muons with a matching reconstructed track that may have addition cuts on the reconstructed track properties.

Figure 4.5 shows the efficiency versus muon d_{xy} for three track finding algorithms. The denominator consists of cosmic ray muons passing through the barrel muon detectors with a

$p_T > 20$ GeV. The numerator requires a matching L1 track with reconstructed $p_T > 10$ GeV. The lower threshold for the L1 track is set to prevent the momentum resolution of the reconstruction algorithms from affecting the efficiency. The first algorithm is the BMTF, which shows inefficiencies due to the p_T LUTs assuming the muon originated at the beamspot. The second is the prompt KBMTF, which uses the vertex constrained measurement. This also shows inefficiencies due to the vertex constraint, but is better than the BMTF as the vertex constraint isn't applied until the last step of track reconstruction. Lastly, the displaced KBMTF uses the vertex unconstrained measurement and shows substantial improvements compared to both prompt algorithms.

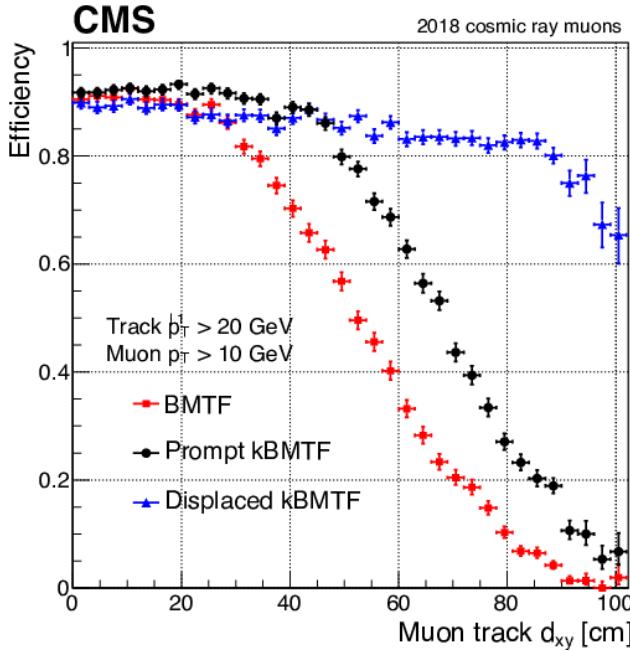


Figure 4.5: Comparison of efficiency vs cosmic track d_{xy} using cosmic data from the 2018 data taking run. The prompt KBMTF takes the reconstructed p_T using the vertex constrained measurement, while the displaced KBMTF uses the veretx unconstrained measurement. The displaced KBMTF shows substantial improvements to the previous BMTF [15].

A reconstructed track must have a p_T above a given threshold to trigger on an event. Thresholds are set to optimize the analysis potential of the data while also minimizing the

additional bandwidth to store events. The rate, which is the frequency at which a trigger decides to store an event, can be calculated as a function of threshold using zero bias data, which is collected by storing random events regardless of trigger acceptance. This creates a representative data set of all pp collisions, hence the name "zero bias". The rate can be calculated from the zero bias data as follows

$$R = f_{\text{rev}} \times \langle N_{\text{bunches}} \rangle \times \frac{N_{\text{events}}(\text{L1 Track } p_T > \text{Threshold})}{N_{\text{events}}} \quad (4.22)$$

Where f_{rev} is the LHC revolution frequency of 11.245 kHz and $\langle N_{\text{bunches}} \rangle$ is the average number of proton bunches that will collide in one revolution. For 2017 data taking these yield a scale factor of 20984 kHz. Figure 4.6 shows the rate versus threshold for the BMTF, vertex constrained KBMTF, and vertex unconstrained KBMTF.

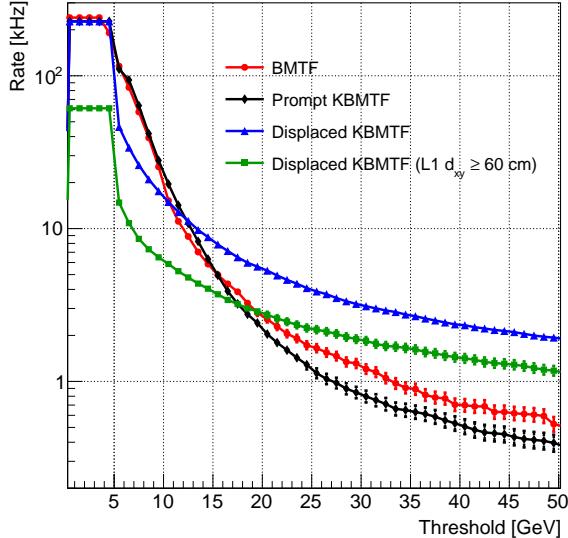


Figure 4.6: Rate comparison calculated using 2017 zero bias data of the Phase I BMTF, vertex constrained (prompt) KBMTF, vertex unconstrained (displaced) KBMTF, and displaced KBMTF for displaced tracks.

4.4.2 Firmware and Emulator Agreement

The performance of the KBMTF is tested using a dedicated C++ software emulator which is integrated into the CMS analysis framework. This emulator follows similar logic to the HLS code, but is not a one-to-one replica. Several modifications are made to the HLS code in order to make it compatible with the synthesis into HDL and pipelining. As such, a test bench is used to ensure that the output of the firmware and emulator are identical. The emulator runs the algorithm over several events and outputs a text file containing information on the input stubs and the reconstructed tracks. This text file is used as input for the test bench, which runs the stub information through the HLS code and compares the properties of the reconstructed tracks to those found in the emulator.

Disagreements between the emulator and firmware are used for a variety of debugging purposes. They frequently point to logic errors in the simplification of the algorithm for HLS synthesis, or in other cases show flaws in the emulator. During Run 2, the disagreements were almost entirely due to differences resulting from fixed point calculations in the firmware compared to floating point calculations in the emulator. These differences have been drastically reduced due to the addition of C++ classes that mimic fixed point calculations in the emulator, which now has a >99.9% agreement to the firmware. Figure 4.6 shows a comparison of the outputs of the firmware and emulator using events from a 2017 Z tag-and-probe sample.

4.5 The Phase-2 KBMTF

The modifications to the KBMTF algorithm for the HL-LHC result from increased precision from the input trigger primitives and modifications to the trigger architecture. These changes mainly affect the firmware implementation and do not require conceptual changes to the logic used to propagate and update tracks. The propagation coefficients and Kalman Gain must be recalculated as the number of bits allocated to k , ϕ , and ϕ_b change. Additionally,

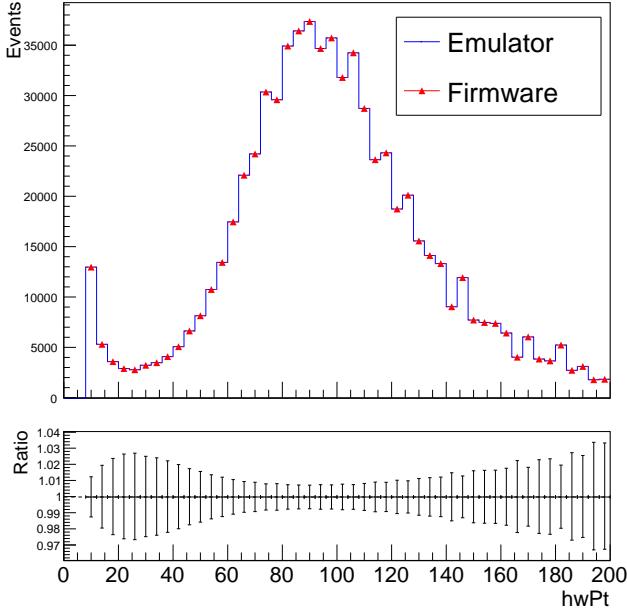


Figure 4.7: Comparison of the track p_T between the firmware and emulator using data from a 2017 Z tag-and-probe sample. hwPt represents the integer value of p_T used in hardware. The emulator and firmware output show 100% agreement across 500000 events.

the firmware will receive primitives from the entire barrel muon system simultaneously as opposed to processing each sector in ϕ independently. The firmware will be time-multiplexed with a factor of 18, meaning data from each event will be sent to one of 18 processors running the KBMTF algorithm in parallel. Additionally, the FPGA will be changed to a ZYNQ Ultrascale+ and synthesized at a clock frequency of 200 MHz. Simulations show a resource utilization of 17% LUT, 3% FF, 11% DSP, and 7% BRAM [14].

The KBMTF algorithm performance is robust to the high pileup conditions of the HL-LHC. The efficiencies remain the same as Phase-1, and the rate scales approximately linearly with pileup as expected. Figure 4.8 shows the efficiency versus generated muon p_T , single muon rates versus threshold, and single muon rates versus pileup for the Phase-2 KBMTF and two proposed track finding algorithms. As discussed in section 3.2.7.1, the HL-LHC track trigger introduces new track primitives that will be used in novel muon triggers. These new

algorithms have higher efficiency and precision for prompt muons due to the momentum resolution of the tracker, but will be unable to trigger on displaced muons as the track primitives do not include displaced tracks.

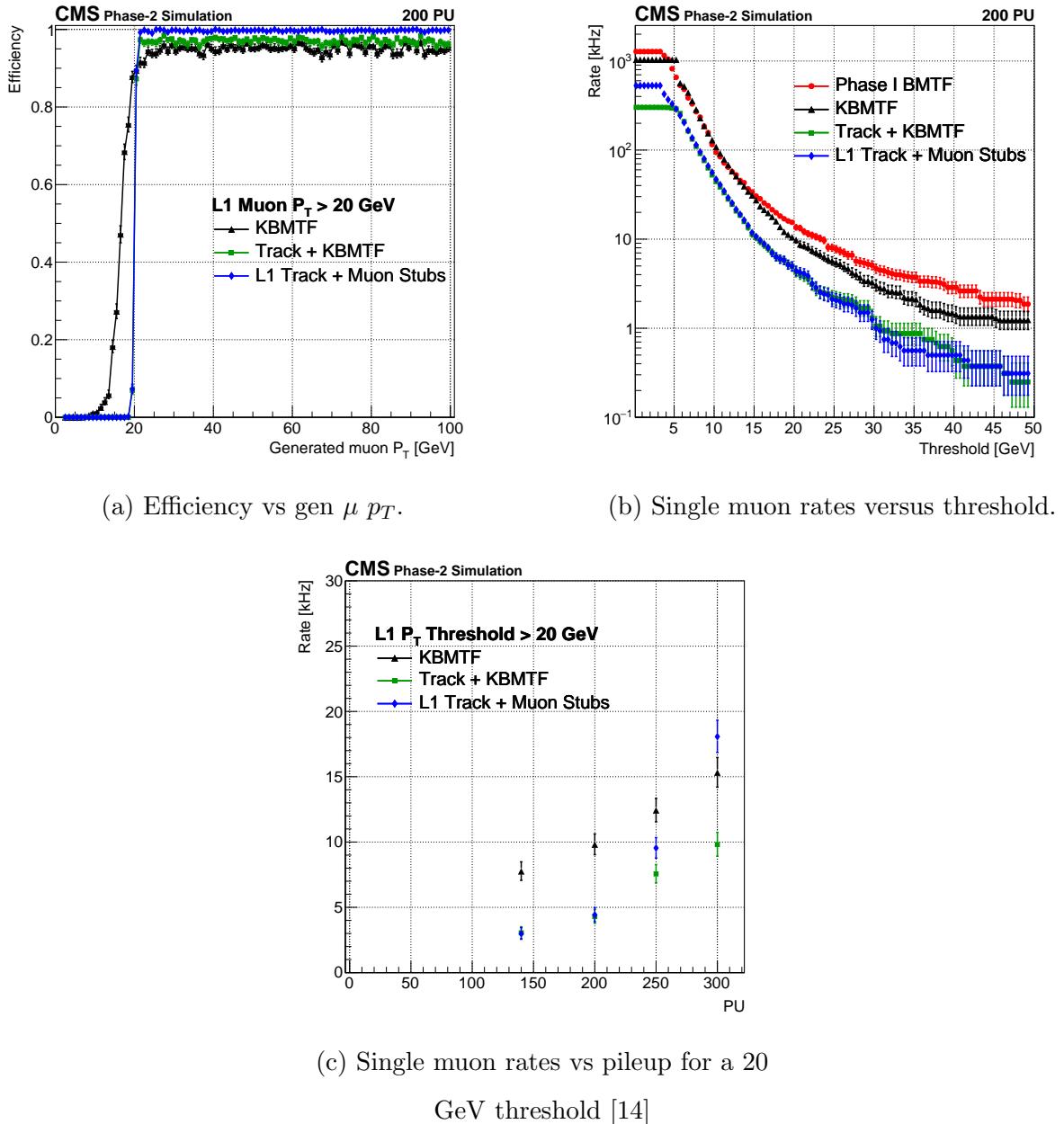


Figure 4.8: Performance of the L1 Barrel Muon algorithms using simulated samples with HL-LHC run conditions. The Track + KBMTF algorithm matches tracks to loose KBMTF muons, while the L1 Track + Muon Stubs algorithm matches tracks to muon stubs.

CHAPTER 5

Search for a Long Lived Scalar Boson

5.1 Introduction

This chapter describes the search for beyond the standard model (BSM) particles using data taken by the CMS detector from 2016-2018. The following sections outline the choice of datasets and simulated samples, event selection criteria, methods of background estimation and statistical analysis for signal extraction, and results.

5.2 Data and Simulated Samples

5.2.1 Data Samples

As discussed in chapter 3, interesting events that pass at least 1 HLT path are stored for future analysis. These events are collected and sorted into datasets based on the number, flavor, and type of physics objects used in the HLT path. It can be noted that these datasets are not disjoint; one event can be sorted into multiple datasets if it satisfies more than a single HLT path. This data is maintained and updated by the Physics Data and Monte Carlo Validation (PdmV) group, which also provides recommendations on optimal usage for physics analysis [30]. This search utilizes the PdmV recommended datasets for Run-2 analyses, referred to as Ultra-Legacy (UL), which has been reprocessed using the most current calibration for detector responses and alignment. Although the characteristic signature of this analysis is two displaced photons, the HLT double photon triggers present in Run-2 are

unsuitable for the phase space of signal photons. Therefore we utilize the leptonic decay of the associated Z which provides a robust trigger. Since we are interested in Z decays to either electrons or muons, we use the datasets labeled as SingleMuon or SingleElectron (renamed to EGamma in 2018).

The reprocessed datasets must be cross referenced with the "Golden JSON": a certificate created by the Data Quality Monitoring (DQM) group within CMS by monitoring each sub-detector system in order to veto collisions that occur during sub-optimal conditions. The total integrated luminosity recorded by CMS during Run-2 was 164 fb^{-1} ; however, only 138 fb^{-1} of that was certified for analysis. The total integrated luminosity and uncertainty per year for collisions recommended by DQM can be seen in table 5.1.

Table 5.1: Table of integrated luminosity per year [16–18].

Year	2016	2017	2018	2016-2018
Luminosity [fb^{-1}]	36.31	41.48	59.83	137.62
Uncertainty [%]	1.2	2.3	2.5	1.6

5.2.2 Simulated Samples

Simulated samples use Monte Carlo methods to generate events in several steps, starting from the matrix elements that determine cross sections and hadronization processes and ending with the processing of digitized signals from each subdetector in CMS. These samples are generally created to model specific physics processes such as Drell-Yan, QCD, $t\bar{t}$, etc. Simulated events are not created one-to-one to match data events; instead they are used to create weighted events that are normalized to match the luminosity of the data. For this analysis we use both simulated background events as well as simulated signal samples.

5.2.2.1 Monte Carlo Background

Although this analysis employs a data driven method of background estimation, simulated samples are used to inform on properties of background events and tune cuts to discriminate between signal and background. For this analysis, we use simulated Drell-Yan (DY) events with 0, 1 and 2 additional jets, simulated at next-to leading order (NLO). Studies were done with inclusive (any number of additional jets) samples but found to have worse agreement and statistical power compared to the jet categorized samples.

5.3 Event Selection

REFERENCES

- [1] CMS Collaboration. CMS luminosity - public results, 2018. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [2] CMS Collaboration. The high-luminosity lhc (hl-lhc). URL: <https://voisins.web.cern.ch/en/high-luminosity-lhc-hl-lhc>.
- [3] Tai Sakuma. Cutaway diagrams of CMS detector. 2019. URL: <http://cds.cern.ch/record/2665537>.
- [4] The Tracker Group Of The CMS Collaboration. The cms phase-1 pixel detector upgrade, 2020. URL: <https://arxiv.org/abs/2012.14304>, doi:10.48550/ARXIV.2012.14304.
- [5] S Chatrchyan et al. Alignment of the CMS silicon tracker during commissioning with cosmic rays. *JINST*, 5:T03009, 2010. URL: <https://cds.cern.ch/record/1211825>, arXiv:0910.2505, doi:10.1088/1748-0221/5/03/T03009.
- [6] Badder Marzocchi. Simulation of the cms electromagnetic calorimeter response at the energy and intensity frontier. *Journal of Physics: Conference Series*, 1162:012007, 01 2019. doi:10.1088/1742-6596/1162/1/012007.
- [7] Albert M Sirunyan et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *JINST*, 15:P05002, 2020. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-18-001> (CMS Public Pages). URL: <https://cds.cern.ch/record/2691403>, arXiv:1910.00079, doi:10.1088/1748-0221/15/05/P05002.
- [8] A.M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/MUO-16-001> (CMS Public Pages). URL: <http://cds.cern.ch/record/2313130>, arXiv:1804.04528, doi:10.1088/1748-0221/13/06/P06015.
- [9] CMS Collaboration. Muon drift tubes. URL: <https://cms.cern/detector/detecting-muons/muon-drift-tubes>.
- [10] Isabelle Helena J De Bruyn. Electronics upgrade for the CMS CSC muon system at the High Luminosity LHC. Technical report, CERN, Geneva, 2019. URL: <http://cds.cern.ch/record/2797803>.

- [11] Priyanka Kumari. Improved-rpc for the cms muon system upgrade for the hl-lhc, 2020. URL: <https://arxiv.org/abs/2005.11396>, doi:10.48550/ARXIV.2005.11396.
- [12] Albert M Sirunyan et al. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 15(10):P10017, 2020. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/TRG-17-001> (CMS Public Pages). URL: <https://cds.cern.ch/record/2721198>, arXiv: 2006.10165, doi:10.1088/1748-0221/15/10/P10017.
- [13] Benjamin Radburn-smith. Overview of the hl-lhc upgrade for the cms level-1 trigger. page 639, 12 2022. doi:10.22323/1.414.0639.
- [14] The Phase-2 Upgrade of the CMS Level-1 Trigger. Technical report, CERN, Geneva, 2020. Final version. URL: <https://cds.cern.ch/record/2714892>.
- [15] Aram Hayrapetyan et al. Development of the CMS detector for the CERN LHC Run 3. Technical report, CERN, Geneva, 2023. URL: <https://cds.cern.ch/record/2870088>, arXiv:2309.05466.
- [16] Albert M Sirunyan and other. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81(9):800, 2021. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/LUM-17-003> (CMS Public Pages). URL: <https://cds.cern.ch/record/2759951>, arXiv:2104.01927, doi:10.1140/epjc/s10052-021-09538-2.
- [17] CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2018. URL: <https://cds.cern.ch/record/2621960>.
- [18] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2019. URL: <https://cds.cern.ch/record/2676164>.
- [19] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. URL: <https://cds.cern.ch/record/782076>, doi: 10.5170/CERN-2004-003-V-1.
- [20] Esma Mobs. The CERN accelerator complex - august 2018. complexe des accélérateurs du CERN - août 2018. Aug 2018. General Photo. URL: <https://cds.cern.ch/record/2636343>.
- [21] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, oct 2014. URL: <https://dx.doi.org/10.1088/1748-0221/9/10/P10009>, doi:10.1088/1748-0221/9/10/P10009.

- [22] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi:10.1093/ptep/ptac097.
- [23] E. Tournefier. The preshower detector of cms at lhc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 461(1):355–360, 2001. 8th Pisa Meeting on Advanced Detectors. URL: <https://www.sciencedirect.com/science/article/pii/S0168900200012432>, doi: 10.1016/S0168-9002(00)01243-2.
- [24] G. Baiatian et al. Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges. Technical Report 1, CERN, Geneva, 2008. URL: <http://cds.cern.ch/record/1049915>, doi:10.1140/epjc/s10052-008-0573-y.
- [25] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the cms detector. *Journal of Instrumentation*, 12(10), October 2017. URL: <http://dx.doi.org/10.1088/1748-0221/12/10/P10003>, doi:10.1088/1748-0221/12/10/p10003.
- [26] W Adam, R Fröhwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, July 2005. URL: <http://dx.doi.org/10.1088/0954-3899/31/9/N01>, doi:10.1088/0954-3899/31/9/n01.
- [27] A. Triassi, M. Bellato, J.M. Cela Ruiz, J. Ero, C. Fernandez Bedoya, G. Flouri, C. Foudas, L. Guiducci, N. Loukas, A. Navarro Tobar, S. Mallios, G. Masetti, F. Montecassiano, E. Paradas, I. Redondo Fernandez, D. Redondo Ferrero, J. Sastre, P. Sphicas, S. Ventura, and C. Wulz. The cms barrel muon trigger upgrade. *Journal of Instrumentation*, 12(01):C01095, jan 2017. URL: <https://dx.doi.org/10.1088/1748-0221/12/01/C01095>, doi:10.1088/1748-0221/12/01/C01095.
- [28] R. Fröhwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874>, doi:10.1016/0168-9002(87)90887-4.
- [29] Michail Bachtis. Upgrade of the CMS Barrel Muon Track Finder for HL-LHC featuring a Kalman Filter algorithm and an ATCA Host Processor with Ultrascale+ FPGAs. Technical report, CERN, Geneva, 2019. URL: <https://cds.cern.ch/record/2648953>, doi:10.22323/1.343.0139.
- [30] CMS Collaboration. Pdmvgroupdescription. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVGroupDescription>.