University of Padova

Department of Electrical & Computer Engineering

Master Thesis in ICT for Internet and Multimedia

Field Measurements to Study the Effect of the Common Mode Current Noise in the Cabling for Phase 2 Upgrade of CMS Drift Tubes

Supervisor
Prof. Andera Triossi
University of Padova

Co-supervisor
Dr. Antonio Bergnoli
Padova University

Master Candidate
Elham Norouzimehmandoustolia

Student ID
2052056

Academic Year
2023-2024

"To my loving Reza"

Throughout this journey, you have been my unshakable pillar of support. Your unending support, love, and sacrifices molded me into who I am today. This thesis is a testimonial to your unwavering trust in me and your steadfast belief that I can do everything I set my mind to. Thank you for always being there for me, encouraging me, and believing in my aspirations.

With all my heart, Elit

Abstract

Contents

Listing of figures

Listing of tables

Listing of acronyms

FFT . . . . . Fast Fourier transform

LSTM . . . . . Long Short-Term Memory

STFT . . . . . Short-time Fourier transform

1 Introduction

1.1 Large Hadron Collider

Particle physicists use accelerators [1][2][3]to investigate particle interactions and find new phenomena, improving our understanding of matter in the Universe. The Large Hadron Collider (LHC), based at CERN in Geneva, Switzerland, is the most advanced particle accelerator currently in use.

The LHC is a CERN accelerator that collides proton beams with a center-of-mass energy of up to 14 TeV and a luminosity of 10^{34} $cm^{-2}s^{-1}$. It can also accelerate lead ions to 2.76 TeV per nucleon pair at its center of mass. The LHC is useful for exploring scientific phenomena at the TeV energy scale, such as the Higgs mechanism, and verifying the Standard Model's consistency (SM). It also investigates alternate possibilities and attempts to combine nature's fundamental forces. It uses the current CERN accelerator chain Fig.1.1 to obtain an injection energy of 450 GeV. Utilizing current facilities offers technical and economic benefits but also creates architectural limits due to the LEP tunnel's fixed curve.

High-field superconducting magnetic dipoles, capable of reaching a stable field of 8.3 T, are required to bend the beam trajectory within the existing tunnel. Acceleration to the design energy is achieved using superconducting radio-frequency cavities, and the entire system is maintained at a temperature of 2.1 K using superfluid helium to sustain the superconducting regime.

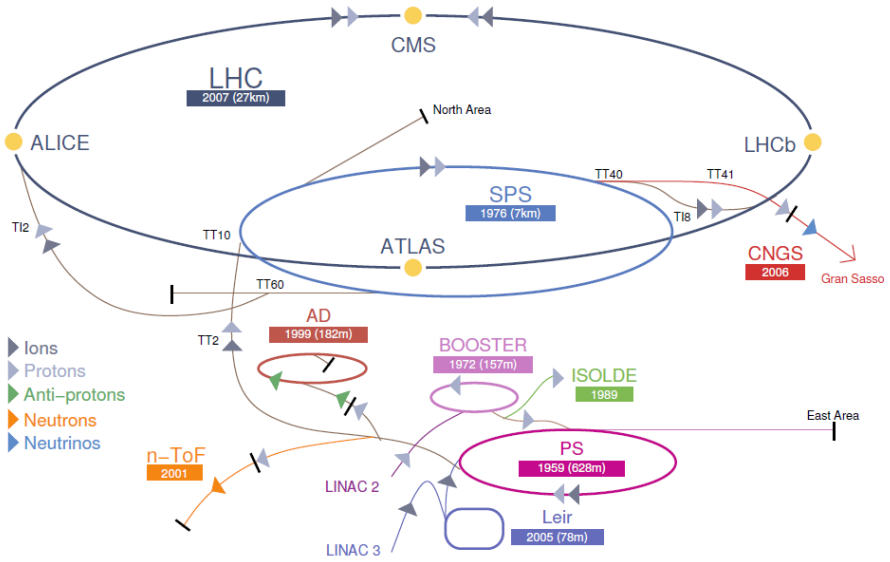


Figure 1.1: An exploded view of the CMS at CERN

The LHC collides bunches of approximately 1.1×10^{11} protons each, with a beam spot size of $\sigma_z \approx 53$ mm and $\sigma_{xy} \approx 15\mu$m, at a design bunch crossing (BX) frequency of 40 MHz. The design option for collision proton-proton (pp) beams provides numerous advantages:

energy range flexibility: Hadrons, like protons, can experiment with various energies utilizing fixed-energy beams. Protons are not simple particles; their interactions need components (quarks and gluons) that carry different amounts of proton energy, making them valuable in discovery experiments.

Production Efficiency: At high energies, proton-proton and proton-antiproton ($p\bar{p}$) total cross sections are similar. However, protons are easier and faster to create than antiprotons. This results in increased brightness and better beam stability. Two independent beam pipelines are required to manage proton bunches circulating in opposite directions.

Reduced Synchrotron Radiation: In a circular motion, energy loss due to synchrotron radiation is inversely proportional to the fourth power of the particle mass. Being almost 2000 times heavier than electrons, Protons experience significantly less synchrotron radiation. Using this method, the LEP tunnel can be reused to build a collider capable of reaching higher energy levels than its predecessor through the re-use of the tunnel.

Considering an average magnetic field of 5.4 T, this leads to a maximum energy of 14 TeV,

roughly seven times that of the Tevatron. Table 1.1 summarizes some relevant LHC parameters.

Parameter	**p-p**	**Pb-Pb**	**M.U.**
Energy per nucleon (TeV)	E	7	2.76
Design Luminosity ($cm^{-2}s^{-1}$)	L	10^{34}	10^{27}
Bunch Separation (ns)		25	100
Number of bunches	N_B	2080	592
Particles per bunch	N_P	1.15×10^{11}	7.0×10^{7}
β value at IP (m)	β^*	0.55	0.5
RMS of beam radius at IP (μm)	σ^*	16.7	15.9
Luminosity lifetime (h)	τ_X	15	6
Number of collisions/crossing	20	-	
Bunch length (σ_z) (mm)		53	
Beam current (mA)		560	

Table 1.1: Beam parameters of the LHC.

1.2 The Compact Muon Solenoid

The Compact Muon Solenoid [4] is one of the two general-purpose experiments that will operate at the LHC. It is a detector designed to investigate the physics phenomenology of p-p collisions, but it will operate in heavy ions mode. The experiment is located 100 meters underground along the LHC tunnel near the French village of Cessy. The requirements of good reconstruction of charged particles, high electromagnetic energy resolution, precise missing transverse energy and jet measurements, and good muon identification and pt reconstruction operate the CMS design. To achieve these goals and maintain the detector's compactness simultaneously, a high solenoidal magnetic field of 4 T has been chosen to provide large bending power. The CMS Experiment is one of the LHC's four major experiments, and it is intended as a general-purpose detector for researching the Standard Model of particle physics and pursuing novel physics beyond it.

It comprises several subdetectors arranged in a cylinder around the beam pipe. Each subdetector detects a specific physical attribute of a certain particle type. Endcap detectors are used to instrument the flat portions of the cylinder. This shape gives virtually flawless coverage throughout four solid angles with an active detector area. CMS is the second-biggest experiment at the LHC, with a length of around 29 m and a diameter of 15m.

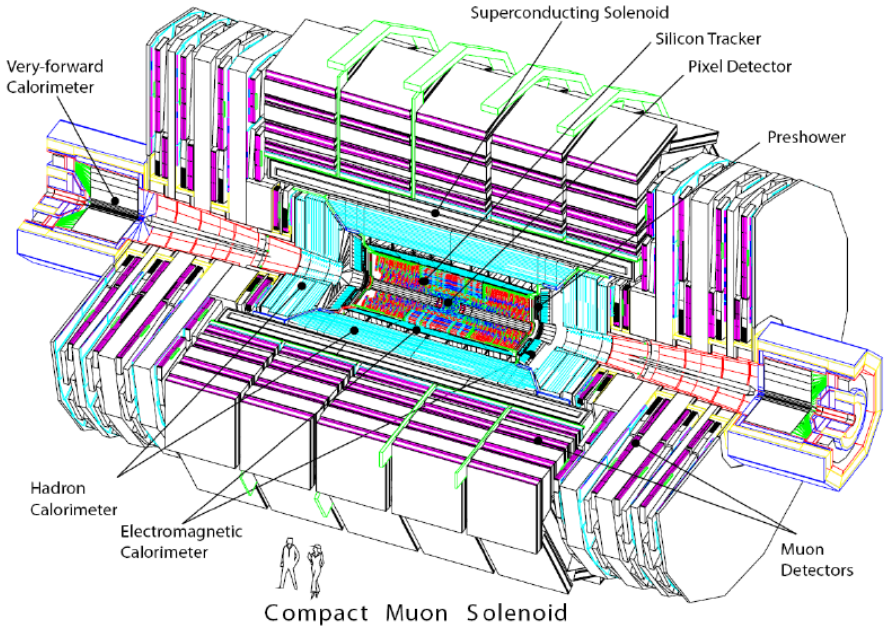


Figure 1.2: Overview of the CMS detector.

1.2.1 Overall Design Physical Requirement

The CMS detector [5, 6] (shown in Fig.1.2) has a cylindrical barrel of five slices and two disk-like endcaps. The overall length of the detector is 21.6 meters, its diameter is around 15 meters, and its total weight is approximately 12500 tons. Because of the difficult operational environment, high-granularity detectors with good time resolution must be utilized to limit occupancy. Furthermore, the large radiation flux projected at the LHC design luminosity (1-2 kGy/year) necessitates the inclusion of radiation-hard components, particularly in the central tracking system.

To achieve the goals of the LHC physics program, the detector must meet the following requirements:

- Excellent muon identification and high-resolution momentum measurement over a wide range of angles and momenta allows for precision reconstruction of di-muon invariant masses (1% resolution at 100 GeV) and clear muon charge determination up to 1 TeV/c.
- High reconstruction efficiency and precise momentum measurement for charged particles in the inner tracking system, with particular emphasis on charm, bottom jet triggering, and offline tagging.

- High electromagnetic resolution at 100 GeV allows di-photon and di-electron invariant mass reconstruction with 1% precision. Excellent geometrical coverage, efficient photon/lepton isolation at high luminosity, and strong π^0 rejection capacity.

- Hadron calorimeters with a high degree of hermeticity, comprehensive geometrical coverage, and precise segmentation to ensure accurate missing transverse energy measurements and high-resolution dijet mass reconstruction.

The compactness of the detector is ensured using a high field superconducting solenoid, 16 m long and of 6 m inner diameter, able to generate a field up to 4 T. A silicon-based inner tracking system, a homogeneous $PbWO_4$ scintillating electromagnetic calorimeter and a high hermeticity brass/scintillator sampling hadron calorimeter are accommodated in the solenoid bore. The magnetic field outside the solenoid is strong enough to saturate the iron return yoke, where a complex muon spectrometer, based on four layers of Drift Tubes detectors and Cathode Strip Chambers, respectively, is placed in barrel and endcaps. A Resistive Plate Chamber complements the other muon subdetectors, ensuring redundancy and improving trigger abilities. A longitudinal view of a quarter of the detector and a transversal view of the barrel region are given in Fig.1.3.

In CMS, a right-handed coordinate system, centered at the nominal collision point, is defined: the x-axis points radially inward to the center of the accelerator ring, the y-axis points upward, and the z-axis is parallel to the beam pipe (pointing to the Jura mountains). The polar angle θ is measured from the z-axis using a $0 \leq \theta \leq \pi$ range, while the azimuthal angle φ is measured in the x-y plane from the x-axis in a $0 \leq \varphi \leq 2\pi$ range. Usually the polar angle is replaced by the pseudorapidity (η) defined as follows:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{1.1}$$

This is because particle production is loosely constant as a function of rapidity (y), and η is the ultra-relativistic limit of y:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \approx \eta = \frac{1}{2}\ln\left(\frac{|\vec{p}|+p_z}{|\vec{p}|-p_z}\right) = -\ln\left(\tan\frac{\theta}{2}\right). \tag{1.2}$$

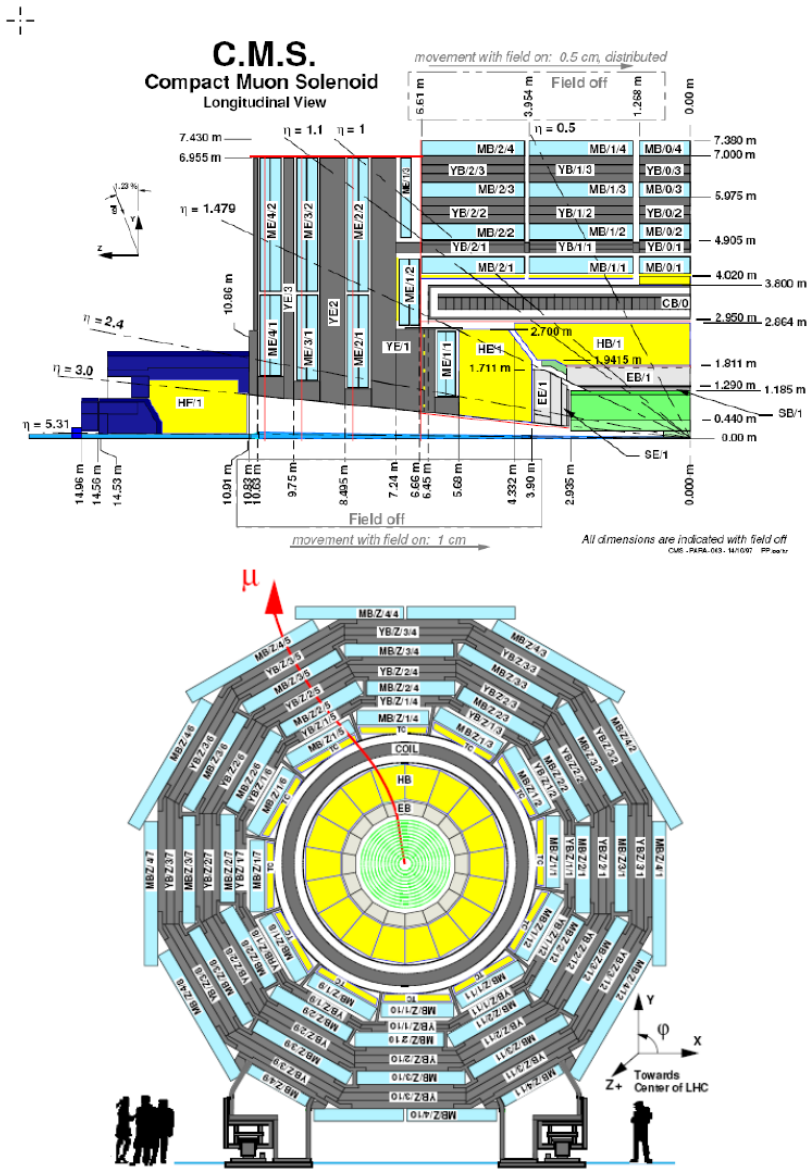


Figure 1.3: Longitudinal view of one quarter of the CMS detector(top); transversal view of the barrel region of apparatus (bottom).

In the above formulas, E, $\vec{p}$, and p_z represent, respectively, energy, 3-momentum, and z-axis momentum components of a particle coming out from the interaction point. In the following sections, a brief overview of every subdetector is given. Particular attention is also devoted to describing the CMS trigger and data-acquisition systems.

1.2.2 The Tracking System

The tracker [7, 8] is the innermost subdetector of the CMS experiment, measuring 5.8 m in length and 2.5 m in diameter. It efficiently detects and measures the trajectory of charged particles with $p_T > 1$ GeV/c, and precisely reconstructs their secondary vertices for jet-flavor tagging. It plays a crucial role in electron and muon track reconstruction and is heavily used in the high-level trigger. At LHC design luminosity, around 1000 charged particles from p-p interactions are produced every 25 ns, necessitating a high granularity and radiation-hard system. The CMS tracker comprises a fine granularity pixel detector system in its innermost parts and silicon strip modules of different pitch in the central and external parts, maintaining track occupancy around 1% during high luminosity p-p collisions and ensuring reasonable levels during Pb-Pb collisions (1% in pixels, 20% in silicon microstrip detector). High granularity results in elevated power consumption and requires efficient cooling infrastructure to prevent radiation damage. The tracker's material is minimized to reduce multiple scattering and interactions.

1.2.3 The Pixel Detectors

The pixel detector system[9] consists of finely segmented silicon pixels placed on a silicon substrate with a cell size of $100 \times 150\ \mu m^2$. It is built to ensure precise 3D vertex reconstruction to allow efficient τ and b jets identification, and it covers a pseudorapidity range up to $|\eta| < 2.5$. The small pixel size allows single channel occupancy per bunch crossing around 10^{-4} even in the expected high flux scenario (10^7 particles/s at 10 cm).

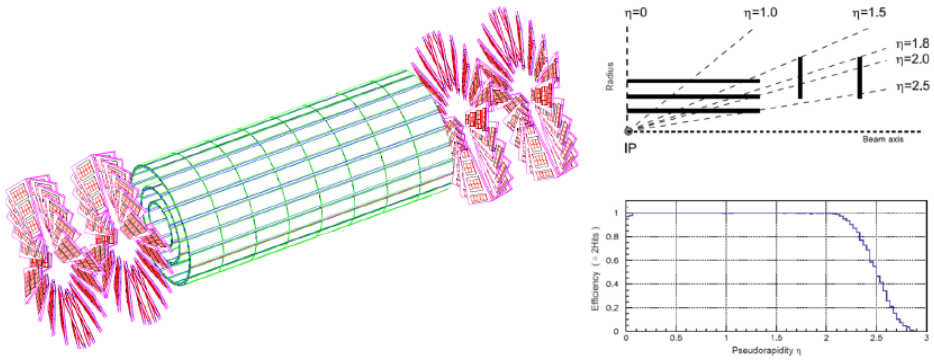


Figure 1.4: Overview of the pixel detector layout (left); longitudinal view of one-quarter of the pixel detector and its hit coverage as a function of (right).

The final layout of the pixels is shown in Figure1.4. It consists of three barrel layers (located at a mean radius of 4.4, 7.3, and 10.2 cm) and two disks (located in a radial region extending from 6 to 15 cm) in the endcaps. However, only two barrel layers and one endcap disk will be placed during the low luminosity phase. An interpolated measurement, based on the analog readout of charge deposited in nearby pixels, will ensure a hit resolution of 10 μm in the φ plane and 15 μm in the zone for the barrel layer. Correspondingly, a 15 μm and 20 μm respectively in φ and z is expected in the endcaps.

1.2.4 The Silicon Microstrip Detectors

The outermost regions of the tracking system [10]contain several layers of silicon microstrip detectors, as shown in Fig.1.5. The system is divided into the Tracker Inner Barrel (TIB), which consists of the four innermost barrel layers, and the Tracker Outer Barrel (TOB), which consists of the six outer barrel layers. The Tracker Inner Disk (TID) includes three layers of disks on each side of the TIB, while the Tracker End-Cap (TEC) is made up of nine detector disks perpendicular to the beam axis, placed after the TOB and TID. The silicon microstrip detectors cover up to $|\eta| < 2.5$, as high track density and radiation levels prevent their use beyond this range. In the TIB/TID regions, the minimum cell size is 10 cm $\times$ 80 μm, with an occupancy of about 2-3% per BX, a cell thickness of 320 μm, a pitch size of $80 - 120$ μm, and a spatial resolution of $23 - 35$ μm. In the TOB, the minimal cell size is 25 cm $\times$ 180 μm, with an occupancy of about 1% per BX, strip thickness of 500 μm, pitch size of $122 - 183$ μm, and a spatial resolution of $35 - 53$ μm. In the TEC, the four innermost rings have a strip thickness of 320 μm, while the five outermost ones have a thickness of 500 μm, with pitch sizes ranging from 97 to 184 μm.

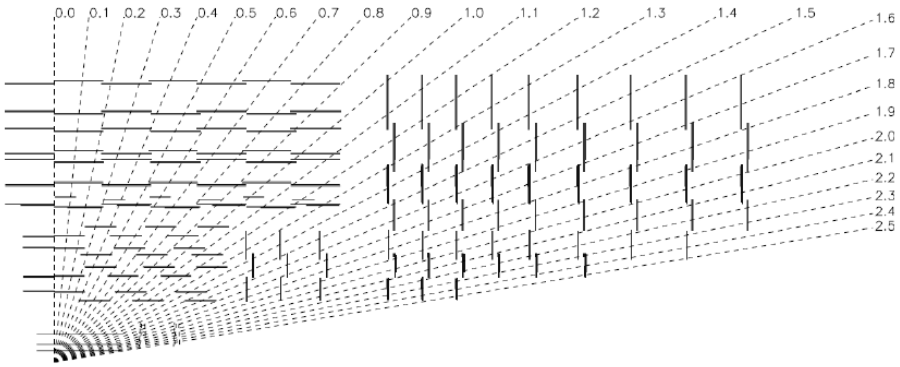


Figure 1.5: Longitudinal view of one-quarter of the silicon strips and pixels detectors.

1.2.5 The Electromagnetic Calorimeter

The Electromagnetic Calorimeter[11] (ECAL) is designed to accurately reconstruct electron and photon positions and energies, working alongside the Hadron Calorimeter for precise hadronic jet measurements. Fig. **??** is A longitudinal view of the HCAL layout. Its primary goal is to achieve 1% resolution for the $H \rightarrow \gamma\gamma$ decay channel. The ECAL is built using lead tungstate ($PbWO_4$) crystals, chosen for their radiation hardness, small Molière radius, and short radiation length, ensuring good shower containment and compactness. These crystals, which have a rapid scintillation decay time, enable 80% light collection within 25 ns. Photodetectors, specifically Vacuum Photodiodes (VPT) and Avalanche Photodiodes (APD), maintain performance in high magnetic fields and require thermal stability. The crystals are trapezoidal, with dimensions of 230 mm in the barrel and 220 mm in the endcaps. The ECAL system extends to $|\eta| < 3.0$. A pre-shower detector, consisting of a two-layer sampling calorimeter, is used for better photon and electron position measurements. The ECAL energy resolution for energies below 500 GeV can be parameterized as follows:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2$$

with $S = 2.8\%$, $N = 0.12\%$, and $C = 0.30\%$.

1.2.6 The Hadron Calorimeter

The Hadron Calorimeter [12] (HCAL) works with the ECAL to measure the direction and energy of hadronic jets and estimate missing transverse energy (E_T). It also aids in identifying electrons, photons, and muons. HCAL uses a sampling calorimeter system with brass absorbers and plastic scintillators designed for precision and compactness. The barrel (HB) covers $|\eta| < 1.4$ with segmentation for optimal di-jet separation and mass resolution. An outer calorimeter (HO) enhances hadron shower containment.

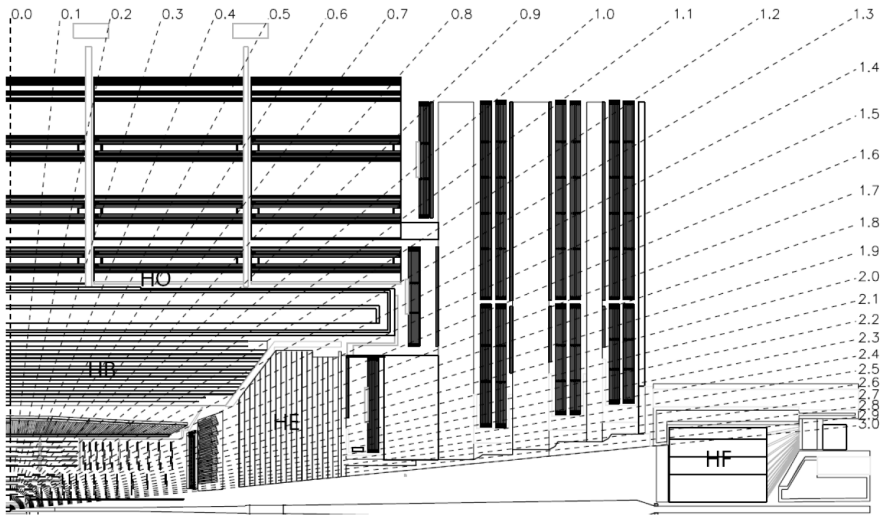


Figure 1.6: Longitudinal view of one-quarter of the HCAL subsystem.

The endcap (HE) covers $1.4 < |\eta| < 3.2$, and the forward calorimeter (HF) extends coverage up to $|\eta| < 5.2$. HCAL's energy resolution varies from $\sigma/E \approx (65/\sqrt{E} \oplus 5)\%$ in the barrel to $\sigma/E \approx (83/\sqrt{E} \oplus 5)\%$ in the HE and $\sigma/E \approx (100/\sqrt{E} \oplus 5)\%$ in the HF.

1.3 The Magnet

The CMS magnet system [13] is designed to identify the charge of high p_t muons and measure the momentum of charged particles at the LHC with high precision. It requires a momentum resolution of $\Delta p/p \approx 10\%$ for momenta up to 1 TeV/c and a 1% resolution for 100 GeV/c particles. To meet these requirements and ensure detector compactness, a high-field, modest-sized, superconducting solenoid has been developed. The CMS magnet is 12.9 m long, has

an inner diameter of 5.9 m, and generates a 4T magnetic field. The magnetic flux is returned using an iron yoke, which also hosts the muon spectrometer. The 2T residual field in the iron provides sufficient bending power for efficient p_T-based muon trigger selection in the $|\eta| < 2.4$ region. Table 1.2 lists the magnet parameters.

Parameter	**Value**
Field	4 T
Residual field in the yoke	2 T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Operation temperature	4.5 K
Nominal current	19.14 kA
Stored energy	2.6 GJ
Hoop stress	64 atm

Table 1.2: Parameters of the CMS superconducting solenoid.

1.4 The Muon System

Many significant physical processes anticipated at the LHC will result in final states involving high p_T muons. Therefore, a robust and redundant muon spectrometer is crucial for precise muon identification, high-resolution p_T measurements, and effective trigger capabilities.

The muon system [14] is the outermost group of subdetectors in the CMS experiment, covering a pseudorapidity range up to $|\eta| < 2.4$. Its layout is depicted in Fig. 1.7. The system includes three types of gaseous detectors, selected based on the extensive area to be covered and the varying conditions of radiation and magnetic fields across different η regions.

Drift Tube Chambers (DTs) are employed in the barrel region ($|\eta| < 1.2$), where they manage low track occupancy and residual magnetic fields. For the endcaps ($0.8 < |\eta| < 2.4$), Cathode Strip Chambers (CSCs) are used due to their capability to handle high particle fluxes and the non-uniform magnetic field present in these areas. To enhance redundancy and improve trigger capabilities, Resistive Plate Chambers (RPCs) supplement both DTs and CSCs in the barrel and endcaps, covering up to $|\eta| < 2.1$. Although RPCs offer only coarse spatial resolution, their fast response and excellent time resolution ensure unambiguous bunch-crossing

identification for the muon trigger.

For muons with transverse momentum (p_T) up to approximately 200 GeV/c, the system resolution is primarily limited by multiple scattering before the particle reaches the first spectrometer station. At higher p_T, the precision of chamber measurements becomes more dominant. Figure 1.8 illustrates the overall p_T resolution as a function of η, including resolutions obtained using information solely from the tracker or the muon system. It is clear that at low transverse momentum, tracker precision prevails, whereas at higher p_T, combining data from both the tracker and the muon system enhances the overall resolution.

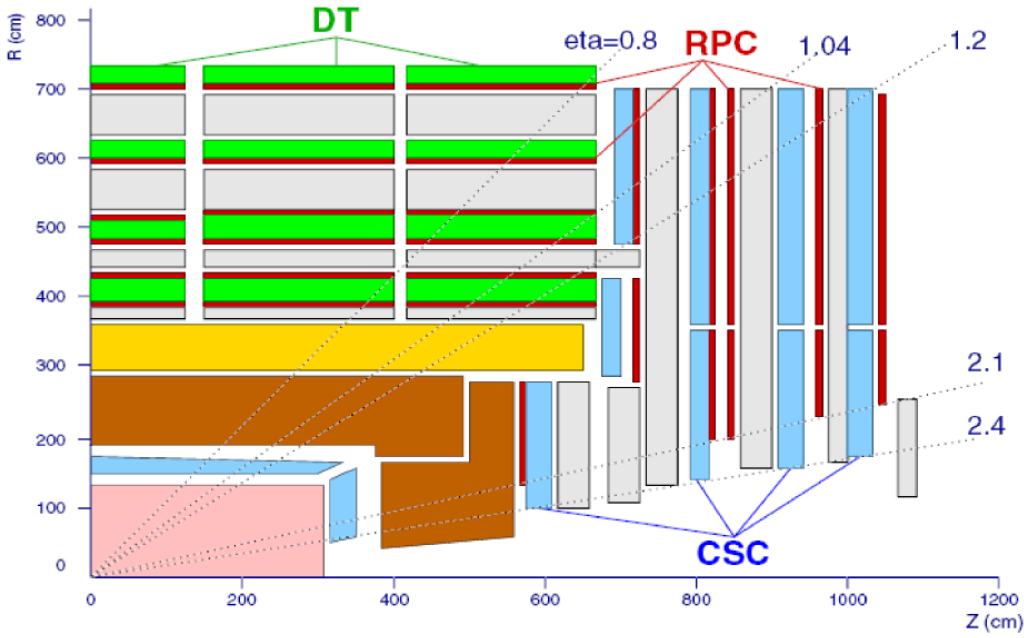


Figure 1.7: Longitudinal view of one-quarter of the muon spectrometer.

1.4.1 The Drift Tube Chambers

The barrel region of the muon system [15] is characterized by low residual magnetic field and occupancy, making Drift Tube (DT) technology suitable for precise spatial measurements and extensive surface coverage.

The layout of the DT system is depicted in Figure 1.9. It follows the segmentation of the yoke and consists of five iron wheels, each divided into 12 azimuthal sectors covering approximately 30° each. Each wheel contains four concentric rings of chambers, referred to as stations,

and named MB1, MB2, MB3, and MB4 (where MB stands for Muon Barrel). Each station comprises 12 DT chambers, except for MB4, which includes 14 chambers.

The fundamental detector element of the DT system is a rectangular drift tube cell with a transverse size of 4.2 cm × 1.3 cm and a length ranging from 2 to 4 m. These cells are filled with an 85%/15% Ar/CO_2 gas mixture and arranged parallel to form detection layers. Groups of four layers form a superlayer, and a block of two superlayers measuring the φ coordinate, along with one superlayer measuring the z coordinate, constitute a DT chamber. The only exception is the MB4 stations, where only the two φ superlayers are present. A schematic layout of a single DT chamber is shown in Fig. 1.10.

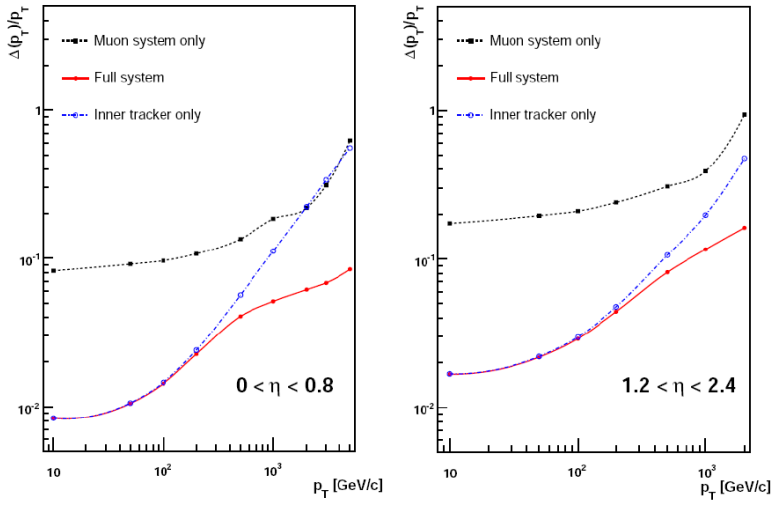


Figure 1.8: Muon transverse momentum resolution as a function of pT in the barrel (left) and endcap (right) region; results obtained using tracker only, muon system only, and a combination of muon spectrometer and tracker information are shown.

DT cells exhibit an efficiency of 99.8% and a spatial resolution of around 200 μm, resulting in a radial resolution of 100 μm for an 8-point φ reconstructed segment. A single DT chamber schematic layout is shown in 1.10.

1.4.2 The Cathode Strip Chambers

The high magnetic field [16] and particle rate expected in the muon system endcaps do not allow the use of drift tube detectors for measurements at large η values. Therefore, Cathode Strip Chamber (CSC) detectors have been adopted. CSCs are gaseous ionization detectors

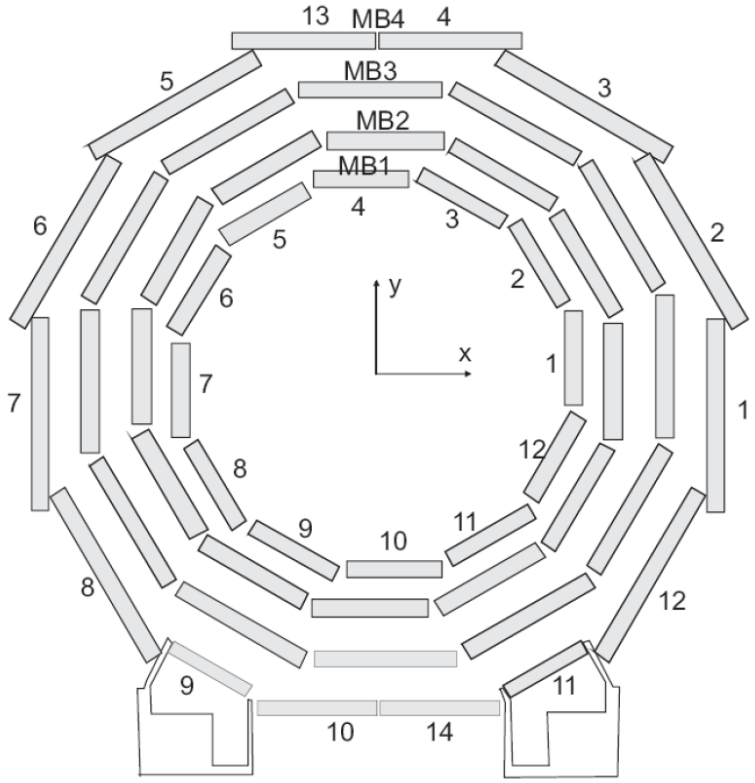


Figure 1.9: Transversal view of the CMS DT system. Station and sector numbers are shown in the figure.

operating in avalanche mode and are characterized by a short drift length. Information about the position of the incoming particle is collected both in the anode wire and on a set of finely segmented cathode strips. The cathode strips allow for the center of gravity measurements, ensuring high position resolution, while the anode wires provide a fast signal suitable for trigger purposes.

The layout of the CSC subsystem is shown in Fig. 1.11(left). Trapezoidal detector chambers are arranged to form four disks of concentric rings placed between the endcap iron yokes. All rings, except for the third of the first endcap disk (ME1/3), are arranged staggered in φ to avoid azimuthal dead regions. The innermost ring of each disk hosts 18 chambers, except for the first one, which hosts 36 chambers like the rest of the system.

One single chamber comprises six layers of 9.5 mm thick arrays of anode wires enclosed between two cathode planes, as shown in Fig. 1.11(right). One of the cathodes features fine strip segmentation, enabling precise radial measurements, while the anode wires, placed perpendic-

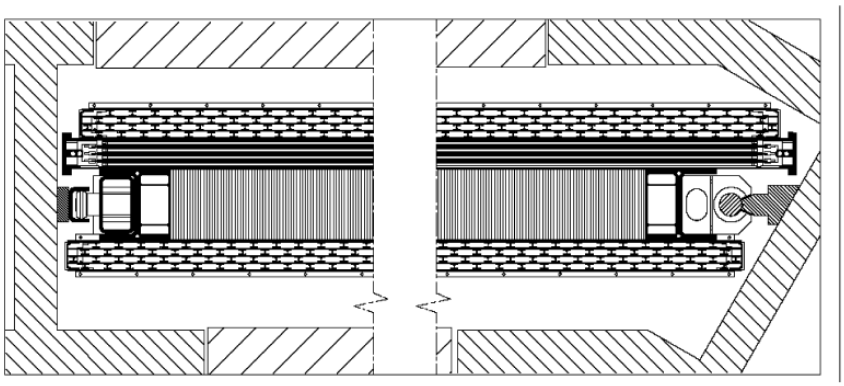


Figure 1.10: Schematic view of a CMS drift tube chamber.

ular to the strips, measure the r coordinate. Anode information is collected in groups of 5 to 16 wires to reduce the number of readout channels. The chambers are filled with a 30%/50%/20% $Ar/CO_2/CF_4$ gas mixture. The ME/1 disk operates in higher magnetic field conditions than the rest of the system, so the anode wires are tilted by 20° to compensate for the Lorentz drift effect. The gap thickness is also reduced to 6 mm, doubling the number of strips in the $|\eta| > 2.0$ region.

Strip measurement resolution varies from approximately 70 μm for the innermost stations to approximately 150 μm for the outermost ones. In contrast, the r coordinate can be determined with a precision of approximately 0.5 cm.

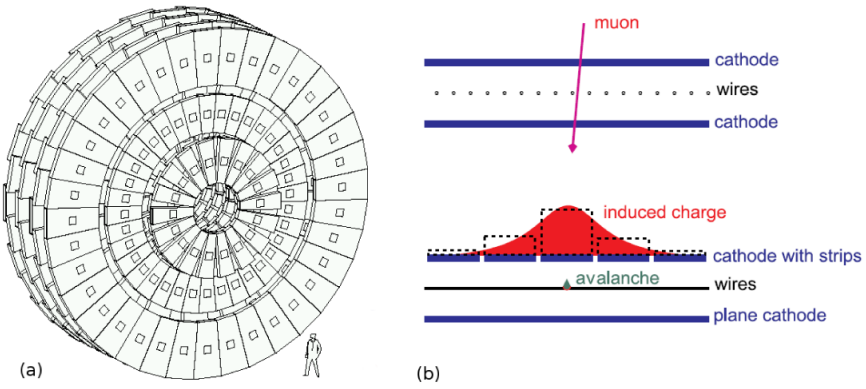


Figure 1.11: Layout of the CSC subsystem (left). Schematic overview of a CSC chamber (right).

1.4.3 The Resistive Plates Chambers

Resistive plate chambers (RPCs)[17] are utilized in both the barrel and endcaps, supplementing the Drift Tubes (DT) and Cathode Strip Chambers (CSC) systems to ensure robustness and redundancy in the muon spectrometer. RPCs are gaseous detectors characterized by their coarse spatial resolution, but they excel in precise time measurements comparable to those provided by scintillators. This enables accurate bunch crossing (BX) identification for the muon trigger system.

The CMS uses double-gap RPC chambers constructed from four bakelite planes, which alternate to form two 2 mm thick gas gaps, as illustrated in Fig. 1.12. These planes are coated with graphite to create electrodes set at a potential difference of 9.5 kV. The central part of the chamber is equipped with insulated aluminum strips that collect signals generated by crossing particles. The design choice of using double-gap chambers enhances the induced signal.

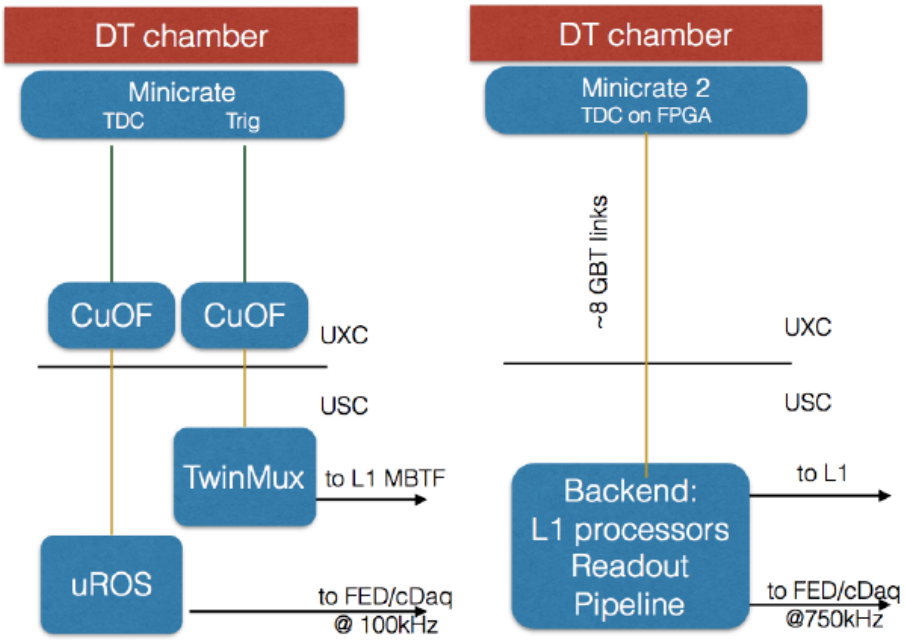


Figure 1.12: Schematic view of a CMS double gap RPC.

In the barrel region, the strips are rectangularly segmented (12.1 to 41 cm wide and 80 to 120 cm long) and aligned along the beam axis. At the same time, the endcaps feature trapezoidal-shaped strips covering an approximate range of $\Delta\varphi = 5 - 6°$ and $\Delta\eta = 0.1$. The strip length constraint makes measurements in the η coordinate impossible. The gas mixture in the gap

comprises 90% freon ($C_2H_2F_4$) and 5% isobutane (i-C_4H_{10}). The detector operates in avalanche mode instead of the more common streamer mode to sustain higher rates, necessitating improved electronic multiplication due to reduced gas multiplication.

The system layout in the barrel follows the DT segmentation. Two RPC stations are attached to each side of a sector's two innermost DT chambers, while a single RPC is attached to the inner side of the third and fourth DT chambers. This setup extends the low p_T reach of the trigger system in the barrel, allowing the detection of low p_T muons using four RPC layers before they stop in the iron yoke. This configuration is unnecessary in the endcaps since higher η values correspond to higher total momentum for the same p_T.

1.5 The Upgraded OBDT board Electronics Architecture

The Phase-2 upgrade of the CMS detector at the Large Hadron Collider (LHC) involves significant improvements to the Drift Tubes (DT) sub-detector. This upgrade is necessary to accommodate the increased data and trigger processing demands posed by the High Luminosity LHC (HL-LHC) environment. The existing on-detector electronics, known as MiniCrate (MiC), will be replaced by a more advanced and simplified system called the On-Board Electronics for Drift Tubes (OBDT) board. The OBDT board is designed to enhance data processing capabilities, improve maintainability, and ensure reliability in the high-radiation environment of the HL-LHC . Figure 1.13 shows the schematic design of the drift tube OBDT board at INFN Padova.

1.5.1 Design and Functionality of the OBDT Board

The OBDT board is specifically designed for the HL-LHC environment and will serve as the sole type of on-detector electronics for the DT subdetector. Built around a Microsemi PolarFire flash-based FPGA, the OBDT board hosts 240 Time-to-Digital Converters (TDCs) with nanosecond resolution, ensuring the precise timing measurements necessary for muon tracking [18].

Figure 1.13: Schematic OBDT board on the drift tube INFN Padova.

1.5.2 Key Features of the OBDT Board

- **High Precision Time Digitization:** The FPGA on the OBDT board digitizes signals from the DT front-end comparators, utilizing its TDCs to achieve nanosecond timing accuracy [18].
- **High-Speed Data Transmission:** Data processed by the OBDT board are transmitted via high-speed optical links, using QSFP+ transceivers capable of up to 12 Gbps per lane, ensuring rapid data transfer to the backend electronics [18].
- **Control and Timing:** The integration of GBTx and SCA chips allows for comprehensive control over slow signals, time distribution, and analog signal generation, enhancing the flexibility and reliability of the system [18].
- **Radiation Hardness:** The main component of the OBDT board is the flash-based PolarFire FPGA from Microsemi, selected for its good resistance to radiation. This is crucial for the LHC's high-radiation environment [].

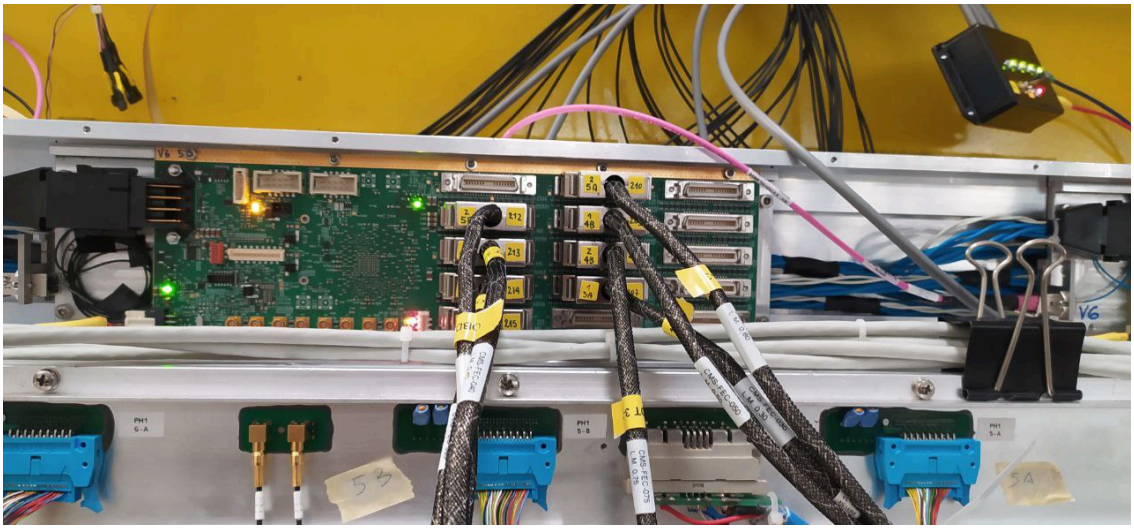


Figure 1.14: Schematic view of an OBDT board at INFN Padova.

1.5.3 Radiation Testing and Validation

Given the high-radiation environment of the HL-LHC, radiation hardness is a critical requirement for the OBDT board. Extensive radiation tests were conducted at the INFN-TIFPA Protontherapy Centre in Trento, Italy, where the OBDT board was exposed to proton beams with energies typical of the LHC environment. The total dose delivered to the electronics during these tests was 191.6 Gy, much higher than the 0.5 Gy expected over ten years of HL-LHC operation [19].

During testing, the OBDT board demonstrated excellent resilience to radiation-induced errors. The main issues identified were single event upsets (SEUs) in the FPGA and occasional failures in the safety comparator circuit, which were within acceptable limits for the operational environment. These errors were manageable within the system's design, demonstrating the OBDT board's capability to handle the HL-LHC's demanding conditions [19].

1.5.4 Development and Future Implementation

Initial prototypes of the OBDT board have been developed and tested using a Virtex7 FPGA to simulate the new electronics environment. The full assembly and further testing of the OBDT boards are planned as the next steps, with installations in a DT sector scheduled during LS2 (2019-2020) [18].

The full deployment of the OBDT boards will simplify the CMS DT electronics system

by eliminating the need for separate trigger paths and allowing the L1 Trigger Track Finders direct access to complete datasets. This integration will optimize performance and enhance the reliability of muon detection in the CMS experiment [18].

2

Theoretical Basis and Mathematical Tools

2.1 Fourier Transform and Signal Analysis

The Fourier Transform [20] is a fundamental mathematical tool used in signal processing to analyze the frequency components of a signal. Named after the French mathematician Jean-Baptiste Joseph Fourier, this transform converts a time-domain signal into its frequency-domain representation. The importance of the Fourier Transform lies in its ability to decompose complex signals into a sum of sinusoidal waves, each characterized by a specific frequency, amplitude, and phase.

Mathematically, the continuous Fourier Transform $\mathcal{F}\{x(t)\}$ of a signal $x(t)$ is defined as:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

where:

- $X(f)$ is the frequency-domain representation of the signal.
- $x(t)$ is the time-domain signal.
- f is the frequency.
- j is the imaginary unit.

The magnitude of $X(f)$ at a specific frequency f indicates the strength of that frequency component in the original signal. The angle of $X(f)$ gives the phase shift of the sinusoidal component relative to the origin.

It provides a complete description of the signal in terms of its frequency components, making it easier to analyze and manipulate. Signals are often discrete and of finite length in practical applications. Therefore, the Discrete Fourier Transform (DFT), a sampled version of the continuous Fourier Transform, is commonly used.

2.1.1 Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) [20] is used to analyze discrete signals. Given a discrete signal $x[n]$ with N samples, the DFT transforms it into a sequence of complex numbers $X[k]$ representing the signal's frequency spectrum. The DFT is mathematically expressed as:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1$$

Here, k corresponds to the frequency index, and N is the total number of samples in the signal. While the DFT is a powerful tool, its computational complexity can be a limitation, especially for large datasets. This leads to the need for a more efficient algorithm, the Fast Fourier Transform (FFT).

2.1.2 Fast Fourier Transform (FFT)

The Fast Fourier Transform (FFT)[20] is an optimized algorithm for efficiently computing the Discrete Fourier Transform (DFT). Introduced by Cooley and Tukey in 1965, the FFT reduces the computational complexity of the DFT from $O(N^2)$ to $O(N \log N)$, making it feasible to analyze large signals in real-time applications.

The FFT exploits the symmetry and periodicity properties of the DFT to reduce the number of calculations required. Several variants of the FFT algorithm, including the Radix-2, Radix-4, and Split-Radix algorithms, each designed to optimize performance under different conditions.

For a sequence $x[n]$ of length N, where N is a power of 2, the FFT recursively divides the DFT into two smaller DFTs of length $N/2$ until the problem size is small enough to be solved directly. This approach dramatically reduces the computational complexity from $O(N^2)$ to $O(N \log N)$.

2.2 Short-Time Fourier Transform (STFT)

The **Short-Time Fourier Transform (STFT)**[21] is a fundamental signal processing technique used to analyze the frequency content of a signal over time. It allows for the examination of localized changes in the spectral properties of a signal by dividing the signal into short, overlapping segments and applying the Fourier transform to each segment. This process provides a time-frequency representation of the signal, particularly useful for non-stationary signals where the spectral characteristics vary over time.

2.2.1 STFT as a Transform

The STFT of a signal $x[n]$ is defined as the discrete Fourier transform (DFT) of a windowed segment of the signal. Mathematically, it can be represented as:

$$X[k, m] = \sum_{n=m}^{m+(N-1)} w[n - m]x[n]e^{-j\omega_k(n-m)}$$

Where:

- $X[k, m]$ is the STFT of the signal at frequency bin k and time index m.
- $w[n - m]$ is the window function applied to the signal segment.
- $\omega_k = \frac{2\pi k}{N}$ represents the discrete angular frequency.
- N is the length of the windowed segment.

The window function $w[n]$ plays a crucial role in the STFT by determining the trade-off between time and frequency resolution. Common choices for the window function include the Hamming, Hanning, and rectangular windows, each offering different spectral properties.

2.2.2 STFT as a Linear-Frequency Filterbank

The STFT can also be interpreted as a bank of bandpass filters that decompose the signal into various frequency components. Each filter is centered around a specific frequency, and the output of each filter corresponds to the energy of the signal at that frequency over time. This perspective is useful for understanding how the STFT captures the signal's frequency content and can be visualized as a series of overlapping bandpass filters.

In this filterbank interpretation, the STFT is given by:

$$X[k, m] = x[m] * h_k[-m]$$

where:

- $h_k[m] = w[m]e^{j\omega_k m}$ is the impulse response of the filter centered at frequency ω_k.
- $*$ denotes the convolution operation.

2.2.3 Inverse STFT

The inverse STFT (ISTFT) allows for reconstructing the original signal from its STFT representation. This process involves applying the inverse DFT to the STFT coefficients and overlapping the resulting segments to recover the time-domain signal. The ISTFT is essential for applications that require time-frequency manipulation and subsequent reconstruction of signals.

The ISTFT can be expressed as:

$$x[n] = \frac{1}{NW(0)} \sum_{m=n-(N-1)}^{n} \sum_{k=0}^{N-1} X[k, m]e^{j\omega_k(n-m)}$$

Where $W(0)$ is the normalization factor for the overlap and windowing effects.

2.3 Autocorrelation

Autocorrelation [22][23], also known as serial correlation, refers to the correlation of a time series with its past and future values. In time series analysis, understanding autocorrelation is crucial as it can indicate patterns and inform the development of more accurate predictive models.

2.3.1 Mathematical Definition

Autocorrelation measures the relationship between time series observations separated by various lags. Mathematically, the autocorrelation function (ACF) at lag k is defined as:

$$\rho_k = \frac{\sum_{t=1}^{T-k}(X_t - \mu)(X_{t+k} - \mu)}{\sum_{t=1}^{T}(X_t - \mu)^2}$$

Where:

- X_t is the value of the time series at time t,
- μ is the mean of the time series,
- T is the total number of observations,
- k is the lag.

2.3.2 Practical Visualization Using Plots

In practice, one of the most effective ways to identify and understand autocorrelation in time series data is through visual analysis, particularly using plots such as the correlogram. A correlogram plots the autocorrelation coefficients at different lags, providing a visual representation of how observations in a time series are related to each other over time.

The correlogram helps to identify patterns of positive or negative autocorrelation quickly:

- **Positive Autocorrelation**: When the correlogram shows positive spikes at certain lags, it suggests that high values tend to be followed by high values and low values by low values, indicating a trend or persistence in the data.
- **Negative Autocorrelation**: Conversely, negative spikes indicate that high values are likely followed by low values and vice versa, suggesting a mean-reverting behavior.

2.3.3 Importance in Time Series Analysis

Understanding and addressing autocorrelation is critical in time series analysis. By visually identifying autocorrelation using plots like correlograms and interpreting the patterns they reveal, we can improve the accuracy and reliability of our models. This approach allows for better predictions from time series data, as it accounts for the temporal dependencies often present.

2.4 Inter-Arrival Times Analysis

Inter-arrival times analysis is a crucial tool in time series analysis, particularly when studying the timing and frequency of events. The inter-arrival time is the elapsed time between consecutive events in a time series. Analyzing these intervals helps identify patterns, trends, and dependencies, providing information about the processes that generate the observed data.

2.4.1 Definition of Inter-Arrival Times

In the time series context, the inter-arrival time [23][24]is defined as the time difference between successive events. If we denote the times of events as $T_1, T_2, \ldots, T_n$, then the inter-arrival times ΔT_i are calculated as:

$$\Delta T_i = T_{i+1} - T_i, \quad \text{for } i = 1, 2, \ldots, n-1$$

where:

- T_i is the time at which the i-th event occurs,
- ΔT_i is the inter-arrival time between the i-th and $(i+1)$-th events.

2.4.2 Significance of Inter-Arrival Times in Time Series Analysis

Analyzing inter-arrival times [22]can reveal important data characteristics that are not evident when looking solely at the raw time series. Key insights include:

- **Event Regularity**: By examining the distribution of inter-arrival times, we can determine if events occur at regular intervals or if there are periods of clustering and sparsity.
- **Temporal Dependencies**: Patterns in inter-arrival times can indicate dependencies between events, such as bursts of activity or periodic behavior.
- **Process Understanding**: understanding the inter-arrival times helps model the underlying stochastic processes, such as Poisson or renewal.

2.4.3 Distribution Analysis of Inter-Arrival Times

To analyze inter-arrival times, it is common to study their statistical distribution. Key steps include:

- **Histogram of Inter-Arrival Times**: Plotting a histogram of the inter-arrival times visually represents their distribution. For example, exponential inter-arrival times suggest a memoryless process typical of Poisson processes.
- **Fit Statistical Models**: Fit different probability distributions (e.g., exponential, gamma, Weibull) to the inter-arrival times data to find the best model that describes the observed patterns.

Inter-arrival times analysis provides a robust framework for understanding the timing and frequency of events in a time series. By examining the intervals between events, we can uncover patterns and dependencies that are not immediately apparent in the raw data. This analysis is instrumental here because the timing of events is crucial, enabling the development of better models and more accurate predictions.

2.5 Machine Learning Techniques

2.5.1 Isolation Forest Algorithm for Anomaly Detection

In iForest [25],[26],[27], "isolation" means separating an instance from the rest of the data. Anomalies are different from other data points and occur rarely. The algorithm works on the idea that anomalies can be separated from the rest of the data faster because they are unique and occur infrequently. This is done by creating a series of binary trees called Isolation Trees (iTrees), which divide the data into smaller groups.

An Isolation Tree is a type of tree that divides data points into smaller groups until each point is separated. The path length from the top to the bottom of these trees shows how easy it is to separate a data point. Anomalies that need fewer divisions to be separated usually have shorter paths. By averaging the path lengths across a group of iTrees, the algorithm assigns a score to each data point, with shorter paths having higher scores.

2.5.1.1 Algorithm

The iForest algorithm has two main stages: training and evaluation.

During the training stage, a forest of trees is created by building each tree using a small, random subset of the data. This process is repeated for a set number of trees (t). The building stops when a point is isolated or a specific tree height (l) is reached. Using small sub-samples and random partitioning allows iForest to quickly build effective models with linear time complexity $O(t\psi \log \psi)$, where ψ is the sub-sample size.

In the evaluation stage, the algorithm calculates an anomaly score for each instance in the test set. The score is based on each instance's expected path length $E(h(x))$ as it moves through the iForest. Instances with shorter path lengths are more likely to be anomalies. The complexity of this stage is $O(nt \log \psi)$, where n is the size of the test set.

iForest offers several advantages, including efficiency with linear time complexity and a low constant factor, making it suitable for large datasets. It also demonstrates scalability by efficiently handling high-dimensional data and large-scale problems with minimal memory usage. The algorithm's sub-sampling approach also helps mitigate swamping and masking, ensuring robust anomaly detection.

Algorithm 2.1 iForest Algorithm

Input: Given data set X, number of trees t, sub-sampling size ψ.
Output: A set of t iTrees.

1: Initialize an empty Forest.
2: Set the height limit $l = \lceil \log_2 \psi \rceil$.
3: **for** $i = 1$ to t
4: $\quad X' \leftarrow \text{sample}(X, \psi)$
5: $\quad \text{Forest} \leftarrow \text{Forest} \cup \text{iTree}(X', 0, l)$
6: **end for**
7: **return** Forest

iForest offers several advantages, including efficiency with linear time complexity and low constant factor, making it suitable for large datasets. It also demonstrates scalability by efficiently handling high-dimensional data and large-scale problems with minimal memory usage. Additionally, the algorithm's sub-sampling approach helps mitigate swamping and masking, ensuring robust anomaly detection.

2.5.2 Local Outlier Factor(LOF)

The LOF algorithm [26],[27], helps identify outliers in data by comparing the density of points to their neighbors. This is achieved by comparing each point's local density to its neighbors'

density. Points with a much lower density compared to their neighbors are considered outliers.

Algorithm 2.2 LOF Algorithm

Input: given data set D, number of neighbors k.
Output: LOF score for each point in the dataset.

1. Find the k-nearest neighbors for each point.
2. Calculate the reachability distance between points.
3. Calculate the local reachability density (LRD).
4. Calculate the LOF score for each point.

The LOF algorithm helps identify local anomalies in datasets when density varies across the data space. However, it needs careful selection of parameters such as k, and its performance might degrade in high-dimensional spaces.

2.5.3 Gaussian mixture model

The GMM algorithm [28] is essential for using Gaussian Mixture Models, which assume data comes from a mix of different Gaussian distributions. This is useful for clustering, where each Gaussian part can show a different group in the data. The EM process helps the model adjust the parameters, like the mean, covariance, and mixing coefficients, to best match the observed data.

The algorithm begins with initial guesses for the means, covariances, and mixing coefficients in the Initialization step. Then, in the Expectation Step, the chances that each data point belongs to a certain part are calculated. In the Maximization Step, the algorithm uses these chances to update the parameters of the Gaussian distributions, making the model fit the data better. The Convergence Check ensures the algorithm stops when the parameters have settled, and further iterations won't improve the model's likelihood by much.

This structured approach to the GMM algorithm allows it to cluster complex data flexibly and efficiently. It accommodates the data's natural changes by modeling it as a mix of Gaussian distributions. We use this as one way to find noise in the frequency data.

Algorithm 2.3 Gaussian Mixture Model (GMM) Algorithm

1: **Input:** Data set X, number of components K, threshold ε, max iterations max_iterations.
2: **Initialize** means μ_k, covariances Σ_k, and mixing coefficients π_k for each k.
3: **Repeat until convergence**:
4: **E-Step:** Compute responsibilities γ_{ik}:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}$$

5: **M-Step:** Update parameters μ_k, Σ_k, π_k:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} x_i, \quad \Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^T, \quad \pi_k = \frac{N_k}{N}$$

6: Compute log-likelihood and check for convergence.
7: **Output:** Parameters μ_k, Σ_k, π_k.

2.6 One-Class Support Vector Machine (One-Class SVM)

One-Class SVM [29] is an unsupervised machine learning technique primarily used for anomaly and outlier detection in datasets where most data belongs to a single class. Unlike traditional classification tasks, One-Class SVM is designed to distinguish the "normal" class from all other possible outliers or anomalies. The One-Class SVM algorithm aims to find a decision boundary that encloses most data points in the feature space, minimizing the boundary's complexity while maximizing the margin around the data. This boundary is created using kernel functions that map the input data to a higher-dimensional space, allowing for more flexible and accurate separation.

2.6.1 Algorithm

The One-Class SVM is used in many fields to find unusual items in data. It works best when there's mostly one kind of data or when rare anomalies are hard to label. It's very flexible because it uses kernel functions for non-linear decision-making. Also, it works well with complex data that has many dimensions.

In conclusion, One-Class SVM is a strong tool for finding anomalies in datasets where most things are of one type. It does this by learning a boundary around the "normal" data, which

Algorithm 2.4 One-Class SVM Algorithm

Input: Given dataset X, kernel function K, regularization parameter ν.
Output: A trained One-Class SVM model.

1: Choose a kernel function $K(\mathbf{x}, \mathbf{x}')$ to map input data to a high-dimensional feature space.

2: Set the regularization parameter ν that controls the trade-off between the fraction of outliers and the margin size.

3: Solve the following quadratic optimization problem to find the hyperplane:

$$\min_{\mathbf{w},\rho,\xi} \left(\frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho \right)$$

4: Subject to:

$$(\mathbf{w} \cdot \varphi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, n$$

5: Calculate the decision function for new data point $\mathbf{x}$:

$$f(\mathbf{x}) = \text{sign}((\mathbf{w} \cdot \varphi(\mathbf{x})) - \rho)$$

6: **return** The trained One-Class SVM model.

helps to recognize outliers and anomalies that don't follow the usual patterns.

2.7 Deep Learning Techniques

2.7.1 Long Short-Term Memory (LSTM) Model

Long Short-Term Memory (LSTM)[30][31][32] networks are a type of recurrent neural network (RNN) architecture created to model sequential data and capture long-range connections. They were developed by Hochreiter and Schmidhuber in 1997. LSTMs solve the vanishing gradient problem found in traditional RNNs, making them highly effective for task like time series forecasting. LSTM networks are designed to hold onto information for long periods, allowing them to capture time-based connections within sequential data. Unlike traditional RNNs, which encounter problems like vanishing or exploding gradients during backpropagation, LSTMs use a gating mechanism to control the flow of information, effectively maintaining and updating the cell state over time. This ability allows LSTMs to hold onto important information for long periods while letting go of unimportant details, which is crucial for understanding long-term connections.

2.7.1.1 LSTM Architecture

The architecture of an LSTM network [33]includes three main components known as gates: the input gate, the forget gate, and the output gate. These gates control the flow of information through the cell state, as described below:

- **Input Gate**: The input gate decides which values from the input should be updated in the cell state. It uses a sigmoid activation function to regulate the amount of information to be added.
- **Forget Gate**: The forget gate determines which information from the cell state should be discarded or retained. It also uses a sigmoid activation function to output values between 0 and 1, where 0 represents "completely forget" and one represents "completely retain."
- **Output Gate**: The output gate controls the output of the cell state to the next hidden state. It combines a sigmoid activation function and a tanh function to output the next hidden state based on the cell state.

2.7.2 Mathematical Formulation

The LSTM unit operates based on the following equations, defining the operations of the gates and the update mechanism for the cell state:

$$
\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{(Forget gate)} \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{(Input gate)} \\
\tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \text{(Candidate cell state)} \\
C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \quad \text{(Updated cell state)} \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{(Output gate)} \\
h_t &= o_t * \tanh(C_t) \quad \text{(Hidden state)}
\end{aligned}
$$

where:

- f_t is the forget gate vector,
- i_t is the input gate vector,
- $\tilde{C}_t$ is the candidate cell state vector,
- C_t is the cell state vector,
- o_t is the output gate vector,
- h_t is the hidden state vector,
- W_f, W_i, W_C, W_o are weight matrices,
- b_f, b_i, b_C, b_o are bias vectors,
- x_t is the input vector at time step t,
- σ is the sigmoid activation function,
- tanh is the hyperbolic tangent activation function.

2.7.3 LSTM Algorithm

LSTM models are an advanced type of recurrent neural network. They are designed to effectively handle data that occurs in a sequence by capturing long-term connections. They have become a standard choice for many tasks involving modeling sequences and can identify machinery abnormalities by analyzing sensor data over time.

Algorithm 2.5 LSTM Algorithm

Input: Input sequence $X = \{x_1, x_2, \ldots, x_T\}$, initial hidden state h_0, initial cell state C_0.
Output: Output sequence $H = \{h_1, h_2, \ldots, h_T\}$.

1: Initialize h_0 and C_0 to zeros (or small random values).
2: **for** each time step t from 1 to T
3: Compute the forget gate:
$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
4: Compute the input gate:
$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
5: Compute the candidate cell state:
$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
6: Update the cell state:
$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$
7: Compute the output gate:
$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
8: Update the hidden state:
$h_t = o_t * \tanh(C_t)$
9: **end for**
10: **return** Output sequence $H = \{h_1, h_2, \ldots, h_T\}$

3

Experimental Setup and Noise Mitigation

3.1 Experimental Setup and equipment

3.1.1 Introduction

The main aim of this thesis is to study the impact of common mode current noise on the CMS Drift Tubes Phase 2 upgrade. Accurate measurement is vital in high-energy physics research to ensure precise experimental findings. By minimizing and managing this type of electromagnetic interference, we can improve the reliability and precision of muon detection, which could significantly impact high-energy physics research.

3.1.2 Cable Design and Configuration

Designing cabling systems for high-energy physics experiments requires a modular system architecture offering several key advantages. First, modular systems provide flexibility, allowing for easy upgrades and adjustments without extensive reconfiguration. This is particularly important in high-energy physics environments where experiments frequently change and may require different setups over time. Modular systems make maintenance easier by quickly identifying and replacing faulty components. In places like CERN, where uptime and reliability are crucial, quickly addressing issues without long downtimes is essential.

We carefully planned and implemented a structured cabling system in this experiment to

ensure the best performance. Due to the systems' complexity, meticulous planning and execution were vital to maintain signal quality and ensure electromagnetic compatibility. We used a modular system architecture, which provided the necessary flexibility to adapt to evolving experimental needs and made maintenance easier. This approach allowed us to select and integrate each component, such as cables, connectors, and related hardware, with precision, ensuring that each part was chosen for its optimal performance. By doing so, we were able to maintain compatibility across the system and minimize potential signal issues, contributing to the overall success of the experiment.

3.1.3 Common-Mode Currents on Cables

Cables must have common-mode currents by EMC rules. Various issues, including EMI, incorrect cable designs, and grounding problems, can cause these currents. These issues can arise from the large and complex cabling systems used in high-energy physics experiments such as CMS drift tube testing. Unwanted currents from EMI might damage cable shielding and outer conductors, and potential differences from incorrect grounding can result in common-mode currents. Additionally, a system may be more susceptible to severe interference from unbalanced cable configurations that result in different current flows. Finding solutions to these problems in key experimental settings is crucial to maintaining signal integrity and reducing noise.

The adverse effects of common-mode currents can significantly impact high-energy physics experiments. These effects include signal degradation leading to errors and data corruption, common-mode currents acting as sources of EMI, and radiating noise affecting other equipment and systems, leading to substantial losses.

In addition to causing overheating[34] and damage to cables and connectors, continuous common-mode currents can shorten equipment lifespans and increase maintenance expenses. These are issues that must be addressed. Solutions to common-mode current mitigation include improved cable design, adequate grounding and joining practices, and, most importantly, more sophisticated signal processing techniques. These methods are not just tools but essential for identifying and eliminating common-mode noise and improving overall signal integrity in complex cable systems. In EMC strategy, they are an indispensable component.

3.1.4 Current Probes

A current probe is a precise tool for measuring the electrical current flowing through a conductor without making direct contact or disrupting the circuit. These devices are crucial for diagnosing and fixing electrical and electronic issues. The most common current probes include clamp meters, which measure current by clamping around a conductor using magnetic induction or Hall effect sensors.

Current probes offer nonintrusive measurement capabilities, ensuring the safety and integrity of electrical systems. When choosing a current probe, the expected current range, frequency bandwidth, measurement accuracy, and the physical constraints of the application environment must be considered to enhance operational safety. One of the most helpful pre-compliance measurements is the common mode current on all cables attached to the board.

The radiation from the cable is directly related to the common mode current on that cable. Common mode current refers to an unbalanced current (current not returned) on the cable. If this current is not returned to the cable, in the case of differential-mode signals, the current flows down one wire of the cable and returns to an adjacent wire. As a result, the net current is zero, and the common-mode radiation is eliminated.

Because cables are always a significant source of product radiation, measuring the common-mode current is one of the most valuable tests for a clamp-on current probe Figure 3.1 and a spectrum analyzer. The current probe [35]is about 2 3/4 in diameter and has a 1-in hole in the center for the cable. The current has a flat frequency response from 2 to 250 MHz. The transfer impedance of the current probe is 5 Ω (+14 dBΩ). Therefore, a 1 μA current will produce a 5 μV output from the current probe.

Most common-mode cable radiation occurs below 250 MHz, making the bandwidth of the F33-1 probe usually sufficient. Its frequency response, from 40 MHz to 1 GHz, allows for a wide range of testing. The transfer impedance of the F-61 probe is 18 Ω (+25 dB Ω), meaning a 1μA current will produce an 18μV output from the probe.

It's crucial to measure the common mode current for all cables. If this test fails, the radiated emission test will also fail, emphasizing the significant role of common-mode current testing. For class B products, the current must be precisely less than 5 μA (15 μA for a class A product). These limits apply to cables that are 1 meter long or longer. For cables shorter than 1 m, the allowable current is inversely proportional to the cable length, underscoring the process's precision. A μA is equivalent to 14 dBμA.

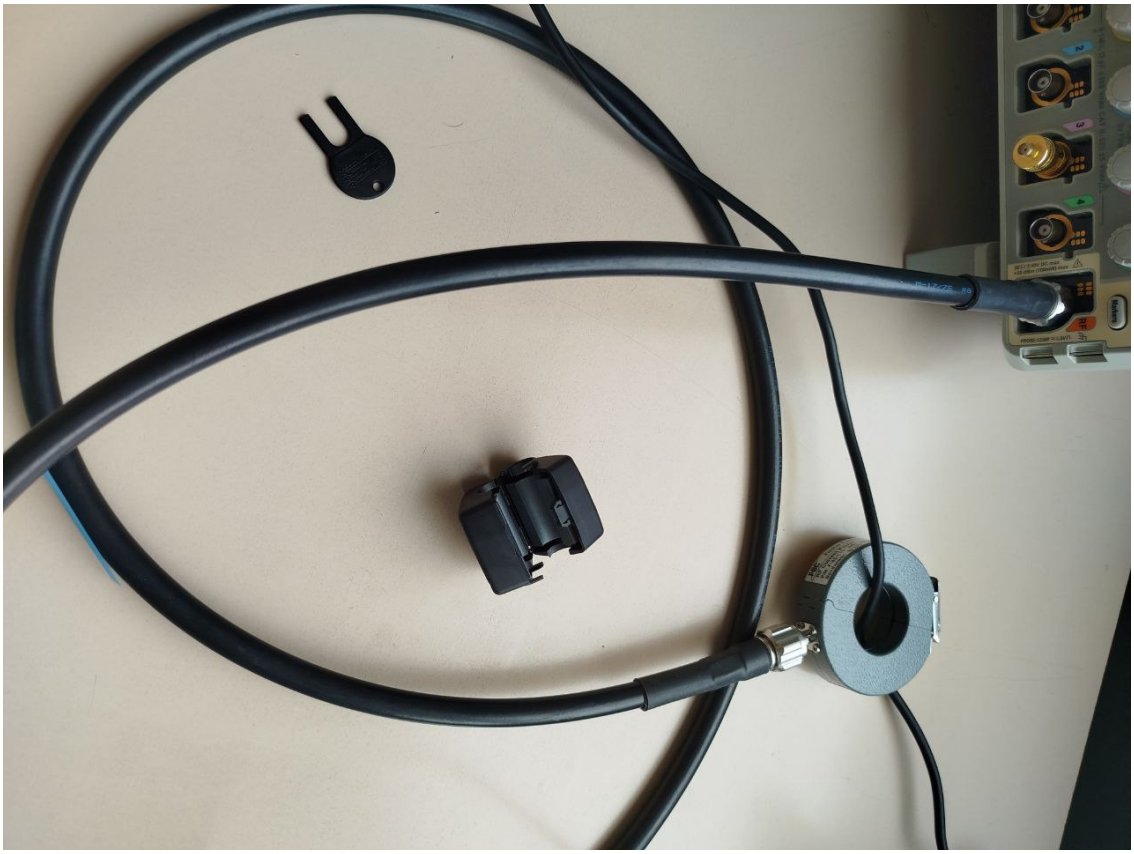


Figure 3.1: A common-mode current clamp.

Therefore, when measuring the current on the cable 1 m long or longer, the probe output voltage, read on a spectrum analyzer, should be less than 28 dBμV for the cable missions to pass the Class B limit. The Class A product's voltage reading should be less than 38 dBμv. This technique is not limited to specific cable types but works equally well on shielded or unshielded cables. It's a versatile tool that can determine the effectiveness of cable shield termination. The current probe measurement will indicate their effectiveness.

3.1.4.1 Test Pocedure

In this study, we utilized current probes to measure the current flowing through conductors without direct contact. These probes detect the magnetic field produced by the current and convert it into a measurable voltage. We employed both Class A and Class B current probes to ensure comprehensive testing and reliable results.

Class A current probes were used for high-performance applications requiring precise and

sensitive current measurements across a wide frequency range, typically from a few kilohertz (kHz) to several gigahertz (GHz). These probes are highly sensitive and can accurately detect low current levels, critical for identifying tiny emissions that could negatively impact electronic performance. With their exceptional linearity, dynamic range, and noise immunity, Class A probes are ideal for professional test facilities and manufacturers needing to meet stringent regulatory standards. In our experiment, we utilized Class A probes for thorough emission testing and analysis, as they provided the precision necessary to evaluate our devices comprehensively.

Conversely, Class B current probes are recommended for general-purpose use, with lower frequency ranges and sensitivity requirements. These probes are suitable for initial compliance checks, internal pre-compliance testing, and instructional purposes, though they lack the sensitivity of Class A probes. We employed Class B probes for our preliminary testing requirements, recognizing that they offer a cost-effective alternative with less precision than Class A probes.

The choice of current probes significantly impacted the testing results. Class A probes, with their higher sensitivity and broader frequency range, were preferred for their ability to identify low-level emissions, ensuring regulatory compliance accurately. In contrast, Class B probes provided a more economical solution with reduced precision.

During this research, I aimed to reduce the common-mode current on each cable to less than 5 μA (15 μA for a Class A product). A critical challenge I encountered was the interaction between cables, where reducing the common-mode current on one cable could inadvertently increase it on another. This potential interaction required careful consideration.

To address this, I measured the current on each cable individually using a common-mode current clamp. I employed common-mode filters, ferrite chokes, and cable shields to reduce the current below the required limit. After adjusting one cable, I proceeded to the next and repeated the process. Once all cables were addressed, I rechecked the currents, as they might have increased on previously adjusted cables. This iterative process continued until the current on each cable was under the limit. Sometimes, I had to repeat the process two or three times per cable. Ultimately, I reduced the current on each cable to under 5 μA (15 μA for a Class A device), thereby eliminating the radiated emission problem.

Cables can radiate energy coupled to them from the product and also pick up energy from external sources. To validate the common-mode current measurement, I conducted a simple test by turning off the electronic system and checking if the signal read zero. If any signal remained, it indicated external pickup. Standing waves were present on the cables under test, and I used various techniques, such as moving the current probe along the cable, to directly detect the maximum. At frequencies above 30 MHz, the current probe was moved approximately

one meter (an arm's length) to locate the maximum.

The iterative process of current reduction was crucial in ensuring that all potential sources of electromagnetic interference were addressed. By carefully monitoring and adjusting each cable, I was able to systematically minimize interference across the entire system. This method ensured compliance with regulatory standards and enhanced the electronic system's overall stability and performance.

3.2 Procedure for Noise Measurement

3.2.1 Spectrum Analyzer

Electromagnetic compatibility (EMC) testing is important for making sure that electronic devices are not affected by electromagnetic interference. The spectrum analyzer is an important tool in these tests since it compares the amplitude of an input signal vs frequency across the instrument's whole frequency range. Spectrum analyzers show us the frequency (horizontal axis) and strength (vertical axis) of signals, helping us to find, study, and control electromagnetic interference.

In my research, the spectrum analyzer played a pivotal role in evaluating the electromagnetic emissions of electronic components, particularly in the context of CMS drift tubes used in high-energy physics experiments. In my research, the spectrum analyzer was crucial for evaluating the electromagnetic emissions of electronic components, especially in the context of CMS drift tubes used in high-energy physics experiments.

The spectrum analyzer's [35] ability to measure frequencies up to at least 1 GHz was crucial for our comprehensive analysis. This wide frequency range allowed for detecting low and high-frequency noise sources that could interfere with our devices' proper functioning. The device's max-hold mode proved valuable, enabling continuous monitoring and recording of peak signal amplitudes over extended periods.

This function was important because it helped capture signals that might have been missed during a single sweep, giving a better overall picture of the emissions. The 50 Ω input impedance ensured that our standard measurement setups worked well, reducing reflections and maintaining the accuracy of our measurements. The spectrum analyzer had both peak and average detector functions. We used the peak detector for most of our initial measurements, which provided a conservative estimation of emission levels.

However, we also used peak and average detectors to assess emission characteristics for con-

ducted emission measurements thoroughly. Although a quasi-peak detector is often used for compliance testing due to its ability to weigh signals based on their pulse repetition rates, we mainly relied on peak and average detectors for initial testing. This allowed us to balance thoroughness with practical considerations, as peak and average detectors were enough to identify significant emissions and guide our efforts to reduce them.

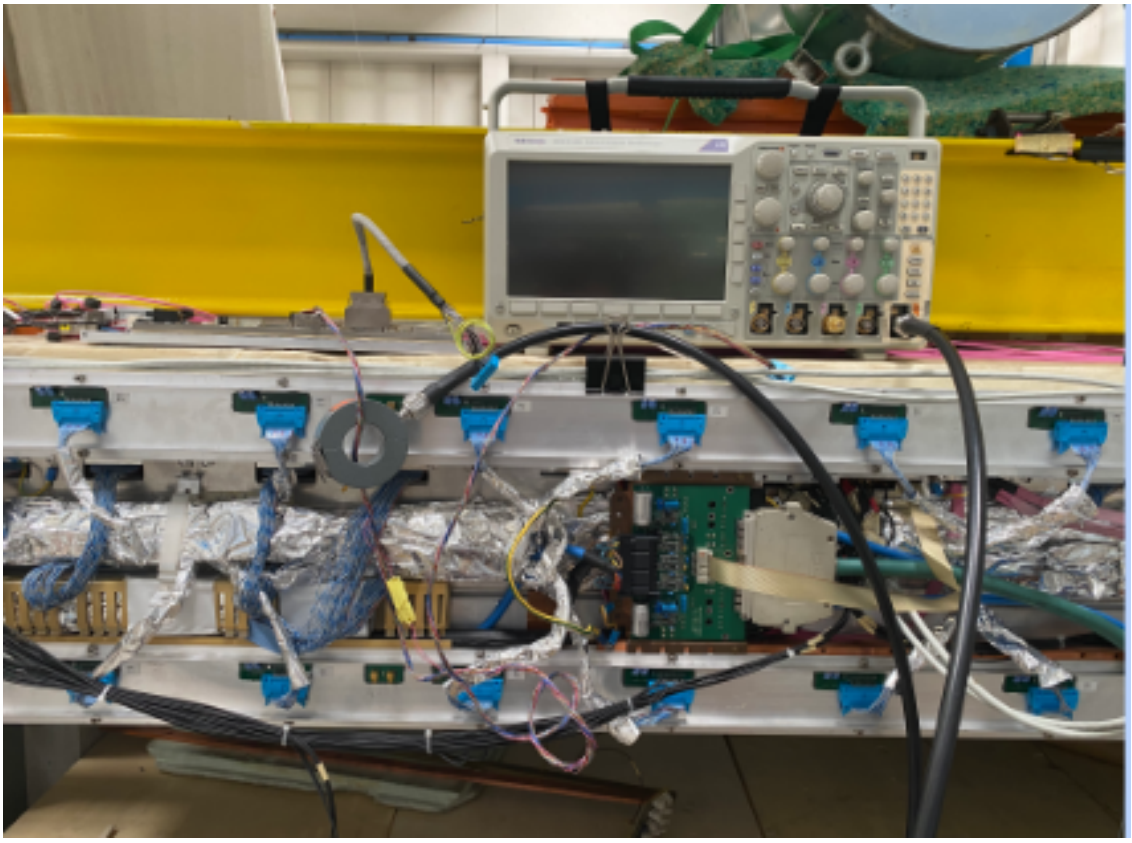


Figure 3.2: Spectrum analyzer used in identifying and quantifying common mode current noise.

My research focused on finding and fixing specific noise frequencies that caused common-mode current and other interference problems. I used a spectrum analyzer to locate these troublesome frequencies accurately. I tested various ways to reduce the interference, such as using ferrite chokes, and shielding the cables. Each time I tried a new method, I used the spectrum analyzer to check that it worked. The max-hold function on the analyzer helped capture occasional noise events, ensuring I didn't miss any major sources of interference.

We used the spectrum analyzer to confirm that our noise reduction strategies were effective. By turning off the electronic system and checking for any remaining signals, we could deter-

mine if they were caused by external noise or issues within our setup. This validation step was important for confirming that our interventions worked.

The spectrum analyzer is an important tool in our EMC testing. Its wide frequency range, max-hold mode, and dual detector functions enable us to accurately measure and analyze electromagnetic emissions, helping us ensure the reliability of our electronic components.

3.2.1.1 Test Procedure

In our research, we used advanced measurement tools, such as spectrum analyzers, to meet EMC standards and ensure that our electronic systems perform effectively. I continued this iterative process until all the cables' currents were under the limit. I had to go through the process two or three times on each cable. When I finished, the current on each cable was under 5µA (15µA for a Class A device), and the cables no longer presented a radiated emission problem. Cables radiate energy coupled to them from the product and pick up energy from external sources. I did a simple validation test for common-mode current measurement, turning the electronic system off and trying to determine if the signal read zero. If it remained, it was due to external pickup. Standing waves existed on the cables under test. Some actions, such as moving the current probe along the cable, directly detected the location of the maximum. At frequencies above 30 MHz, only the current probe was moved about a meter (an arm's length) to detect the maximum.

3.3 Noise Data Acquisition and Analysis Techniques

3.3.1 Data Acquisition

Data is crucial for understanding and addressing noise in the CMS Drift Tubes Phase 2 upgrade. The main goal is to collect accurate noise data to help identify and reduce common mode current (CMC) noise in the cabling system. We used an advanced data collection system (DAQ) to handle the precise data needed for our experiments. The DAQ system included high-resolution converters to accurately convert analog signals from various sensors and probes into digital signals. We used special current probes to measure common mode currents and voltage probes to capture the associated voltage signals. Additionally, we placed environmental sensors to track and record factors such as temperature and humidity, which could impact the measurements.

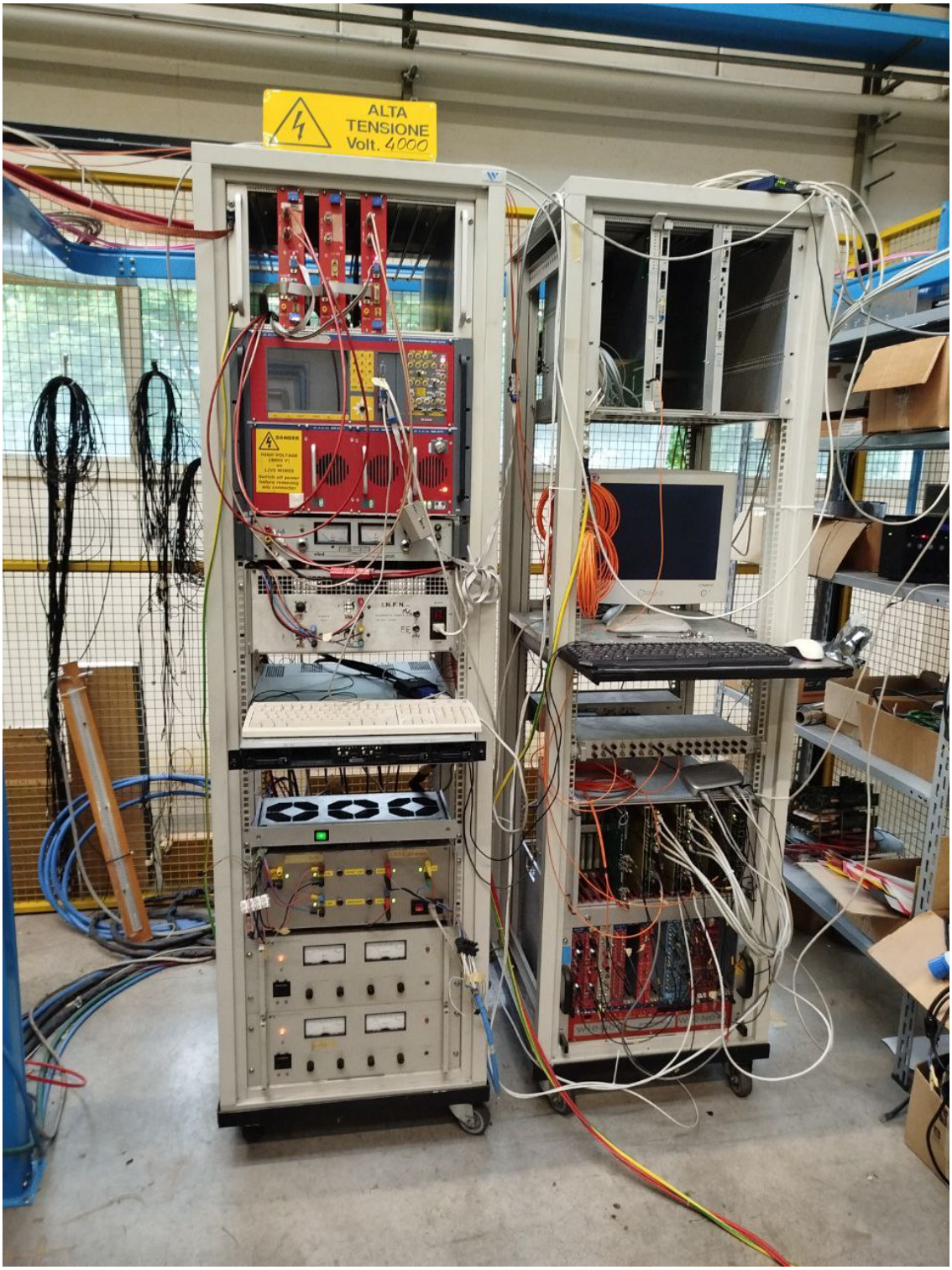


Figure 3.3: high-voltage power supply and data acquisition equipment.

We carefully calibrated all sensors, probes, and the DAQ system to ensure the measurements are accurate and reliable before collecting data. Each DAQ channel was set up to match the input signals from the probes, and we adjusted the sampling rate to capture the full frequency spectrum of the noise. This helps prevent aliasing by following the Nyquis standard. We used both continuous and triggered methods to collect data. Continuous collection helped us gather data over long periods, capturing steady noise. Triggered collection targeted sudden noise events and situations that required precise timing. We made sure to log data carefully to avoid losing any, and we stored the data in multiple places to keep it safe.3.3

Before digitizing the signals, we applied basic filtering techniques to remove unwanted frequency components. We then adjusted the signals to match the input range of the ADCs for optimal resolution and accuracy. To ensure compatibility with analysis software, we stored the data in standard file formats such as CSV and txt.

We needed to keep the environmental conditions consistent during experiments to minimize variability in the noise data. We used electromagnetic shielding to prevent external interference from affecting the DAQ system and sensors and potentially distorting the results. The high-resolution ADCs in the DAQ system allowed us to capture even the smallest signal variations accurately.

We made sure to acquire data at a rate at least twice the highest frequency of interest. In some cases, we even gathered more data than required. This allowed us to collect high-quality, reliable noise data that gave us valuable insights into the characteristics of CMC noise affecting the CMS Drift Tubes cabling. By performing detailed measurements and analysis, we developed effective noise reduction strategies, which improved the performance and reliability of the CMS detector system.

3.3.2 Noise Data Analysis Methods

In our research, we carefully analyzed the noise data collected from the CMS Drift Tubes to identify and address any potential issues that could interfere with their performance. First, we processed the raw data to filter out irrelevant noise and focus on the frequencies that could pose a risk. By converting the data into a frequency format, we could easily spot the critical noise frequencies and understand how they might affect the system. We also examined the data over time to identify sudden noise spikes that could disrupt operations.

After identifying the primary noise sources, we tested various methods to reduce or eliminate these issues. We measured the noise levels again after applying these solutions to see their

effectiveness. This analysis showed that our strategies significantly reduced the noise, especially at the most critical frequencies, which helped improve the overall reliability of the CMS Drift Tubes.

3.4 Noise Reduction Strategies

3.4.1 Shielding Techniques

Shielding is an important way to reduce electromagnetic interference. It involves techniques around cables and essential components to stop outside noise from affecting the measurements. These techniques reflect or absorb electromagnetic waves, thus protecting the signal's integrity.

A shield is a metal barrier that controls the spread of electromagnetic fields between different areas. It's used to contain electromagnetic fields and protect equipment outside the shield. A shield can also block electromagnetic radiation from entering a specific area. Regarding overall system efficiency, it's better to shield the source of noise rather than the receiver.

When making a shield, it's important to prevent electromagnetic energy from entering or exiting through cable penetrations. Cables can pick up and transfer noise from one side of the shield to the other. To maintain the shielded enclosure's integrity, it's crucial to filter out noise voltages from all cables that enter the shield. This applies to both power and signal cables. Additionally, cable shields penetrating a shielded enclosure should be bonded to the enclosure to prevent noise coupling across the boundary.

3.4.1.1 Near Fields And Far Fields

The space around a radiation source is divided into two distinct regions: the near field and the far field. The boundary between these regions is typically defined by a distance of $\lambda/2\pi$ from the source, where λ is the wavelength of the emitted radiation. The characteristics of these fields are determined by several factors, including the source (such as an antenna), the surrounding medium, and the distance from the source to the point of measurement.

In the near field, the parameters of the electromagnetic field are predominantly influenced by the source's characteristics. When the near field is primarily magnetic, it typically indicates a source with low voltage and high current, where the ratio of the electric field E to the magnetic field H is less than 377 ohms $\left(\frac{E}{H} < 377\right)$. Conversely, when the near field is primarily electric, the source tends to have high voltage and low current, with an E/H ratio greater than 377 ohms $\left(\frac{E}{H} > 377\right)$. As the distance from the source increases, the impedance of the near field gradually

transitions towards the impedance of free space, which is 377 ohms, marking the beginning of the far field.

The wave impedance of a magnetic field, such as that produced by a loop antenna, is low near the antenna and increases with distance. In the near field, the electric field strength decreases at a rate of $1/r^2$, while the magnetic field strength diminishes more rapidly at a rate of $1/r^3$. In the far field, the electric and magnetic fields decrease at a rate of $1/r$, forming a plane wave with a constant impedance of 377 ohms. The wave impedance is the ratio of the electric field E to the magnetic field H. In the far field, this ratio equals the characteristic impedance of the medium, which, for air or free space, is $Z_0 = 377$ ohms.

The field's behavior in the far field is significantly influenced by the properties of the medium, affecting both the electric and magnetic components of the wave. Electric and magnetic fields are generally more significant in the near field, while plane waves, characterized by a constant wave impedance, are a feature of the far field.

3.4.1.2 Shielding Effectiveness in Electromagnetic Fields

When evaluating how well a material blocks electromagnetic fields, it's important to consider a few things: the material's properties, how much it reduces electric and magnetic field strength, and how its shape and the frequency of the fields affect its effectiveness. Shielding effectiveness is usually measured in decibels (dB), which show how well the material reduces the field strength. This reduction is described by the shielding effectiveness (S) for electric fields from E_0 to E_1, and magnetic fields from H_0 to H_1:

$$S = 20 \log \left(\frac{E_0}{E_1} \right) \text{ dB}$$

$$S = 20 \log \left(\frac{H_0}{H_1} \right) \text{ dB}$$

Creating a shielded enclosure involves two main factors: the shielding effectiveness of the material used and the impact of any openings or gaps. The first factor provides the baseline performance, while the second factor can greatly affect effectiveness, especially at high frequencies. Shielding effectiveness depends on several different factors:

- **Geometry of the Shield:** The shape and size of the shield influence how well it reduces incident fields.
- **Frequency:** Higher frequencies generally require higher conductivity and permeability materials to maintain effectiveness.

- **Position within the Shield:** The field's measurement location inside the shield can show variations in effectiveness.
- **Type of Field:** Different strategies may be necessary for shielding electric fields versus magnetic fields.
- **Angle of Incidence and Polarization:** The direction and nature of the incident wave can affect shielding performance.

3.4.1.3 Absorption and Reflection Losses

When assessing how well a solid material without openings can block radiation (S), it's important to consider how much is absorbed (A), how much is reflected (R), and a correction factor for multiple reflections (B). The total shielding effectiveness is calculated by combining these factors:

$$S = A + R + B \,\text{dB}$$

When an electromagnetic wave passes through a material, it loses energy due to heat and other factors within the material. The relationship between the initial and final field intensities, considering absorption, is given by:

$$E_1 = E_0 e^{-t/\delta}$$

$$H_1 = H_0 e^{-t/\delta}$$

Here, δ represents the skin depth:

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}}$$

Different materials like copper, aluminum, steel, and mumetal absorb electromagnetic energy differently due to their different skin depths. For example, copper is less effective at higher frequencies and needs to be thicker to shield effectively at lower frequencies. The shielding effectiveness against electromagnetic interference depends on the shield's material, frequency, and construction.

3.4.1.4 Reflection Loss in Electromagnetic Shielding

When an electromagnetic wave encounters a boundary between two materials with different characteristic impedances, some of the wave is reflected back. This reflection loss is important for effective electromagnetic shielding. The intensity of the transmitted wave for electric and magnetic fields (E_1 and H_1) can be expressed as:

$$E_1 = \frac{2Z_2}{Z_1 + Z_2} E_0$$
$$H_1 = \frac{2Z_2}{Z_1 + Z_2} H_0$$

The reflection loss R is given by:

$$R = 20 \log \left(\frac{E_0}{E_1} \right) = 20 \log \left(\frac{Z_0}{Z_s} \right)$$

This indicates that reflection loss is more significant at lower frequencies and for materials with higher conductivity. In the near field, reflection loss depends on the source's characteristics. For point sources, reflection loss can be calculated using specific formulas based on the source's distance to the shield and the wave impedance. The frequency of the wave, the impedance of the materials, and the distance from the source all contribute to the overall outcome. Copper and aluminum, which are highly conductive materials, reflect loss more effectively at higher frequencies. Therefore, reflection loss plays a crucial role in determining electromagnetic shielding efficiency.

3.4.1.5 Effective shielded cabling in the CMS Drift Tubes

I used shielded cables to reduce electromagnetic interference and protect sensitive equipment in the experiment. The results demonstrated that the shielded cables effectively blocked or absorbed electromagnetic waves, significantly reducing unwanted noise in the measurements. In addition, I found that the effectiveness of the shielding was influenced by factors such as the material of the shield, the frequency of the electromagnetic waves, and the positioning of the cables within the shielded area. Overall, the experiment confirmed that using shielded cables can greatly enhance system efficiency by effectively targeting the noise source rather than just shielding the receiver.

3.5 Grounding

Grounding is crucial for connecting electronic devices and reducing unwanted noise to ensure system reliability. In my experiment, especially with the CMS Drift Tubes' cabling systems, a well-designed grounding system is important for reducing interference and protecting sensitive equipment. By focusing on effective grounding techniques, my goal was to enhance electromagnetic compatibility (EMC) and ensure accurate signal transmission, particularly in high-frequency environments.

Grounding can be categorized into two main types: safety grounding, which protects personnel and equipment from electrical hazards, and signal grounding, which serves as a return path for signal and power currents. In my experiment, I concentrated on signal grounding to reduce noise and improve the system's performance.

3.5.1 Effective Grounding Techniques in the CMS Drift Tubes

In the experiment, single-point grounding was used for lower frequencies, ensuring each component had a single ground reference, which reduced ground loops and minimized low-frequency noise. For high-frequency applications, multipoint grounding was employed by connecting multiple ground points to a single ground plane. This approach minimized loop areas, decreased impedance in ground return paths, and reduced the potential for radiated emissions.

Shielded cables with properly grounded shields were crucial in this experiment, as they provided a continuous path for noise to return to the ground, preventing the formation of common mode currents and reducing overall noise levels. The CMS Drift Tubes were designed with multiple grounding points connected to a single ground plane, which helped dissipate high-frequency noise and create low-impedance ground return paths.

Throughout the experiment, grounding techniques were carefully tested and adjusted to optimize noise reduction. By fine-tuning the positions and connections of grounding points, a significant reduction in common mode current noise was achieved, thereby enhancing the signals' reliability and accuracy. These results demonstrate the importance of proper grounding in maintaining complex electronic systems' overall performance and reliability.

4

Analysis of Common Mode Current Noise in CMS Drift Tube Cabling

4.1 Introduction

The CMS Drift Tubes are crucial for tracking muons in the CMS detector at the Large Hadron Collider. However, the performance of these detectors can be significantly affected by common mode current noise, particularly in the cabling infrastructure, which can create false signals. This chapter analyzes common mode noise using signal processing and advanced machine learning techniques. The goal is to identify, describe, and reduce the impact of noise on the Drift Tubes, with a specific focus on maintaining optimal performance during the Phase 2 Upgrade.

4.2 Data Collection and Preprocessing

This study gathered data from field measurements to observe the electrical signals transmitted through the cabling of the CMS Drift Tubes. Our main objective was to identify and analyze noise events, particularly those that could potentially degrade signal quality. The initial step involved decoding the raw data. This decoding process extracted essential features such as channel ID, bunch crossing, orbit count, and fine Time-to-Digital Converter (TDC) values,

making it ready for further analysis.

We created three important plots from the decoded data to fully understand the noise characteristics and how they affect the signal quality. These include the OC value distribution plot, the TDC value distribution plot, and the BX value distribution plot. Each plot gives us unique insights into different aspects of the collected data.

- **Orbit Count(OC) Value Distribution Plot(Unshielded Low Threshold Dataset):** The Orbit Count (OC) value distribution plot graph shows the number of hits recorded at different orbit counts across various channels. This graph is important for identifying noise. Any significant deviations or anomalies, like spikes at specific OC values, can indicate the presence of noise or electronic interference. If a channel shows a disproportionately high number of hits at certain orbit counts, it suggests that noise might be artificially increasing the hit count. This can make it harder to accurately reconstruct muon paths and obscure real signals, degrading the overall data quality. In Figure 4.1, please display the spike and confirm the presence of the noise.

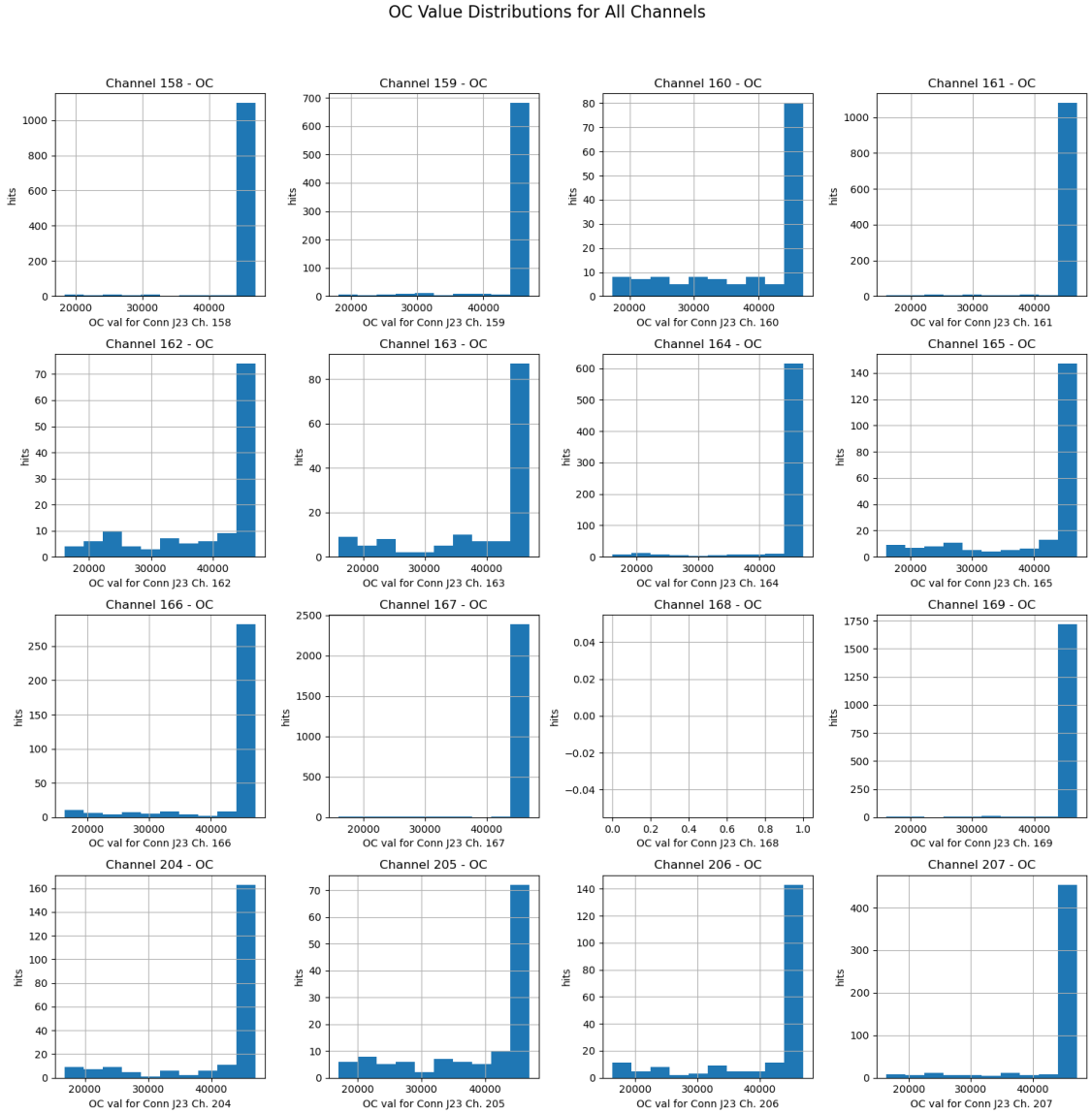


Figure 4.1: Orbit Count(OC) Value Distribution Plot(Unshielded Low Threshold Dataset).

- **Time To Digital(TDC) Value Distribution Plot(Unshielded Low Threshold Dataset):**

  The Time-to-Digital Converter (TDC) value distribution plot, figure 4.2 shows us the exact timing of signals as they pass through the cables. A good TDC plot will have clear, distinct peaks that match the expected timing intervals. However, if the plot shows a more erratic distribution with too much variation or multiple peaks, it means that there is interference affecting the timing measurements. This interference can cause signal timing changes, leading to inaccuracies in muon tracking and ultimately affecting the detector's performance.

Figure 4.2: Time To Digital(TDC) Value Distribution Plot(Unshielded Low Threshold Dataset).

- **Bunch Crossing(BX) Value Distribution Plot(Unshielded Low Threshold Dataset):**

  The Bunch Crossing (BX) value distribution plot, figure 4.3 shows the uniformity of hits recorded across different bunch crossings. Ideally, the graph should be mostly flat, showing that hits are spread out evenly across all bunch crossings. This indicates reliable and unbiased detection. If the graph shows significant variations or trends, like certain bunch crossings consistently having higher or lower hit counts, it could mean synchronization issues are affecting the timing of event detection. These irregularities can lead to data collection biases and affect muon tracking accuracy.

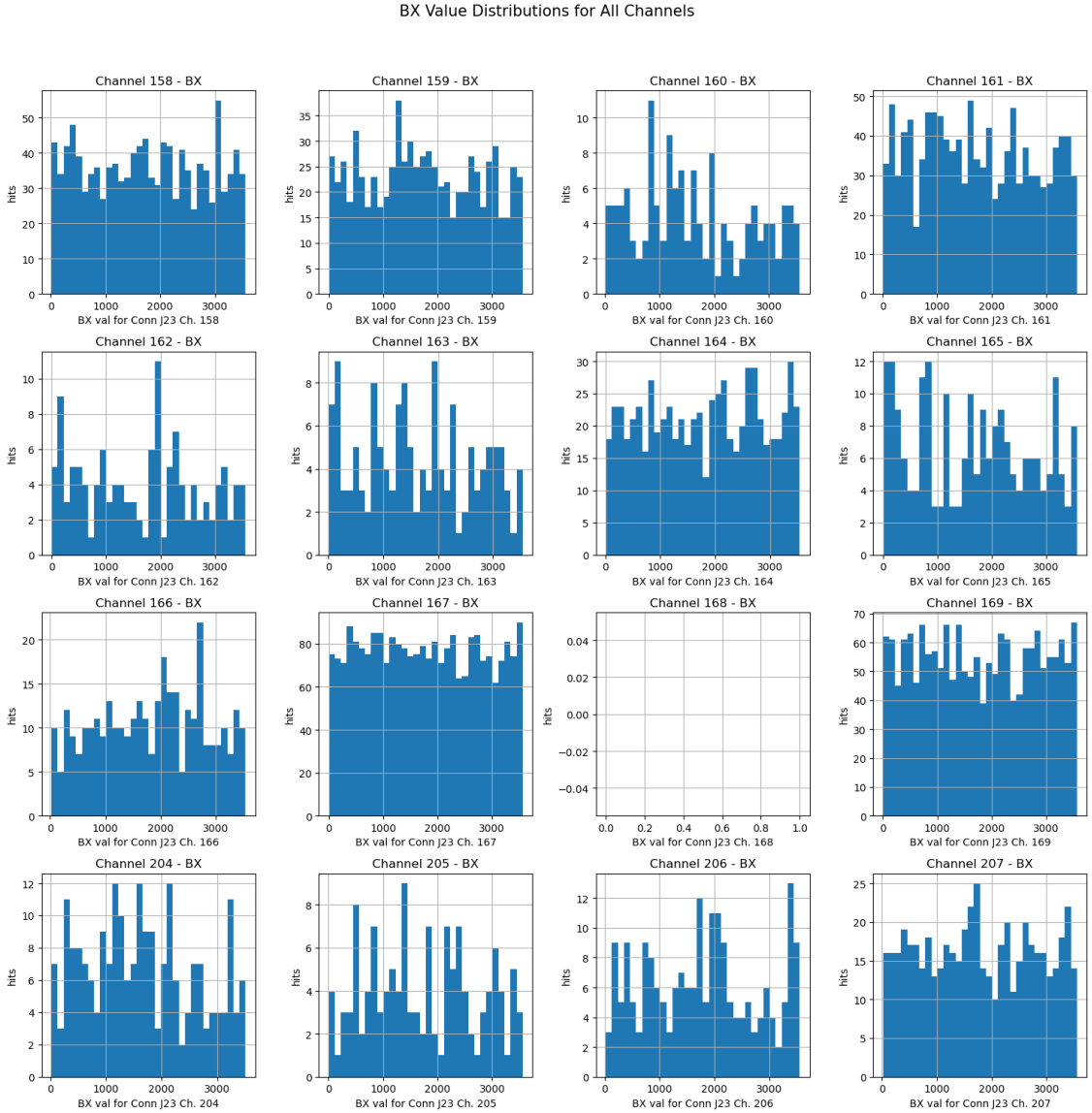


Figure 4.3: Bunch Crossing(BX) Value Distribution Plot(Unshielded Low Threshold Dataset).

Our study used four different cable setups for our research:

- **Unshielded with High Threshold:** This configuration tested the system's performance under conditions where cables were not shielded, and a high detection threshold was applied. The purpose was to understand how much noise was present in an unprotected setup when less sensitive to smaller noise signals.

- **Short Cable:** A shorter cable was employed to evaluate the effect of cable length on

noise propagation. This setup helped identify whether shorter cables reduce noise due to shorter transmission paths.

- **Unshielded with Low Threshold:** Similar to the first setup but with a lower threshold, this configuration was used to capture and analyze more subtle noise events that might be missed at higher thresholds.

- **Shielded:** Shielded cables were used to assess the effectiveness of shielding in reducing noise, particularly in the context of common mode currents that can induce unwanted signals.

The noise from common mode currents in each setup was tested and analyzed to determine the impact of different cable setups on the noise levels in the CMS Drift Tubes.

In addition to these practical setups, advanced signal processing techniques were used to improve our analysis further:

- **Fast Fourier Transform (FFT):** The Fast Fourier Transform (FFT) was used to convert the time-domain signals into the frequency domain. This method enabled us to precisely identify and characterize the dominant noise frequencies in the system. This transformation was imperative for isolating specific noise sources, including those originating from power line interference or other electromagnetic disturbances.

- **Machine Learning and deep learning Models for Noise Prediction:** Machine learning and deep learning models were explicitly used in Long Short-Term Memory (LSTM) networks to detect better and classify common mode current noise in the CMS Drift Tubes cabling. This approach helps us identify noise patterns more accurately, improve noise management, and ensure the reliable operation of CMS Drift Tubes during critical phases.

4.3 Frequency Analysis Across Different Cable Configurations

The analysis of common mode current noise across the four cable setups (Unshielded with High Threshold, Short Cable, Unshielded with Low Threshold, and Shielded) showed that each setup had different levels of noise and frequencies. This demonstrates how different cable setups can affect the suppression of noise.

4.3.1 Shielded and Short Cable Data Setup

The Shielded setup had more high-frequency noise, with the strongest peak at 748.64 Hz for channel 224. This means that while shielding reduces lower-frequency noise, it may still let higher-frequency components, possibly from outside electromagnetic interference or internal cable resonance, remain. Other significant frequencies in the Shielded data included 173.40 Hz for channel 228 and various lower frequencies (ranging from 29.65 Hz to 66.63 Hz) across different channels. These findings show that the Shielded setup captures a wide range of noise, from low to high frequencies, effectively separating noise sources but not eliminating higher frequency disturbances.

On the other hand, the Short Cable setup displayed a prominent peak at 292.76 Hz for channel 224, which is lower than the peak in the Shielded setup. This implies that using shorter cables in the Short Cable setup may reduce high-frequency noise, although it doesn't eliminate it. Most other channels in this setup had frequencies between 29.60 Hz and 39.98 Hz, indicating that short cables mainly pick up lower-frequency noise. These frequencies are depicted in Figure **??**.

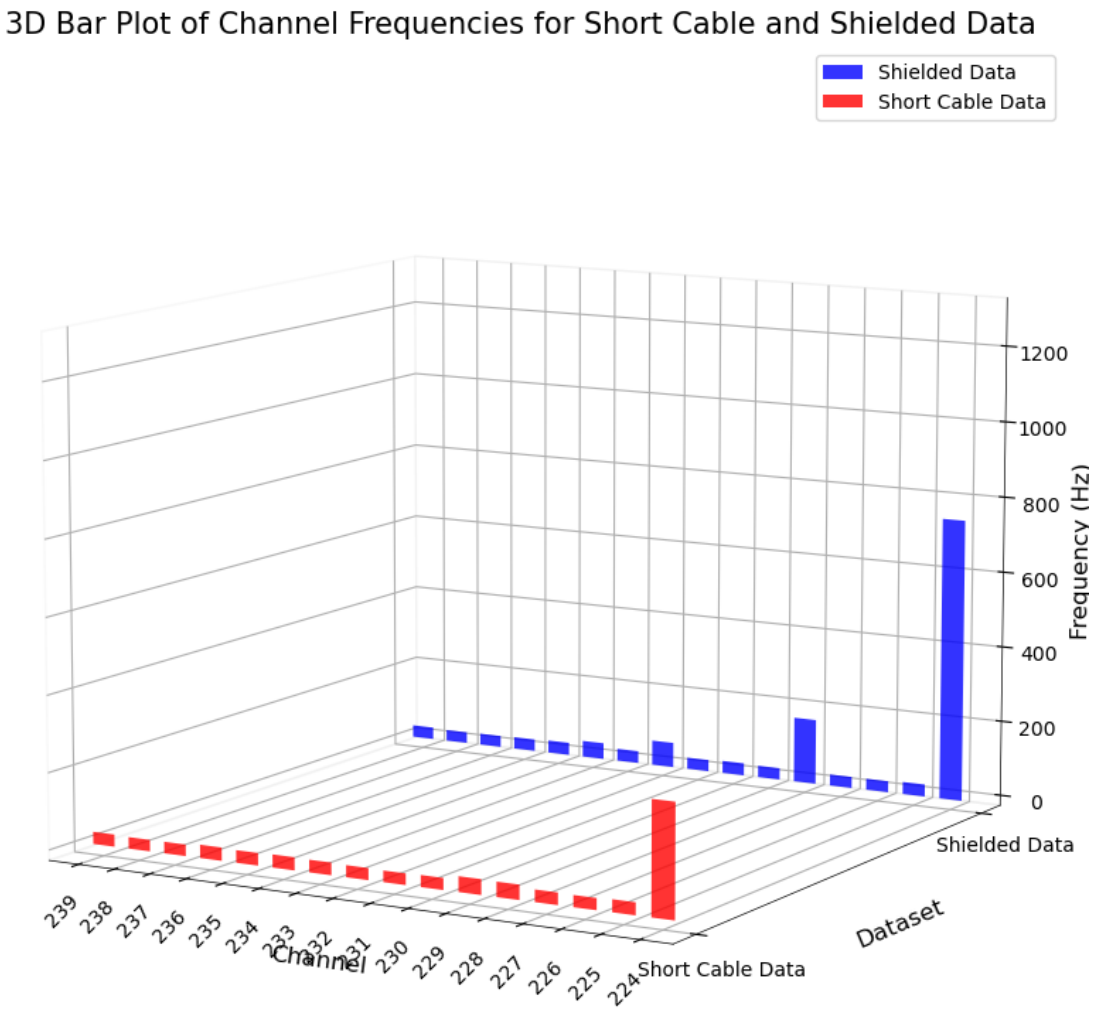


Figure 4.4: 3D Bar Plot Comparing Channel Frequencies for Shielded and Short Cable Data Configurations.

4.3.2 Low and High Threshold Unshielded Data Setup

The frequency analysis of the Low Threshold and High Threshold data reveals a contrasting distribution of noise frequencies between the two configurations. In the Low Threshold setup, the system exhibited significant sensitivity to a broad range of noise frequencies, with notable peaks at 581.50 Hz and 572.50 Hz on channels 158 and 161, respectively. The highest recorded frequency was 1229.50 Hz on channel 167, indicating that lowering the detec-

tion threshold increases the system's sensitivity to a broader array of noise, including higher frequency components. Conversely, the High Threshold data primarily captured lower frequency noise, with the highest frequency recorded at 20.56 Hz on channel 207. This suggests that the High Threshold setting effectively filters out most noise, particularly the high-frequency components. However, it may also miss subtle noise signals due to its selective nature. Consequently, The High Threshold configuration reduces interference from minor noise events, but it may miss lower amplitude noise that could affect the CMS Drift Tubes' performance during critical operations. These frequencies are depicted in Figure 4.5.

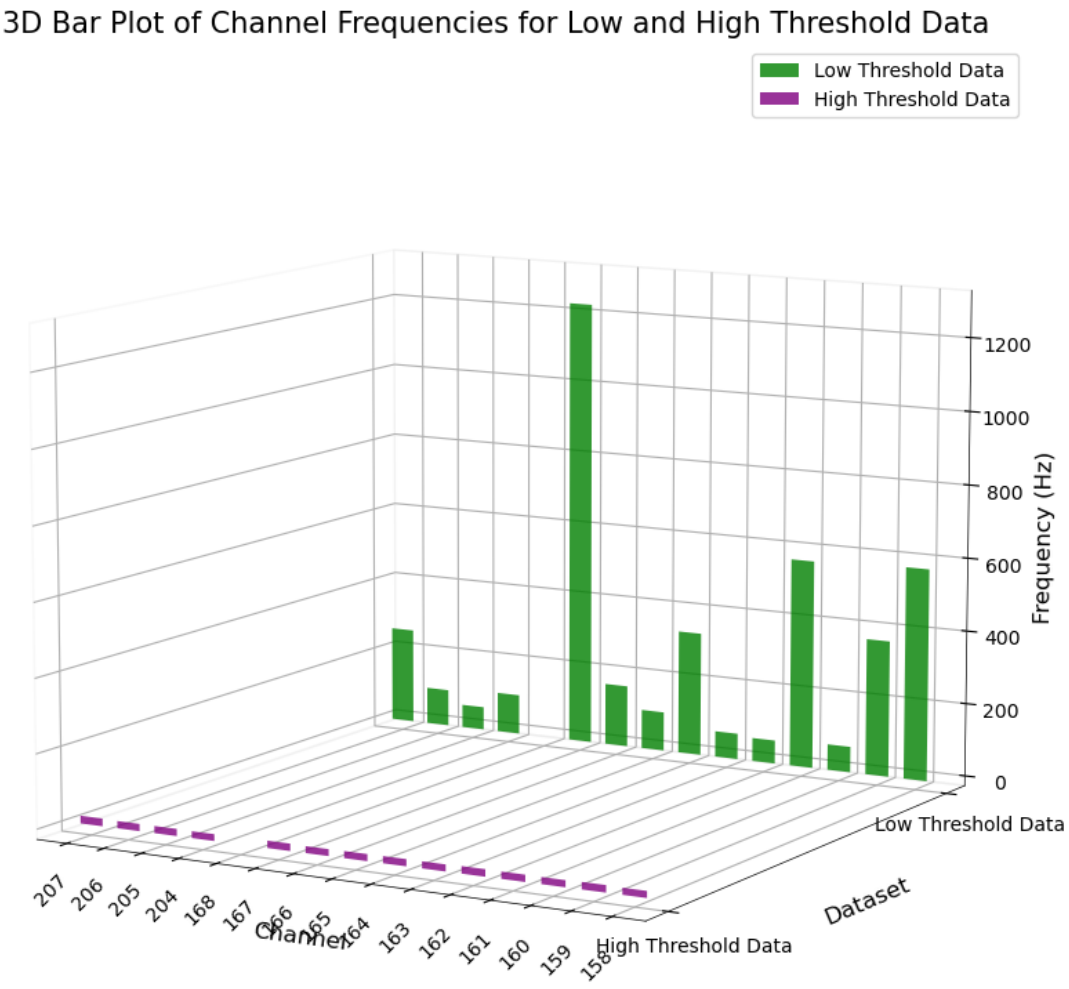


Figure 4.5: 3D Bar Plot Comparing Channel Frequencies for Low and High Threshold Data Configurations.

The analysis of different cable setups shows that each one affects noise differently. Shielded cables reduce low-frequency noise but may still let high-frequency noise through due to outside interference. Shorter cables reduce high-frequency noise but mainly pick up lower frequencies. Unshielded setups with low thresholds detect a wider range of frequencies, including higher ones. On the other hand, high thresholds filter out most noise but may also miss subtle signals. These findings highlight the importance of choosing the right cables and thresholds to reduce noise effectively in the CMS Drift Tubes.

4.4 Multifaceted Time-Series Analysis

The time-series analysis was investigated common mode noise's temporal and spectral characteristics across four datasets. Advanced signal processing techniques were employed, including the Fast Fourier Transform (FFT) to identify dominant noise frequencies, the Short-Time Fourier Transform (STFT) to analyze frequency evolution, and autocorrelation to assess periodicity and self-similarity. Inter-arrival time distributions were also examined to understand the temporal clustering of noise events. Together, these methods provided a comprehensive understanding of the noise processes, allowing us to identify potential interference sources and their impact on the CMS Drift Tubes.

Each dataset was loaded into a data frame, with channels selected based on their relevance to the noise study (channels 158-169 and 204-207 in two datasets and 223-239 in the others). The 'absolute timing' data was used with a consistent time resolution of 50 microseconds, organizing the timestamps into time series representing event counts in equally spaced intervals. This data was then analyzed using FFT to identify key frequency components in the noise, and the results were plotted to detect critical peaks indicating external interference.

Additional techniques were applied to enhance the analysis. Inter-arrival times between events were calculated to reveal timing patterns and clustering. STFT was used to capture the frequency content over time, generating spectrograms that showed how the noise profile varied during the recordings. Autocorrelation was used to examine the temporal structure and randomness of the noise, with plots visualizing recurring patterns or behaviors.

4.4.1 Dataset1:Shielded Analysis

4.4.1.1 FFT Analysis for Shielded Dataset

The frequency spectrum (FFT) for channels 224 through 227 shows distinct characteristics. Channel 224 displays a wider frequency range with higher magnitudes at low frequencies, which supports noise observation. Noise often consists of dominant low-frequency components. Channels 225 and 226 have a consistent signal with less noise, distributing power evenly across frequencies. However, Channel 227 has the least power in the lower frequencies, suggesting it is the least noisy of the four channels. Channel 224 has more low-frequency components, which correlate with irregular event patterns in the inter-arrival times, indicating that this channel is noisier. Figure 4.6 shows these changes.

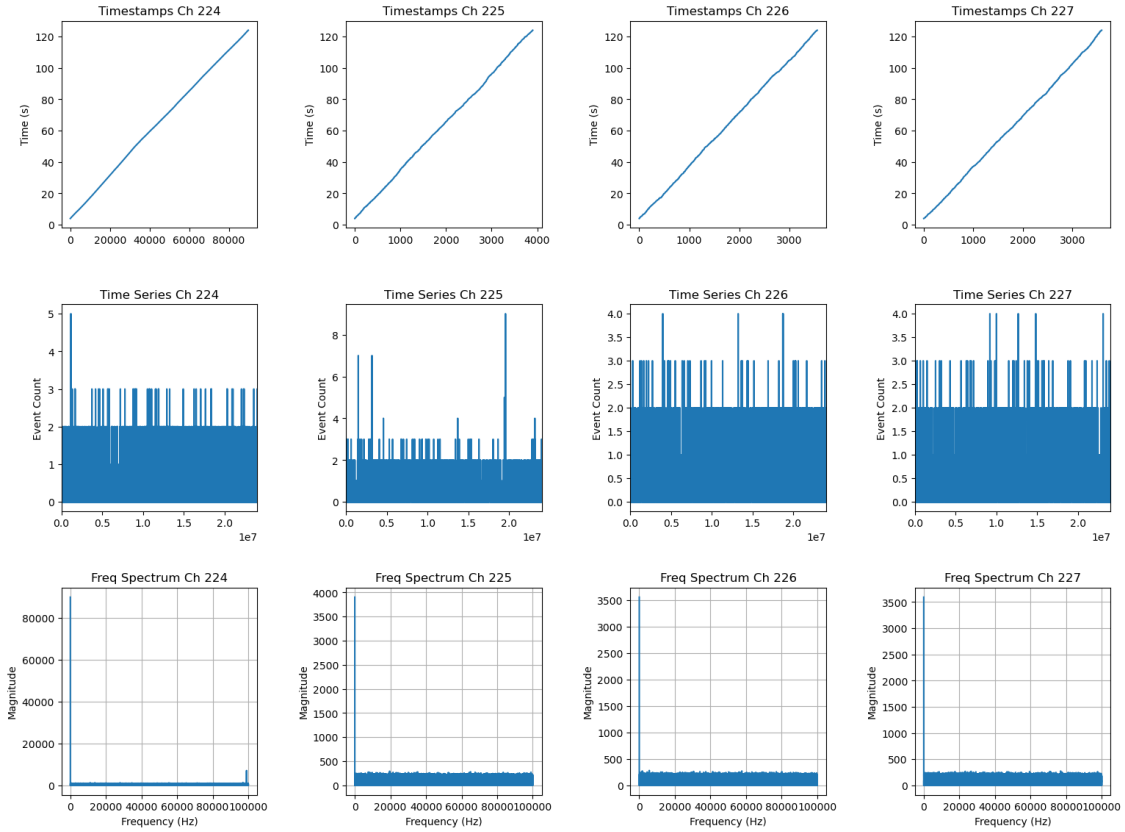


Figure 4.6: Time Series and Frequency Spectrum Analysis for Channels 224 to 227 (shielded).

Figure 4.7, The FFT method analysis shows that Channel 228 has a wider and stronger frequency spectrum, indicating the presence of significant noise. This broad spectrum is characteristic of a noisier signal, where the energy is spread across a wider range of frequencies. In contrast, Channels 229 and 231 have spikes in their FFT plots, but they are more isolated and focused than Channel 228. These spikes show that these channels have specific high-energy frequency components, which could represent periodic signals or recurring artifacts. On the other hand, Channel 230 has a more even frequency distribution with no significant spikes, suggesting a stable and consistent signal without dominant periodic components.

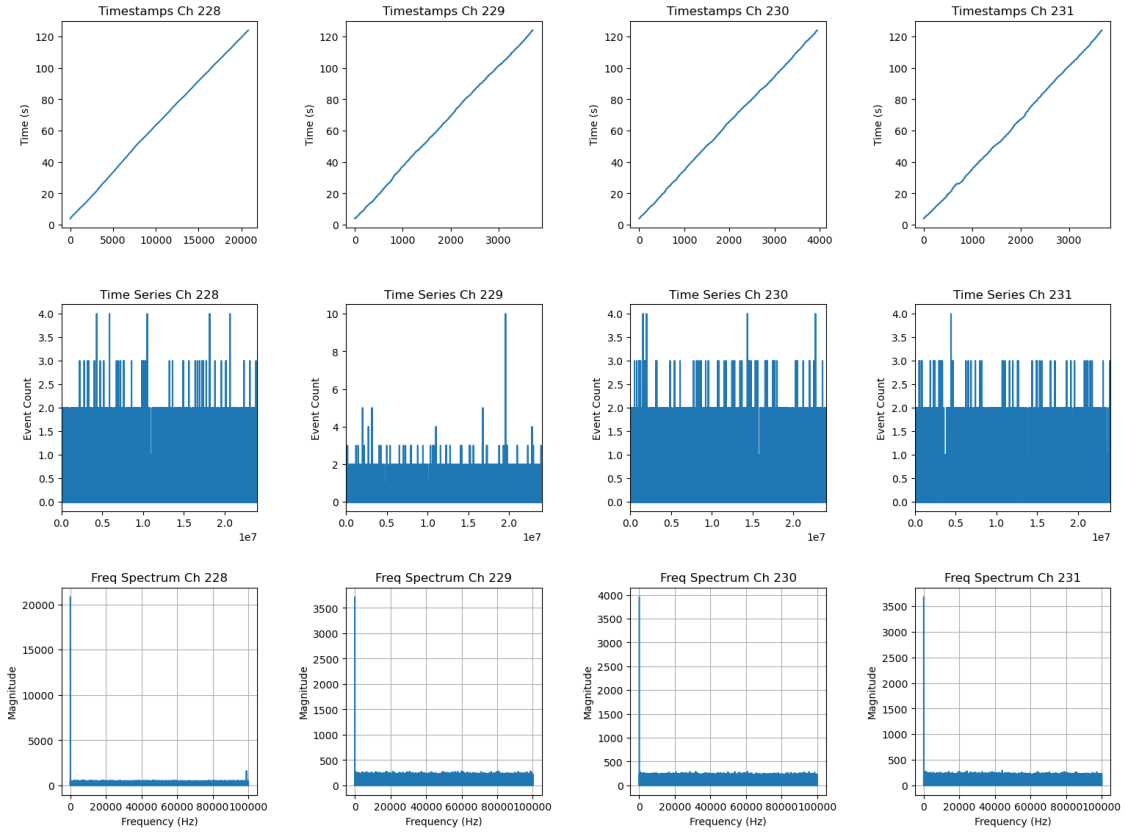


Figure 4.7: Time Series and Frequency Spectrum Analysis for Channels 228 to 231 (shielded).

In the FFT analysis, Channel 232 appears to be the most noisy, with a broader frequency spectrum indicating significant noise. The energy is spread across a wider range of frequencies, characteristic of a noisier signal. Channel 233 also shows some noise but is slightly more concentrated, suggesting less energy dispersion than Channel 232. Channel 234 has a spike with high magnitude at a specific frequency, indicating a dominant periodic component or artifact in the signal. Channel 235, while generally stable, shows a narrower frequency distribution with some noise present, but less than in Channel 233. Figure 4.8 shows these changes.

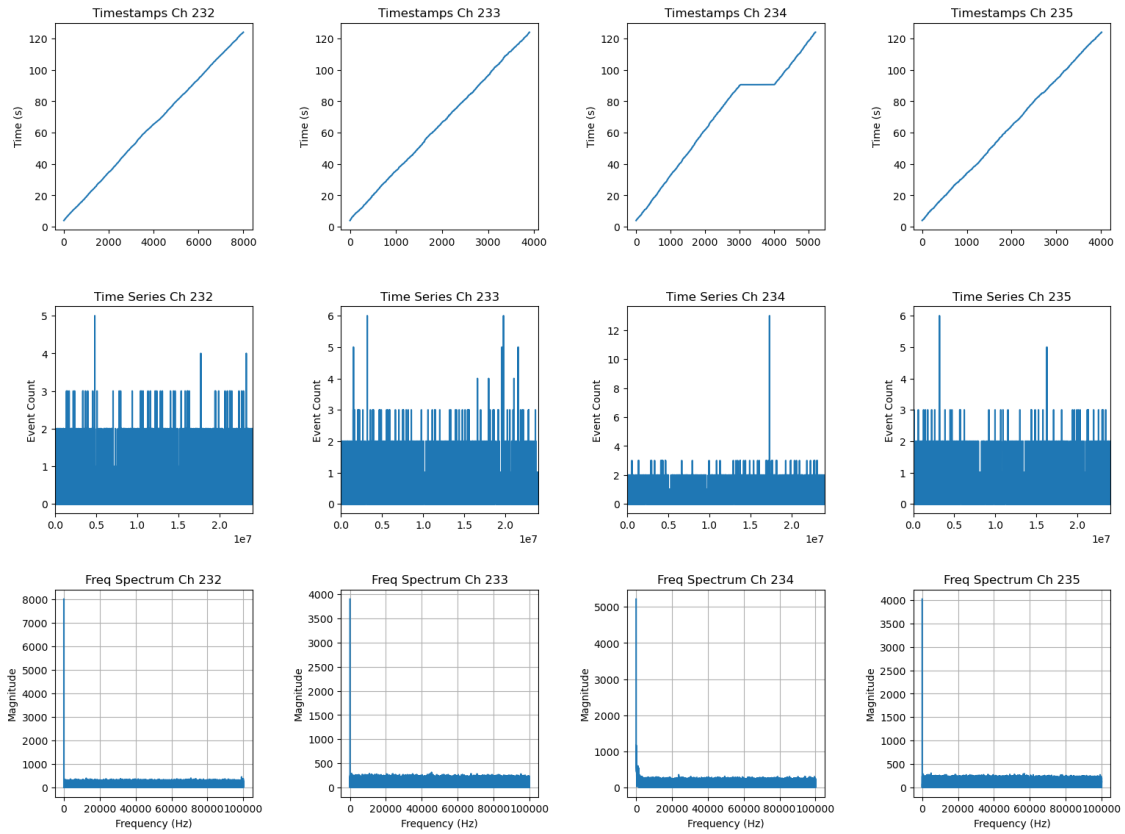


Figure 4.8: Time Series and Frequency Spectrum Analysis for Channels 232 to 235 (shielded).

In the FFT analysis, Channel 237 appears to be the most noisy, with a broader and more pronounced frequency spectrum that indicates significant low-frequency noise. The energy is dispersed across a wide range of frequencies, a characteristic of noisier signals. Channels 239,238 show a notable spike with high magnitude at a specific frequency, suggesting the presence of a strong periodic component or artifact in the signal. Channel 236 also displays some noise, with energy slightly spread across lower frequencies but less pronounced than in Channel 237. Figure 4.9 shows these changes.

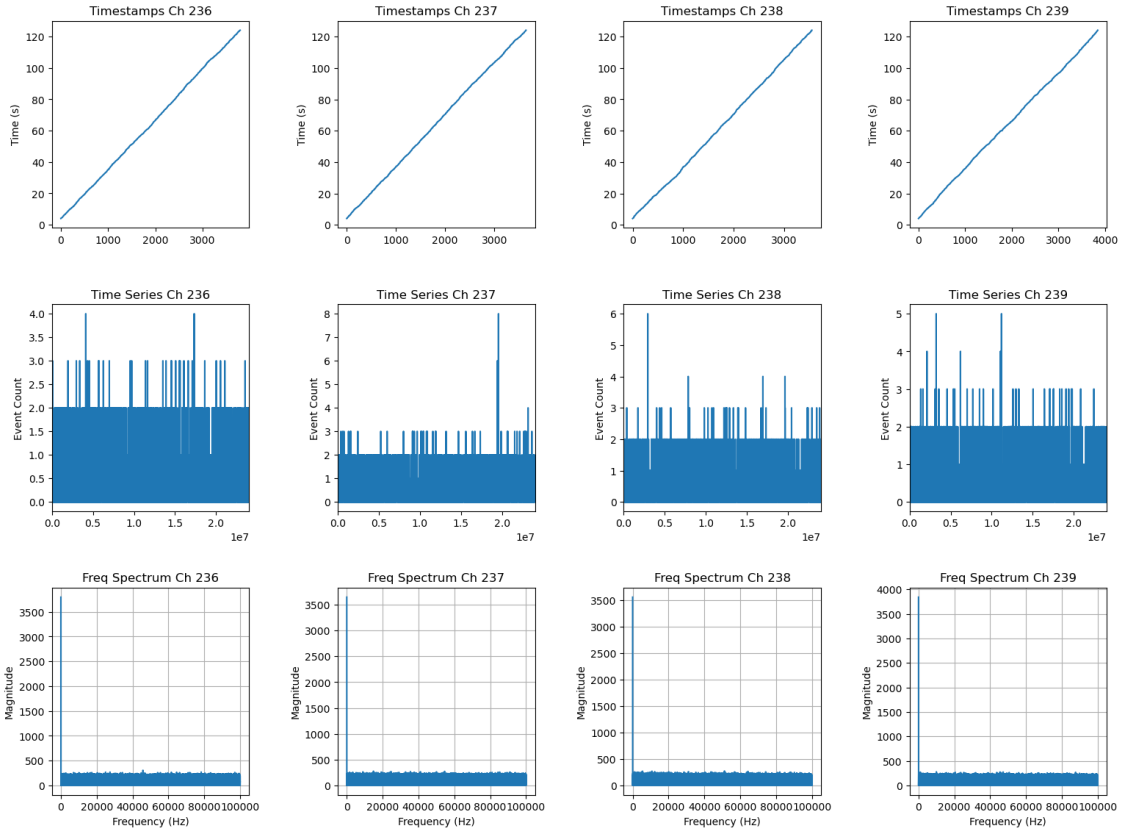


Figure 4.9: Time Series and Frequency Spectrum Analysis for Channels 236 to 239 (shielded).

The analysis of Channels 224 to 239 in the shielded dataset using FFT shows a strong dominance of low frequencies, indicating stable processes with minimal high-frequency noise. However, Channels 224, 228, and 234 show higher noise or irregularities.

4.4.1.2 Inter-Arrival Times Analysis for Shielded Dataset

In figure **??**, The inter-arrival time histograms for channels 224 to 227 exhibit a typical exponential decay, indicative of a Poisson process common in stochastic signal processes. Channel 224 shows higher density at very low time intervals, suggesting that it is more prone to noise, potentially due to an increased rate of spurious events or artifacts. Channels 225 and 226 follow a similar pattern but with slightly lower densities, indicating less frequent noise. Channel 227 presents the cleanest pattern, with the inter-arrival times more spread out, implying a more stable and less noisy signal. A sharp peak at lower intervals in Channel 224 suggests it may be more susceptible to bursty noise, where many events are clustered closely in time.

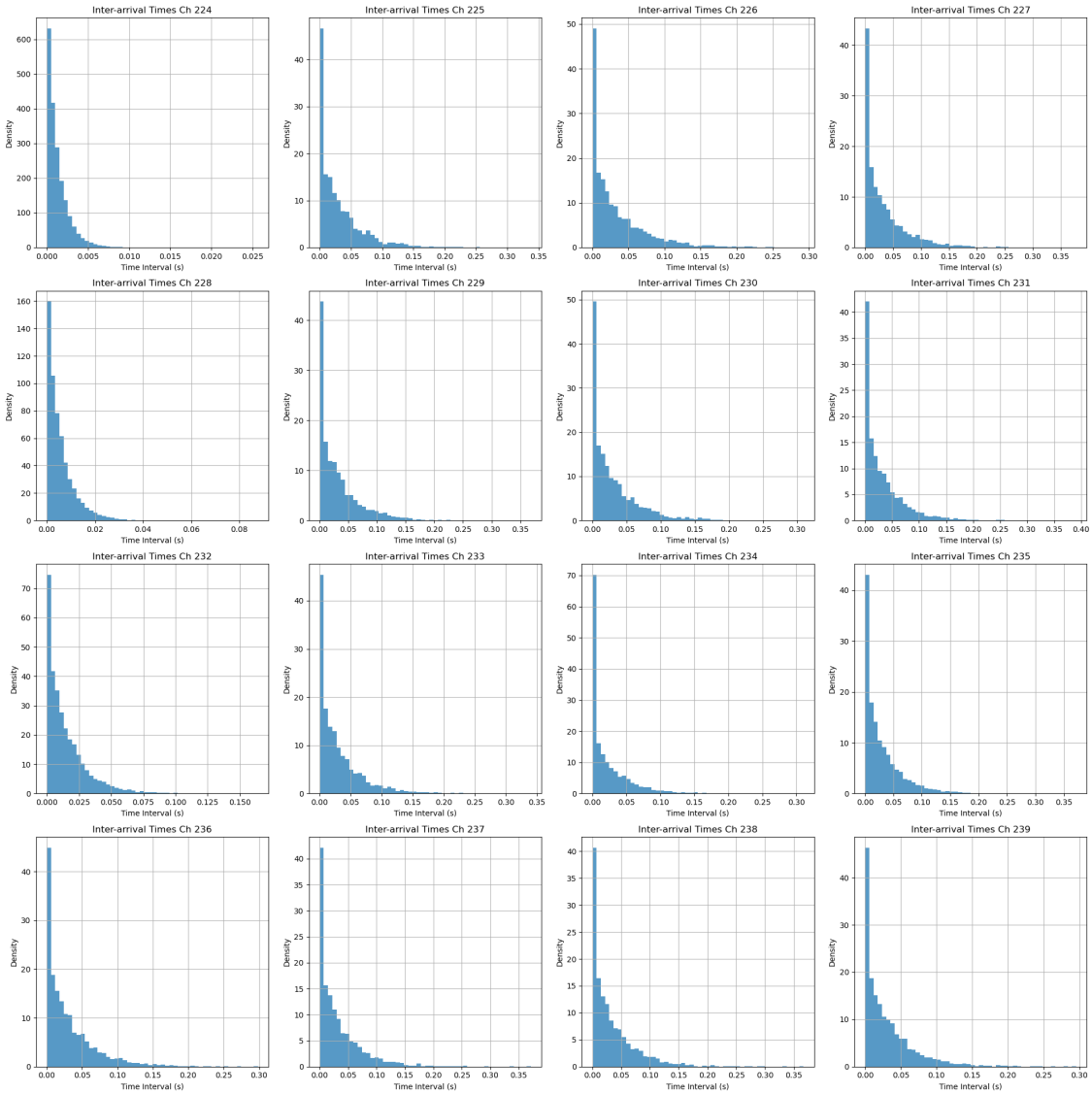


Figure 4.10: Inter-arrival Times Analysis for Shielded Dataset.

The inter-arrival time histograms show a typical exponential decay for all channels, but Channel 228's histogram has a higher density at lower intervals, reflecting its noisiness. Channels 229 and 231 display more spread-out inter-arrival times, corresponding to their spiky nature in the frequency domain.

The inter-arrival time histograms in Channel 232 show a higher density at shorter intervals, indicating more frequent noise events. At the same time, Channel 234 exhibits more spread-out intervals due to its dominant periodic component.

The inter-arrival time histograms reveal that Channel 237 has a higher density at shorter intervals. Channels 239 and 238 show more spread-out intervals, corresponding to the periodic components observed earlier. Channel 236 exhibits a similar pattern to 237 but with a little less density than 237, indicating fewer noise-induced events.

4.4.1.3 Auto Correlation Analysis for Shielded Dataset

On Channel 224 figure**??**, the autocorrelation function is broader and more diverse, suggesting the presence of periodic noise or irregularities in the signal. Channels 225 and 226 have sharper peaks at zero lag, indicating more consistent signals. Channel 227 has the sharpest and most defined peak, suggesting the least noise or periodic disturbances. Channels 228 and 232 show broad peaks in their autocorrelation plots, indicating periodic noise or signal irregularities. Channel 234 has a sharp peak, suggesting a more consistent signal with occasional external noise.

The autocorrelation plot for Channel 228 exhibits a broader peak, indicating periodic noise or irregularities in the signal. This aligns with the noisy characteristics observed in the FFT and STFT analyses. Channels 229 and 231, however, show sharper peaks with additional secondary peaks, suggesting the presence of recurring patterns or spikes within the signal, correlating with the high-magnitude spikes seen in their FFT and STFT analyses. Channel 230 has a sharp autocorrelation peak with minimal side lobes, indicating a more consistent and less noisy signal.

The autocorrelation plot for Channel 232 displays a broader peak, indicating periodic noise or irregularities in the signal. This is consistent with the noisiness observed in the FFT and STFT analyses. Channel 233 also shows a broader peak but is slightly sharper than Channel 232, suggesting some regularity despite the noise. Channel 234 has a sharp peak in its autocorrelation plot, corresponding to the strong periodic component identified in the FFT and STFT analyses. Channel 235 shows a sharp autocorrelation peak.

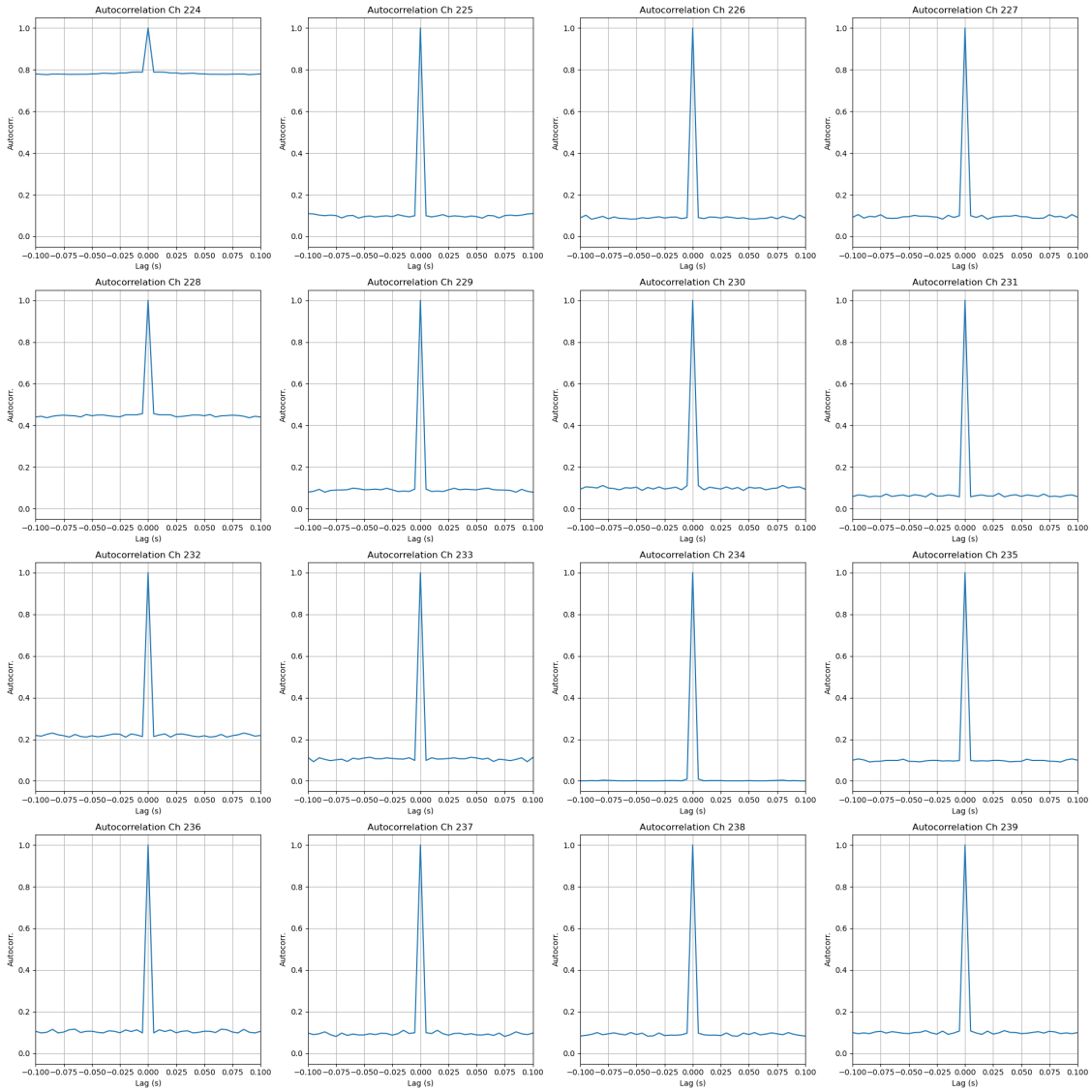


Figure 4.11: Auto-Correlation for Shielded Analysis.

Moreover, Channel 237 shows a broader autocorrelation peak, which suggests periodic noise or irregularities, consistent with its noisiness observed in the FFT and STFT analyses. Channels 239 and 238 display sharp peaks in their autocorrelation plots, indicating strong periodic components aligning with the spikes observed in the FFT and STFT.

4.4.1.4 Short-Time Fourier Transform (STFT) for Shielded Dataset

The spectrograms for these channels show how the frequencies change over time. Channel 224 displays higher intensity across the spectrogram, particularly in the lower frequencies, which matches the findings of the previous FFT analysis and indicates that it is noisy. The sudden spike in energy observed in Channel 234 suggests a sudden burst of energy, possibly caused by external interference. Channels 225 to 227 exhibit more consistent spectrograms with less variation over time, indicating more stable signals.

The STFT spectrograms for these channels further support the FFT findings. Channel 228 shows a higher intensity across the time-frequency domain, particularly at lower frequencies, which aligns with the observation of a noisier signal in the FFT analysis. Channels 229 and 230 display more uniform and less intense spectrograms. Channels 229, 230, and 231, prominent vertical lines in these spectrograms correspond to specific frequencies with high magnitudes, as noted in the FFT analysis. These vertical lines suggest that these channels have strong, consistent frequency components, which could be due to periodic signals or artifacts.

The STFT spectrograms support the findings from the FFT analysis. Channel 232 shows high intensity across time and frequency, especially in the lower frequencies, confirming that it's noisy. Channel 234's spectrogram reveals a prominent vertical line corresponding to the high-magnitude spike in the FFT, indicating a strong and consistent frequency component over time. Channel 233 and 235, with a more uniform and less intense spectrogram with less noise.

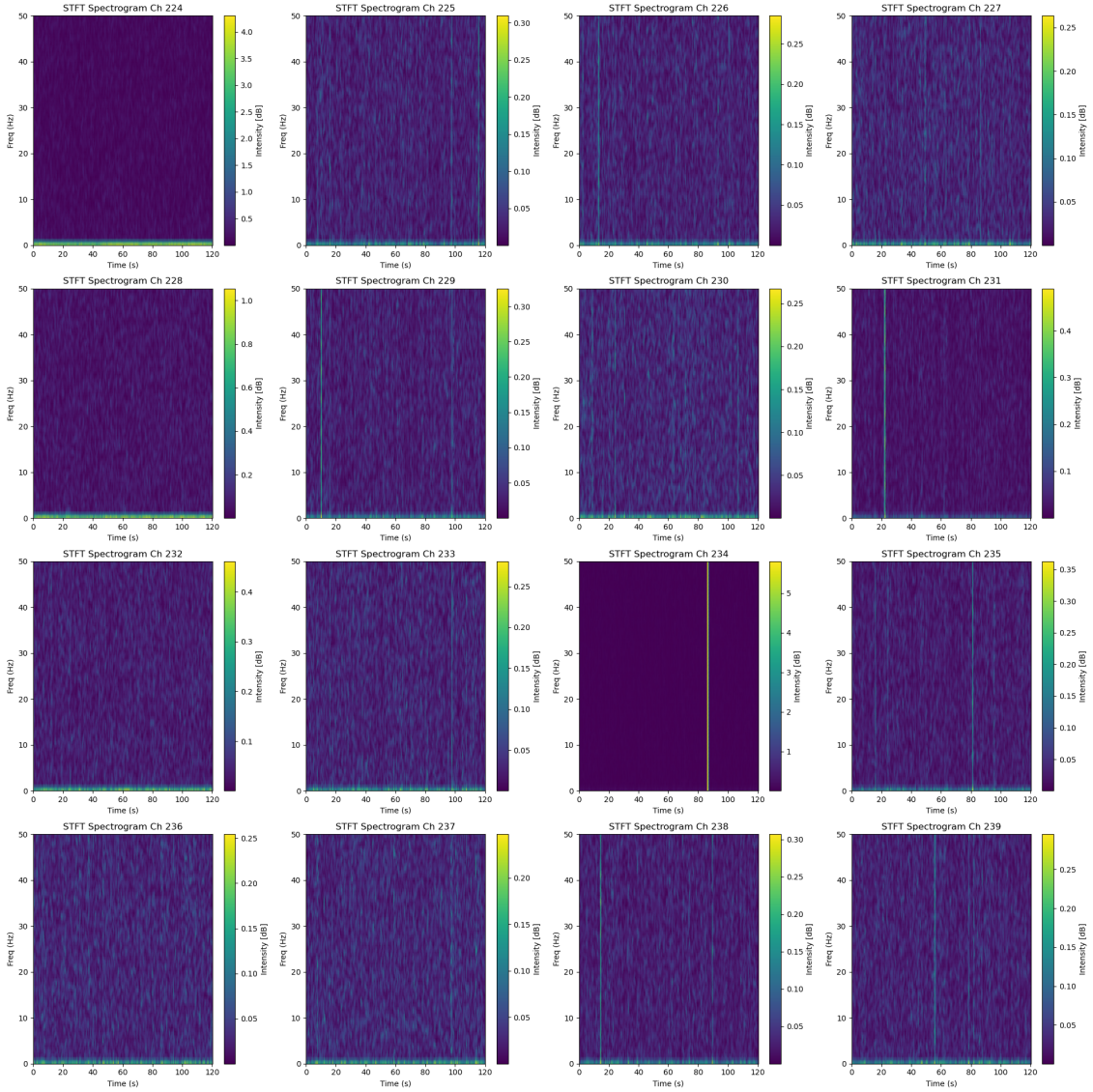


Figure 4.12: STFT analysis for the shielded analysis.

The STFT spectrograms support the findings from the FFT analysis. Channel 237 shows high intensity across the time-frequency domain, particularly in the lower frequencies. The Channels 239 and 238 spectrograms reveal prominent vertical lines corresponding to the high-magnitude spikes seen in the FFT, suggesting strong and consistent frequency components over time. Channel 236 has a moderate intensity in its spectrogram, indicating some noise, but this is less than Channel 237. The spectrogram of Channels is more stable overall, with less variation. Figure 4.12 shows these analyses.

4.4.1.5 Conclusion and Correlation of Shielded Dataset With Spectrum Analyzer Findings

Based on our analysis, Channels 224, 228, 232, 234, and 235 have been identified as the noisiest. Channel 224 is the most affected, showing broad frequency components, high-intensity spectrograms, and irregular autocorrelation. Similarly, Channel 228 exhibits high noise, especially in the low-frequency range. Following this, Channel 232 displays broad autocorrelation peaks and dispersed frequency energy, confirming its classification as noisy. Channel 234, while generally less noisy, exhibits a distinct spike in the frequency domain, likely due to external interference, suggesting a strong but isolated artifact. Channel 235 also shows notable noise, though less than other Channels. The analyses show low-frequency noise, periodic components, and external interference are the main factors affecting signal stability across these channels.

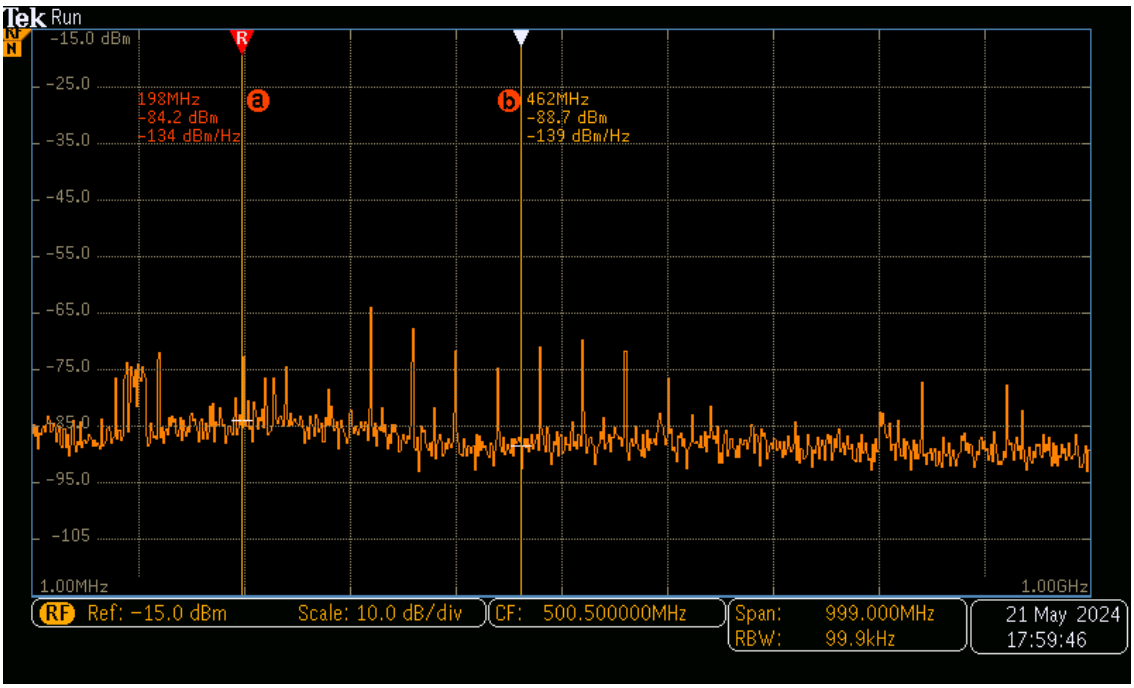


Figure 4.13: spectrom analyzer for the Shielded Analysis.

The spikes observed at 198 MHz and 462 MHz in the spectrogram analyzer closely match the periodic components identified in the FFT analyses of Channels 234, 239, and 238. These channels displayed similar spikes in their frequency patterns, indicating that they could be particularly affected by outside noise or interference at these specific frequencies. This suggests a strong correlation between the spikes in the spectrogram and the noise detected in these chan-

nels, pointing to external electromagnetic interference that the shielding may not fully mitigate.

The spectrogram shows a high background noise level in the lower frequencies, which matches Channels 224, 228, and 232 findings. These channels all had significant low-frequency noise issues. The consistent interference across multiple channels suggests that a common external noise source affects the overall signal quality. The spectrogram analysis further confirms this.

The experiment's use of shielded cables or equipment has led to spikes and elevated noise levels in the spectrogram and channel data. These issues indicate that the current shielding may be ineffective, especially at 198 and 462 MHz frequencies. The data shows that certain frequencies are more likely to pass through the shielding due to possible issues with grounding or design flaws. This means we need better shielding to protect the experiment from outside interference and improve the signal reliability.

4.4.2 Dataset2:short cable anlysis

4.4.2.1 FFT Analysis for Short Cable Dataset

The FFT analysis of Channels 224 to 227, figure 4.14 reveals vital insights into the frequency content of the event data. The study using FFT shows that Channel 224 is very noisy. The frequency plot has a high magnitude, indicating much low-frequency noise. The broad spectrum suggests that the signal in Channel 224 is heavily influenced by noise and unstable. Channels 225, 226, and 227 have less noise than Channel 224. Channel 225 has slightly higher noise levels than Channels 226 and 227. This is shown by more noticeable but still moderate noise in the frequency domain.

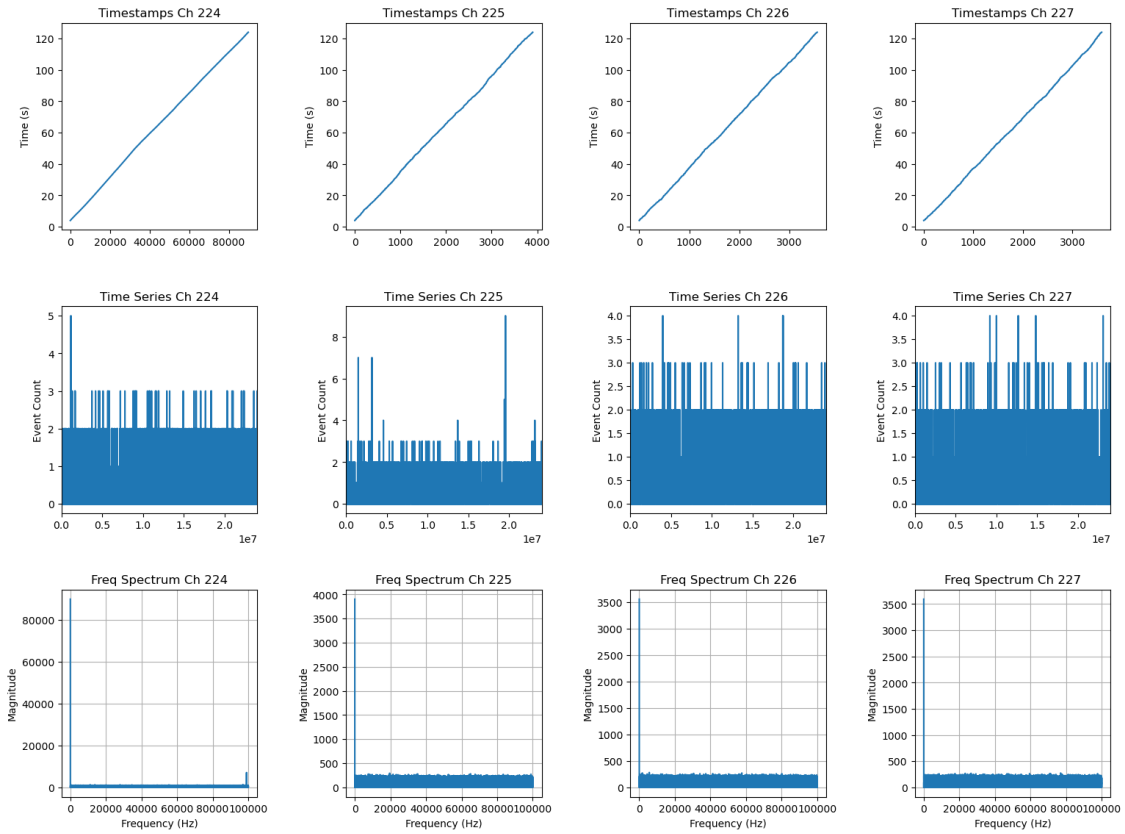


Figure 4.14: Time Series and Frequency Spectrum Analysis for Channels 224 to 227(Short Cable).

The analysis of Channels 228 to 231, Figure 4.15 using FFT shows that Channel 229 is the noisiest among the four channels. This is evidenced by a prominent spike in the frequency spectrum, indicating a strong periodic component or interference that significantly impacts the stability of the signal. Channel 229 has a high noise level with a dense frequency distribution but is slightly less pronounced than Channel 228. Channels 230 and 231 have lower noise levels, with Channel 231 being the least noisy of the four. The noise in these channels has a more stable frequency distribution, which indicates less interference or periodic noise.

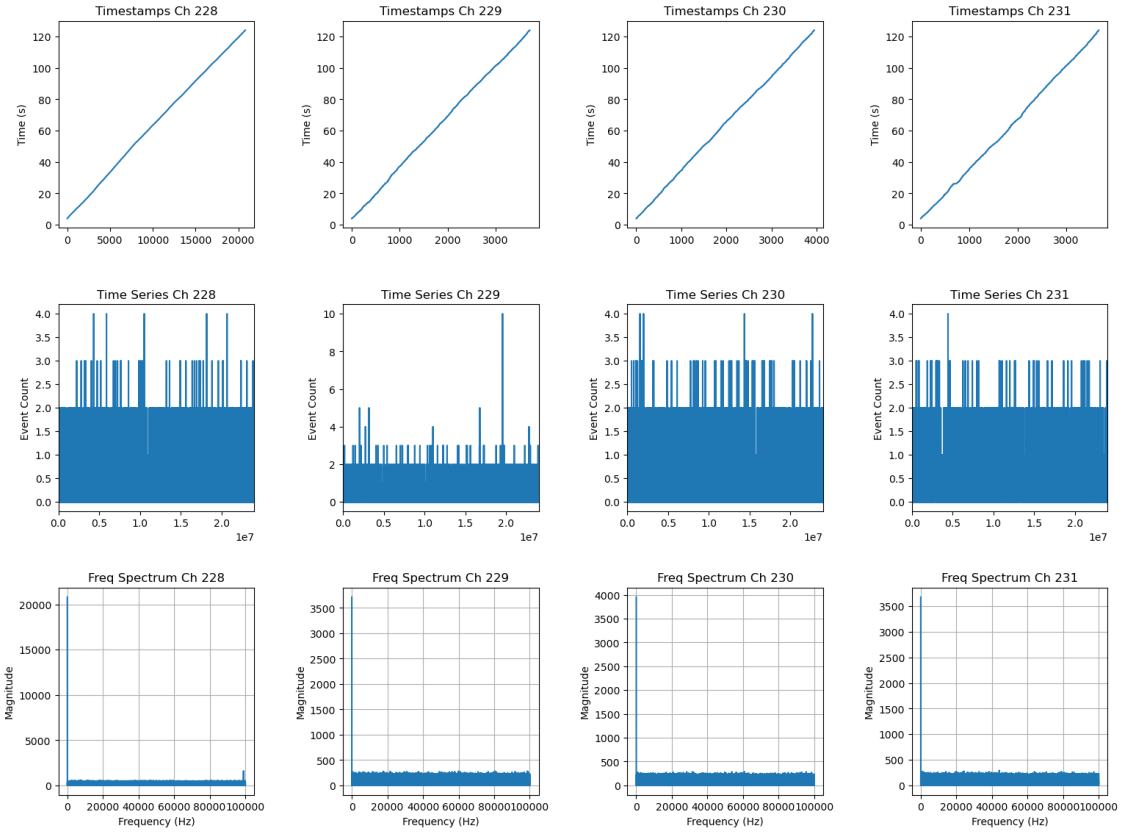


Figure 4.15: Time Series and Frequency Spectrum Analysis for Channels 228 to 231(Short Cable).

Channels 232 to 235 analysis using FFT consistently shows a strong low-frequency pattern across all channels. This suggests the presence of steady, slowly changing processes. Channel 232 has slightly higher low-frequency values; a more robust signal is indicated. Channels 233 and 235 also show significant low-frequency content with minimal higher-frequency noise, indicating stable processes with occasional variations. However, Channel 234 displays more variability, with higher peaks in the frequency spectrum, which may indicate either noise or a more complex process influencing the signal. Overall, the dominance of low-frequency content across these channels suggests consistent, regular processes with minimal rapid fluctuations. Figure 4.16 shows these changes.

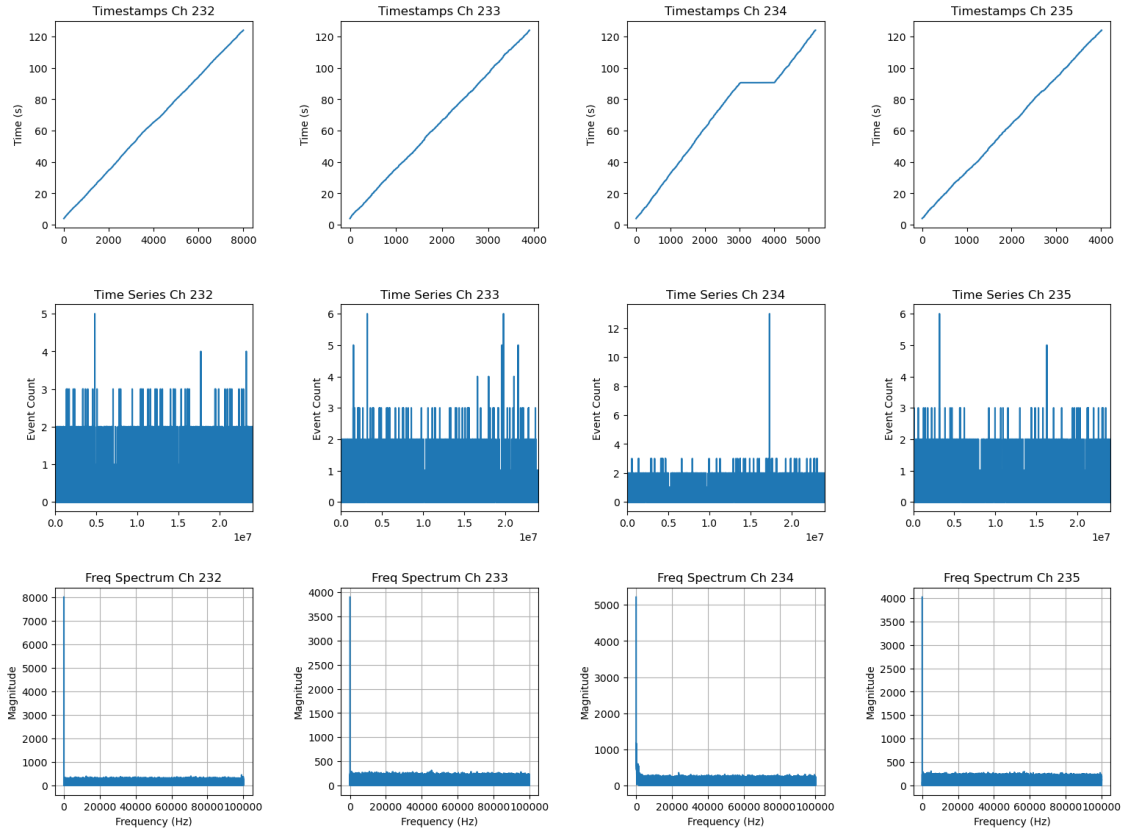


Figure 4.16: Time Series and Frequency Spectrum Analysis for Channels 232 to 235(short cable).

The analysis of Channels 236 to 239 shows that mainly steady and slow changes are happening across all channels, with Channel 236 displaying slightly more pronounced low-frequency patterns. Channels 237 to 239 show similar patterns with minimal high-frequency content, suggesting that consistent, low-frequency signals mainly drive the data. No significant high-frequency peaks across these channels indicate minimal rapid or sudden events. Any variations in the data are likely due to noise or minor fluctuations. Figure 4.17 shows these changes.

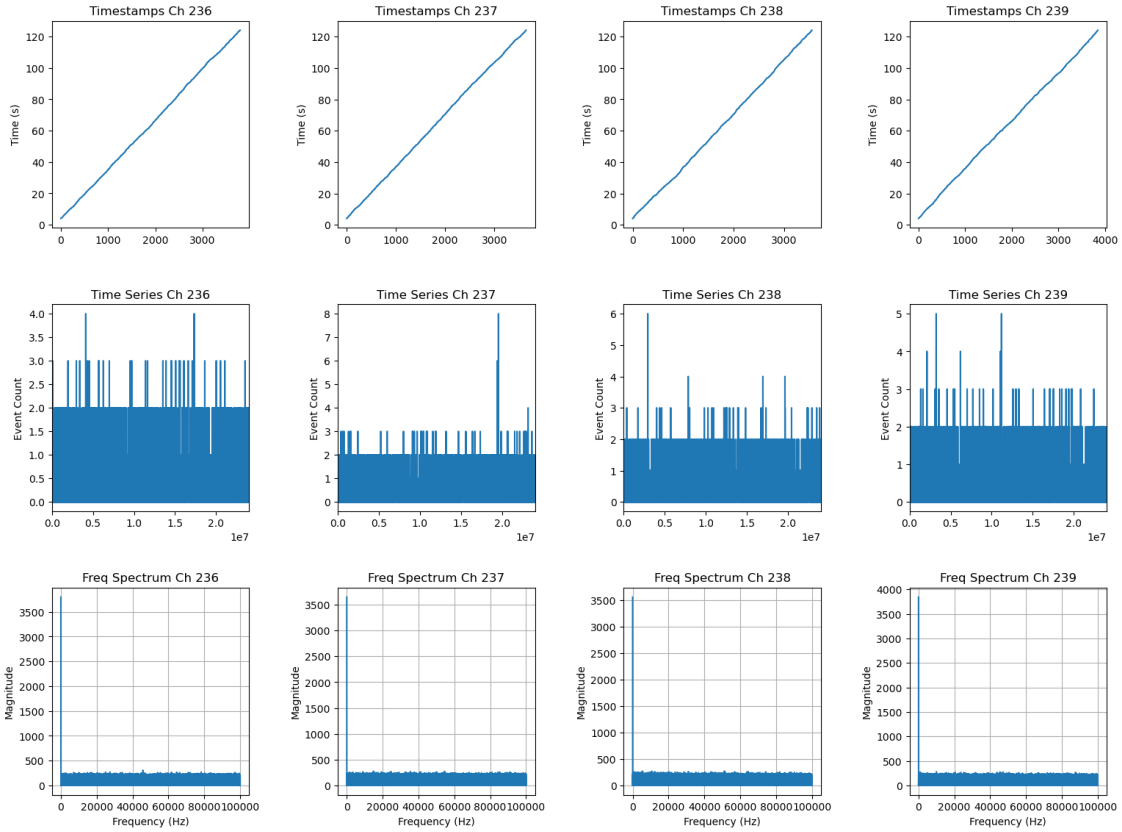


Figure 4.17: Time Series and Frequency Spectrum Analysis for Channels 236 to 239(short cable).

4.4.2.2 Inter-Arrival Times Analysis for Short cable Dataset

This analysis aligns with the observations from the inter-arrival time histograms,figure**??**, where Channel 224 shows a higher density at shorter intervals, indicative of frequent noise-induced events. Meanwhile, Channels 225, 226, and 227 have more evenly distributed inter-arrival times, corresponding to their less noisy signals as reflected in their auto-correlation plots.

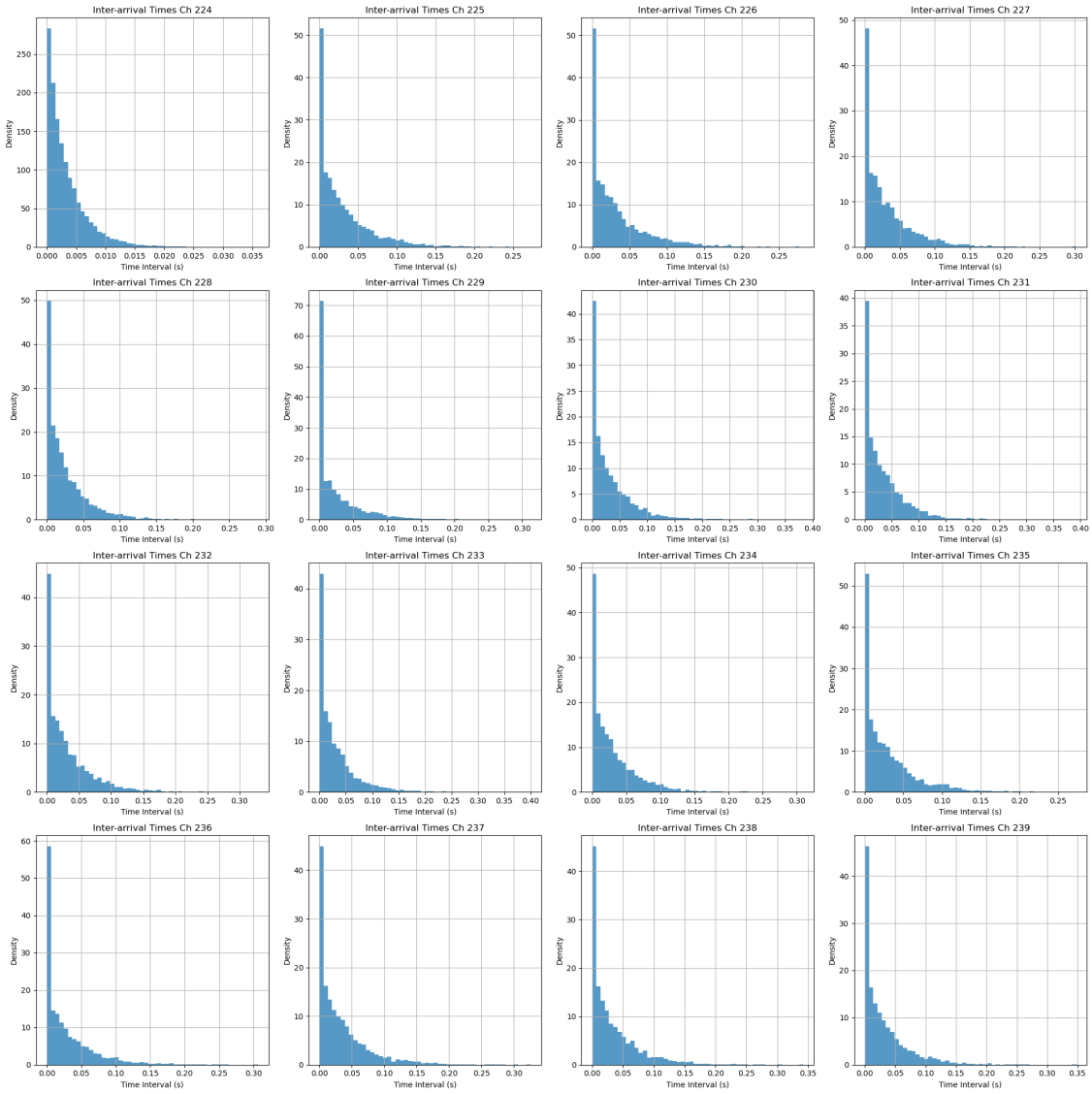


Figure 4.18: Inter-arrival Times Analysis for Shielded Dataset.

The inter-arrival time histograms align with these observations. Channel 229 shows irregular steps and a higher density at shorter intervals, supporting the conclusion that it is the noisiest channel due to frequent noise-induced events. Channel 228 also shows some irregularities but to a lesser extent. Channels 230 and 231 display more evenly distributed inter-arrival times, indicating fewer disruptions and a more stable signal.

The inter-arrival time histograms show that Channel 234 has a higher density at shorter intervals, further supporting its status as the noisiest signal among the four channels. Channels 233

and 232 display more evenly spread inter-arrival times, indicating fewer noise-induced events.

The inter-arrival time histograms show that Channel 236 has a denser concentration of shorter intervals, indicative of more frequent noise-induced events. Channels 237 and 238 exhibit more evenly distributed inter-arrival times, corresponding to their more regular and stable signals. Channel 239 shows a slight concentration at shorter intervals, consistent with its FFT and STFT findings.

4.4.2.3 Auto correlation Analysis for Short Cable Dataset

The auto-correlation plot, figure 4.19 for Channel 224, shows a significant peak around 0.2, indicating irregularities and substantial noise in the signal. This matches the high noise levels seen in the FFT and STFT analyses. In contrast, Channels 225, 226, and 227 have sharp peaks near zero, reflecting more regular and less noisy signals. This disparity highlights Channel 224's instability compared to the more stable signals in the other channels.

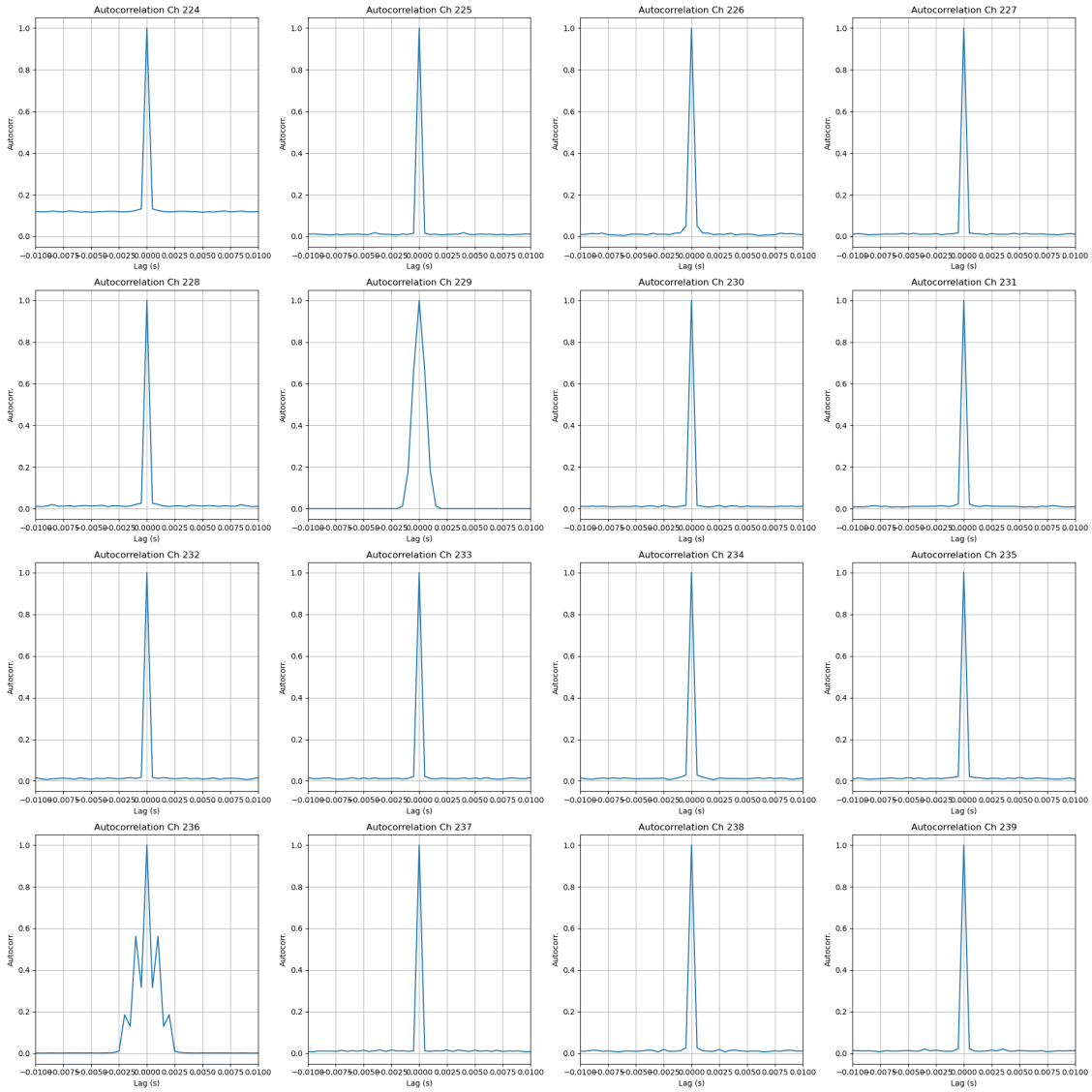


Figure 4.19: Auto-Correlation for Shielded Analysis.

The autocorrelation plot for Channel 229 is distinctly different from the others, showing irregularities and a broader peak, which reflects the strong periodic noise identified in the FFT and STFT analyses. Channel 228 has a higher autocorrelation density than Channels 230 and 231, indicating more noise and less signal regularity. Channel 230 shows a sharper peak, suggesting a more regular and less noisy signal. At the same time, Channel 231 has the sharpest and most defined autocorrelation peak, indicating the cleanest signal among the four channels.

The autocorrelation plot for Channel 234 shows a broad peak, reflecting irregularities and

periodic noise within the signal. Channel 235 also exhibits a broad peak, though less pronounced than Channel 234, indicating some periodic noise. Channels 233 and 232 have sharper peaks in their autocorrelation plots, suggesting more regular signals with less noise.

The autocorrelation plot for Channel 236 is distinctly different, with multiple peaks rather than the sharp single peak seen in the other channels. This indicates a higher level of irregularity and noise within this channel. Channels 237 and 238 have sharp and well-defined peaks, suggesting more regular and less noisy signals. Channel 239, while showing a single peak, shows minor irregularities similar to Channel 236, though less pronounced.

4.4.2.4 STFT Analysis for Short Cable Dataset

The STFT spectrograms support the FFT findings. Channel 224 in figure 4.20 displays intense activity across the time-frequency domain, especially at lower frequencies, indicating that this channel is very noisy. In addition, the spectrograms for Channels 225 and 226 show strong spikes. The spike in 225 is stronger than in 226. Furthermore, 227 shows more noisy patterns, which matches the noise level observed in the FFT analysis.

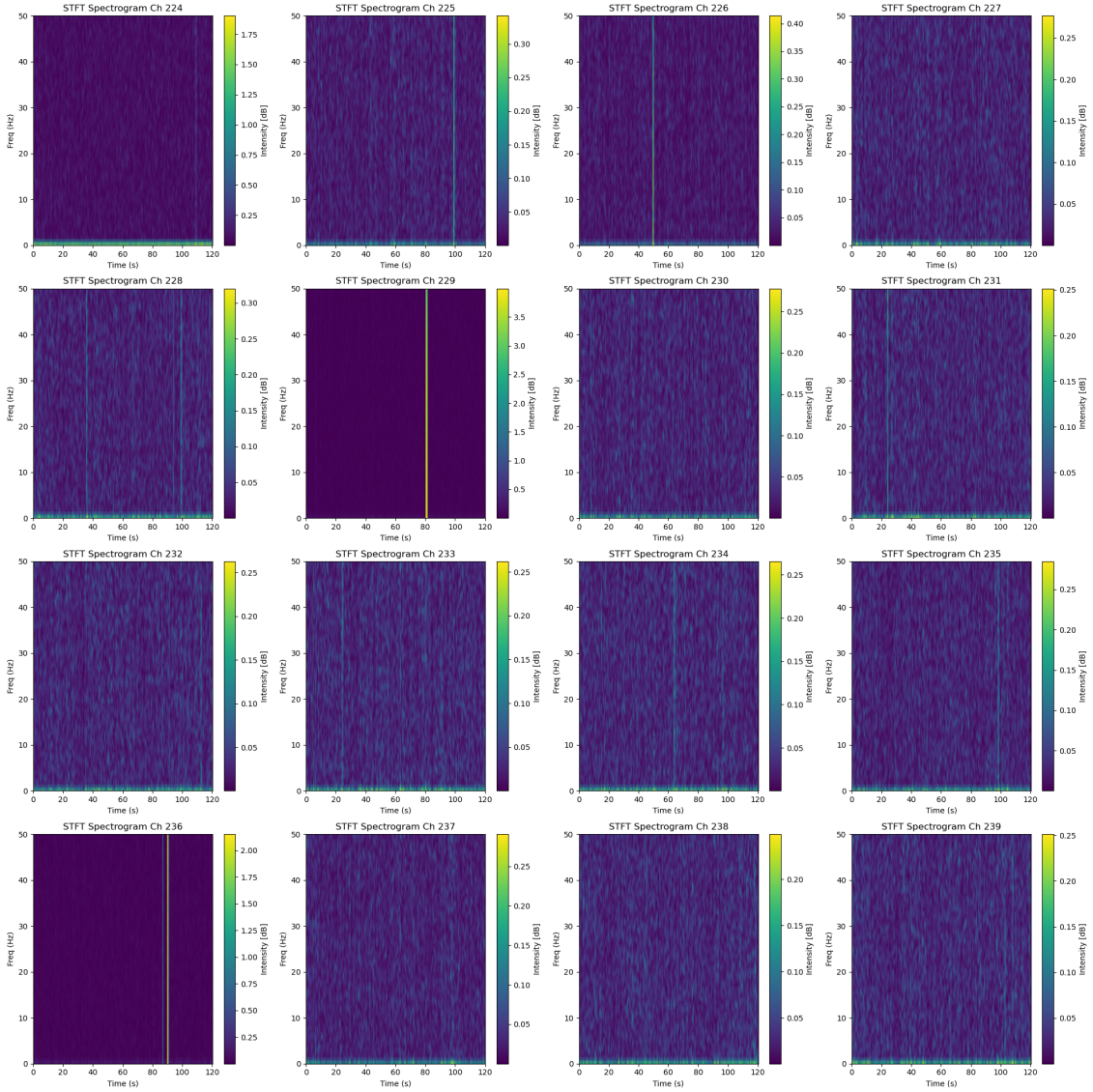


Figure 4.20: STFT analysis for the short cable analysis.

The STFT spectrograms reinforce the findings from the FFT analysis. Channel 229 shows intense activity in specific frequency bands over time, which correlates with the spike observed in the FFT, indicating persistent interference or noise at that frequency. Channel 228 also displays important noise across a wide range of frequencies, though it is less concentrated than in Channel 229. Channels 230 and 231 have more uniform spectrograms with lower intensity, suggesting more stable signals with less noise. The spectrogram for channel 230 shows fewer and less intense frequency components, indicating that this signal is less affected by interfer-

ence.

The STFT spectrograms further support the findings from the FFT analysis. Channel 234 shows high intensity across the time-frequency domain, particularly at lower frequencies, which confirms that this channel is heavily impacted by noise. The spectrogram for Channel 235 also displays notable intensity but to a lesser extent than Channel 234. Channels 233 and 232 have more uniform and less intense spectrograms, indicating more stable signals with reduced noise. The lower intensity in these spectrograms suggests that the signals in Channels 233 and 232 are less affected by external interference.

The STFT spectrograms show that Channel 236 has some irregularities, especially compared to the other channels. It displays a more varied intensity across the time-frequency domain, which aligns with the findings from the FFT analysis. Channels 237 and 238 display more uniform and consistent spectrograms, indicating more stable signals with less interference. Channel 239, while generally stable, shows minor fluctuations in intensity, suggesting the presence of some noise but less than what is observed in Channel 236.

4.4.2.5 Conclusion of Short Cable Dataset

In conclusion, based on the short cable dataset, Channel 224 has been identified as the noisiest. It shows broad frequency components, high-intensity spectrograms, and a broad autocorrelation peak, indicating significant low-frequency noise. Additionally, Channel 229 demonstrates strong interference, especially with a notable spike in the frequency domain and irregular autocorrelation patterns. Channel 238, 237, and 231 are the least noisy and have stable frequencies, consistent spectrograms, and clear autocorrelation peaks. Channel 228 is noisier, with a dense frequency distribution and high autocorrelation density. Channel 236, while not the noisiest, shows a distinct pattern in its autocorrelation, indicating unique irregularities. The analysis shows low-frequency noise and periodic interference significantly impact channel signal stability. Channels 224, 229, and 228 are the most affected.

4.4.3 Dataset3:Unshielded Low Threshold Analysis

4.4.3.1 FFT Analysis For Unshielded Low Threshold

In figure 4.21, After reviewing the FFT plots for Channels 158 to 161, it is clear that Channel 158 exhibits the highest level of noise, with broad energy distribution across the spectrum and a noticeable vertical line at 40 Hz, indicating significant interference. Channel 159 displays

similar characteristics but with even less noise. Channel 161 is noisy noise, too, with a narrower energy distribution and a minimal or absent 40 Hz line. Channel 160 suggests it is the least affected among the four channels.

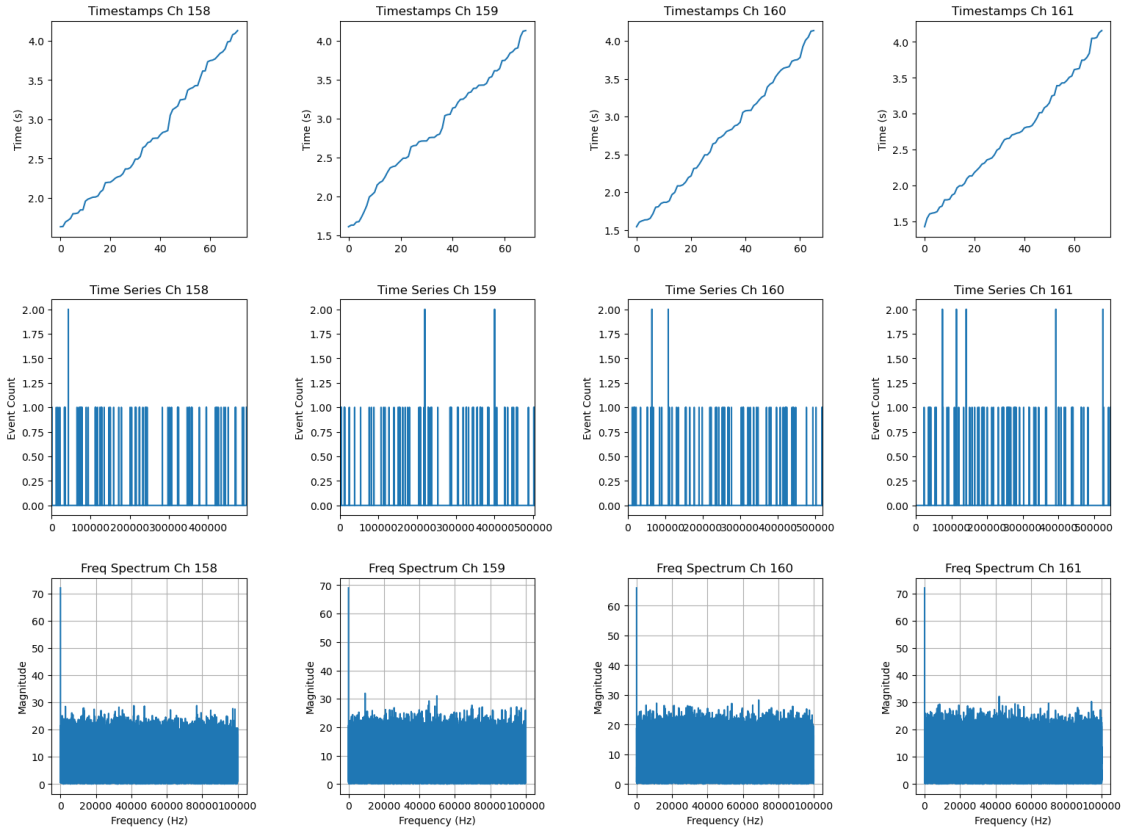


Figure 4.21: Time Series and Frequency Spectrum Analysis for Channels 158 to 161 (Unshielded Low Threshold).

The timestamps maintain a consistent nonlinear trend in the figure 4.22 similar to the previous set. Channel 164 has the most noise, which is evident from the broad and uneven energy distribution across the frequency spectrum. This suggests significant interference or a noisy signal. Channel 165 also exhibits noise, but to a slightly lesser degree than Channel 164, with a somewhat more concentrated energy distribution. On the other hand, Channels 162 and 163 show a much cleaner signal, with less distributed noise and a more defined frequency spectrum. Among these, Channel 162 appears to be the least affected by noise, followed closely by Channel 163.

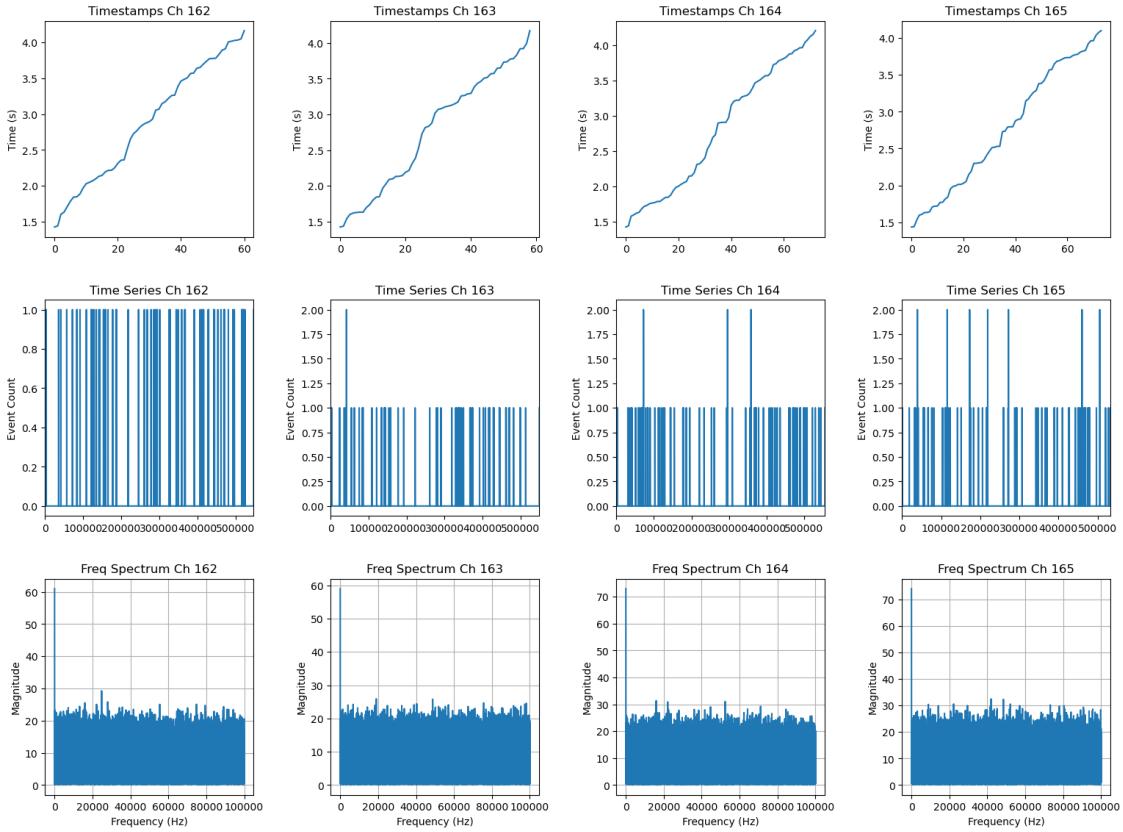


Figure 4.22: Time Series and Frequency Spectrum Analysis for Channels 162 to 165(Unshielded Low Threshold).

The time series histograms for Channels 169 and 167 show sparse events with occasional spikes, and also the FFT plots for Channels 166, 167, and 169 reveal distinct noise characteristics. Channel 167 shows the highest noise level, with a broad and uneven energy distribution across the spectrum, indicating significant interference. This noisy signal is especially pronounced in the lower frequencies. Channel 169 also exhibits considerable noise, though it is slightly less severe than in Channel 167, with a more concentrated energy distribution but still impacted by interference. In contrast, Channel 166 presents a much cleaner signal, with less distributed noise and more defined frequency peaks. This suggests that Channel 166 is the least affected by noise among the three channels. Figure4.23 shows these analyses.

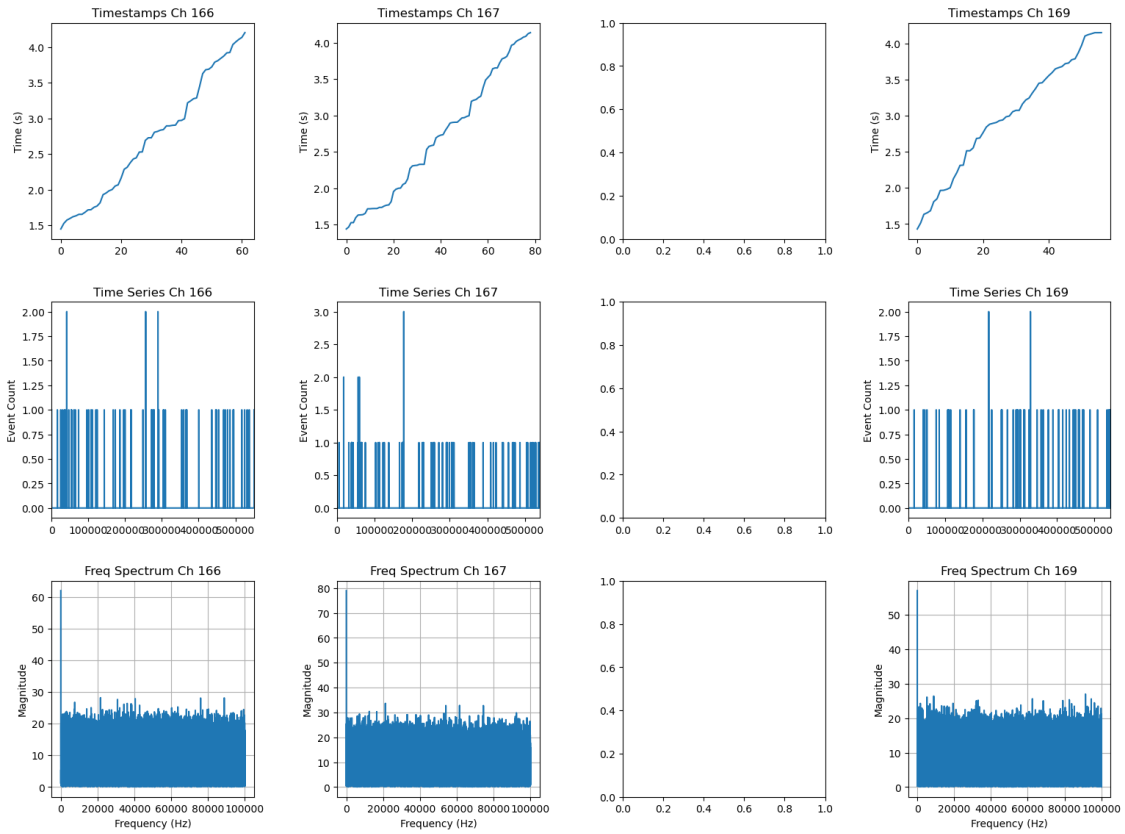


Figure 4.23: Time Series and Frequency Spectrum Analysis for Channels 166 to 169 (Unshielded Low Threshold).

Figure 4.24, Channel 207 shows the highest noise level, as indicated by a broad and uneven energy distribution across the frequency spectrum, offering important interference. Channels 204 and 206 both display lower levels of noise compared to Channel 207, but subtle differences can be observed: Channel 206 has a slightly more concentrated energy distribution, indicating a bit more stability than Channel 204. Channel 205 stands out as the cleanest, with the most defined frequency peaks and the least amount of distributed noise, suggesting it is the least affected by noise among the four channels.

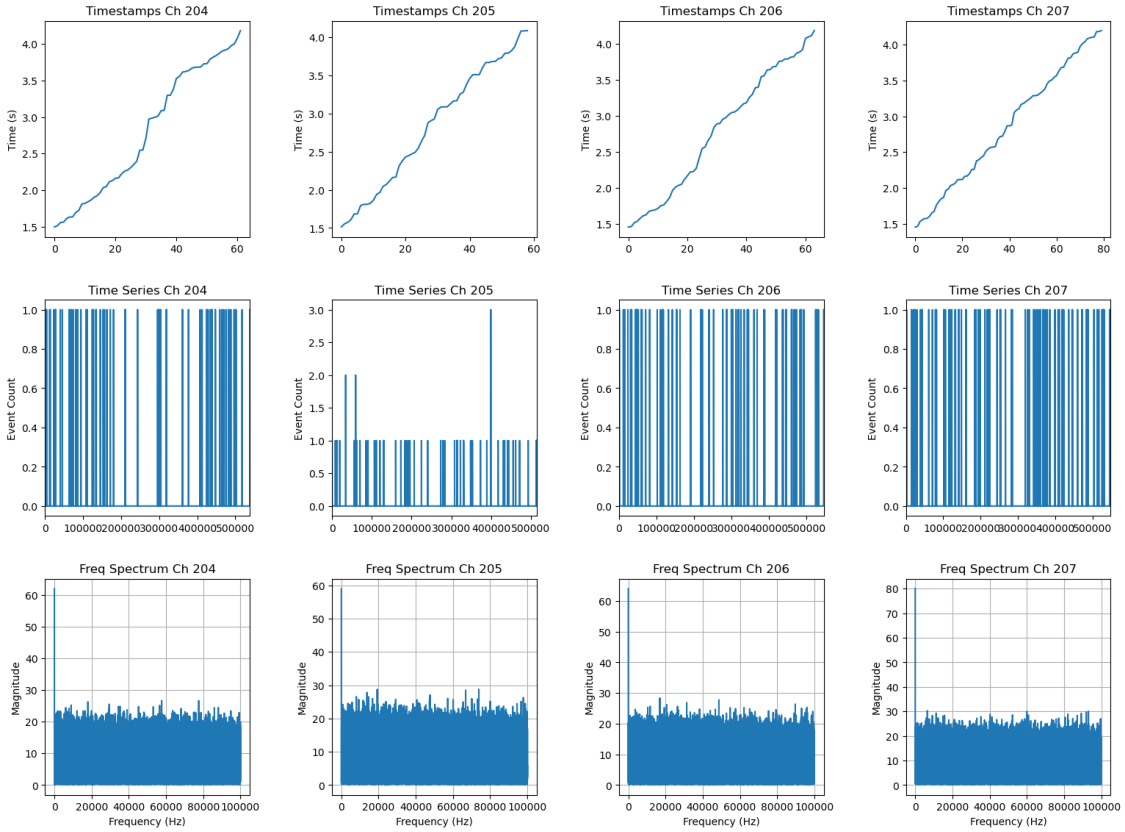


Figure 4.24: Time Series and Frequency Spectrum Analysis for Channels 204 to 207(unshielded low threshold).

4.4.3.2 Inter-arrival Times Analysis for Unshielded Low Threshold Dataset

The inter-arrival time histograms of Channel 158 show a broad and irregular distribution with multiple peaks, indicating highly inconsistent event timing, likely due to significant noise interference. This irregularity suggests that the events are not occurring at regular intervals, which is often a hallmark of a noisy or unstable signal. Channel 161 also displays a somewhat irregular distribution, though it is narrower and slightly more consistent than Channel 158, implying some improvement in event timing but still affected by noise. Channel 159 shows a more focused distribution with fewer peaks, reflecting more regular and consistent intervals between events. Although noise is still present, it is lower than 158 and 161. Channel 160 exhibits the narrowest and most regular distribution of inter-arrival times, with distinct and sharp peaks, indicating the most stable and noise-free signal among the four channels.

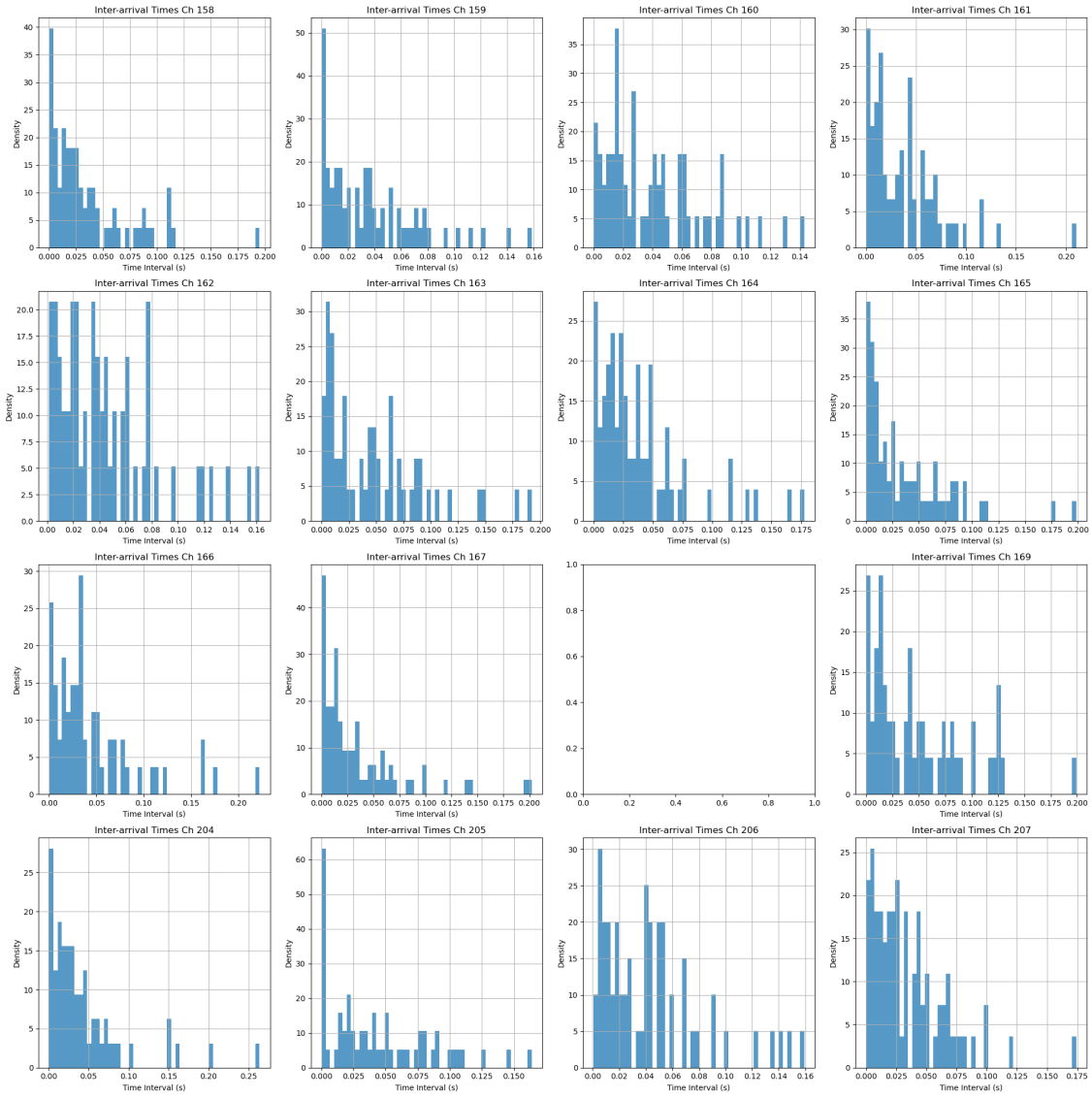


Figure 4.25: Inter-arrival Times Analysis for unshielded low threshold Dataset.

Channel 164, figure4.25 shows a broad and irregular distribution, indicating irregular event timing likely due to noise. Channel 165 also shows a wide distribution but with slightly more regularity than Channel 164. Channels 162 and 163 have more narrowly distributed inter-arrival times, with Channel 162 showing the most regular event intervals, indicating a cleaner and more stable signal. Channel 163 also demonstrates regular event timing but with slight irregularities.

Channel 167 has a wide and irregular distribution with multiple peaks, which suggests in-

consistent event timing and significant noise interference. Similarly, Channel 169 also shows irregularity, but the distribution is slightly narrower, indicating more consistent intervals than Channel 167, although it is still affected by noise. Channel 166 has a more consistent pattern of event timing and less noise interference, as shown by the regular distribution of inter-arrival times and a distinct peak in the analysis. This aligns with the lower noise levels observed in the FFT and STFT analyses for Channels.

The inter-arrival time histograms in Channel 207 display a broad and irregular distribution, stating uneven event timing due to noise. Channel 204 has a similar but slightly narrower distribution, suggesting more regular event intervals. Channel 206 shows further improvement with a more concentrated distribution, indicating more consistent event timing. Channel 205 stands out with the narrowest and most regular distribution, reflecting minimal noise interference and consistent event intervals.

4.4.3.3 Auto Correlation Analysis For Unshielded Low Threshold Dataset

In these plots,figure4.26 a sharp peak at zero lag is expected, indicating the perfect correlation of the signal with itself. However, significant fluctuations around the zero-lag peak are particularly notable across all four channels, indicating a lack of smooth decay in the autocorrelation function. For Channel 158, these fluctuations are pronounced, reflecting a high noise level and instability in the signal. These fluctuations suggest that the signal experiences random variations over time, contributing to the noise and irregularities observed in the inter-arrival time analysis. Channel 161 shows similar fluctuations, though slightly less severe than Channel 158. This indicates a noise reduction but points to underlying instability in the signal's temporal structure. Channel 159 exhibits these fluctuations but is less pronounced, suggesting a more stable signal with reduced noise levels. However, fluctuations still indicate some residual noise or irregularity in the signal. Channel 160 shows the least fluctuation around the zero-lag peak, with a smoother decay in the autocorrelation function. This suggests Channel 160 has the most stable and consistent signal, with minimal noise interference.

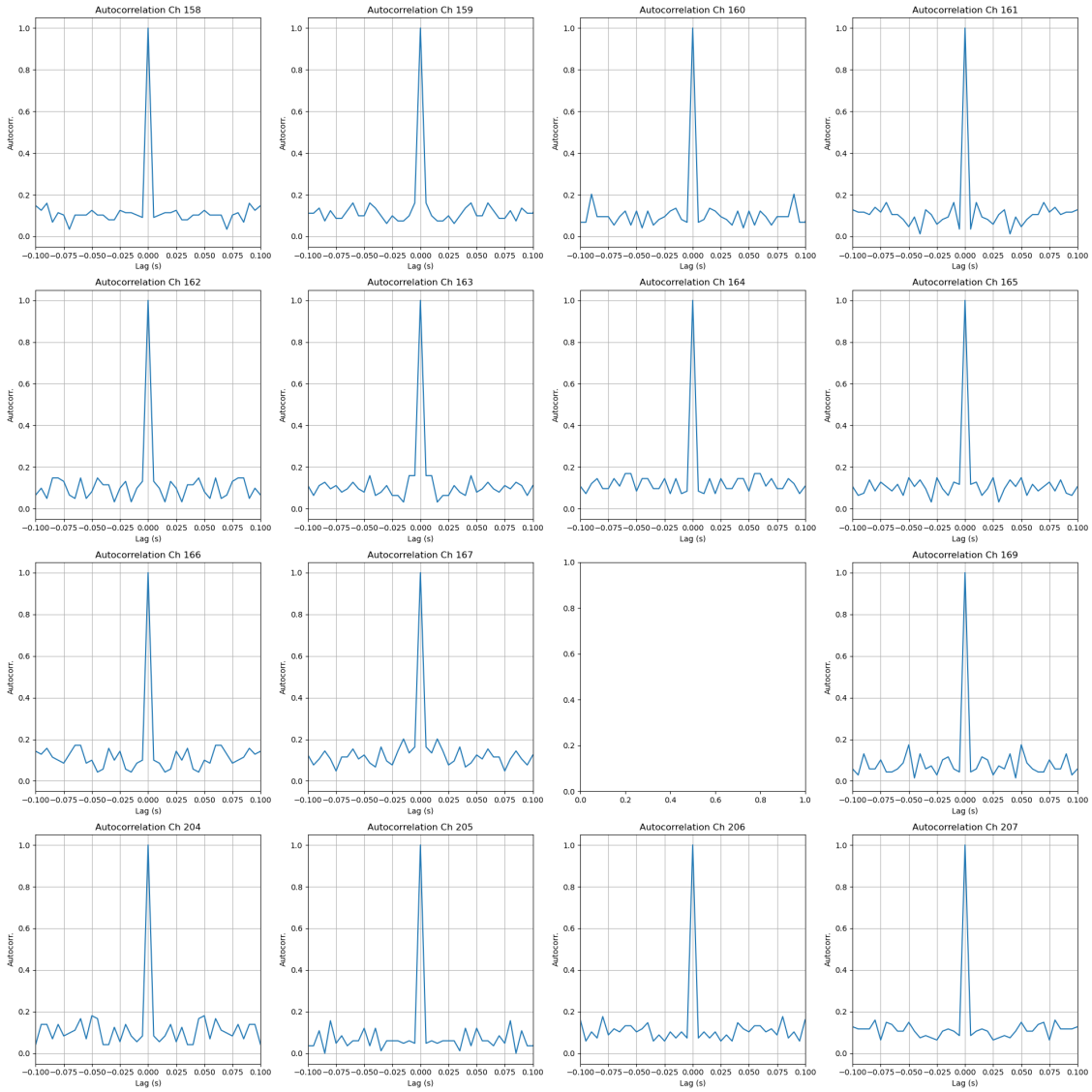


Figure 4.26: Auto Correlation Analysis for Unshielded Low Threshold Dataset

.

Channel 164 displays significant fluctuations around the zero-lag peak, indicating high noise and temporal instability. Channel 165 also shows fluctuations, but they are less severe, reflecting slightly better temporal consistency. Channels 162 and 163 have much smoother autocorrelation plots, with Channel 162 showing the least fluctuation around the zero-lag peak, indicating the most consistent and stable signal over time. However, channel 163 exhibits good stability with marginally more noise than Channel 162.

Channel 167 displays large fluctuations around the zero-lag peak, indicating that the signal's temporal structure is noisy and unstable. These random fluctuations suggest that the signal is highly affected by noise. Channel 169 also has some fluctuations but is less strong than Channel 167. This suggests a more consistent signal but still some noticeable interference. Channel 166 has the smoothest plot, showing minimal fluctuations around the zero-lag peak. This indicates the most stable and consistent signal with the slightest noise interference.

Also, Channel 207 shows significant fluctuations around the zero-lag peak, indicating high temporal instability and noise. Channel 204 also has fluctuations, though they are slightly less pronounced, indicating better signal consistency. Channel 206 exhibits fewer fluctuations than Channel 204, offering a more stable signal with less noise. Channel 205 has the smoothest autocorrelation plot between the four channels, with minimal fluctuations around the zero-lag peak, indicating the most stable and consistent signal among the four channels.

4.4.3.4 Short-Time Fourier Transform (STFT) Analysis For Unshielded Low Threshold Dataset

The STFT analysis for Channels 158 to 161, figure 4.4.3.4 provides information about the time-frequency characteristics of the signals, highlighting the noise levels and temporal stability of the frequency components. Channel 158's STFT reveals a highly noisy signal, with energy spread across a wide range of frequencies over time, showing significant and persistent interference. This noise is consistent throughout the time window, suggesting ongoing instability in the signal.

Channel 161 also shows considerable noise in its STFT, though it is slightly less pronounced than in Channel 158. The time-frequency distribution still reflects significant interference but with more defined structures than Channel 158. Channel 161's STFT reveals a similar noise pattern, though less intense, with frequency components being more stable over time yet still showing some noise-related fluctuations. Finally, Channel 160 displays the cleanest STFT among the four channels, with a more stable and uniform energy distribution across frequencies over time, indicating that this channel is less affected by noise and maintains a more consistent signal.

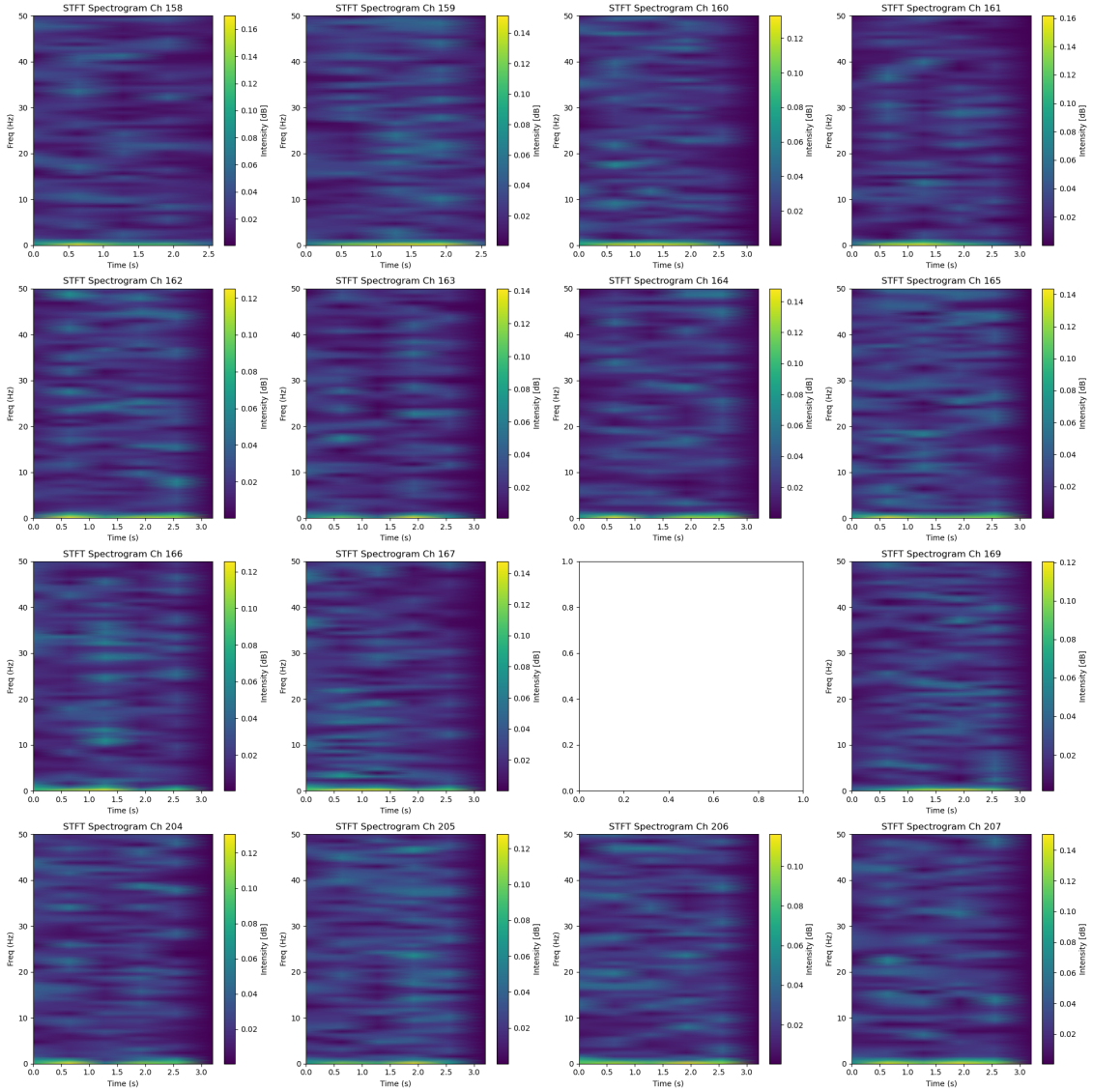


Figure 4.27: Short-Time Fourier Transform (STFT) Analysis for unshielded low threshold Dataset

.

Channel 164 shows a significant energy spread across a wide range of frequencies, indicating considerable noise and temporal instability. The distribution appears chaotic, with no clear, consistent frequency components characteristic of a highly noisy signal. Channel 165 also displays evidence of noise, though it is less severe than in Channel 164. The time-frequency representation shows some structured patterns, but the noise still disrupts the signal, leading to a less stable frequency distribution over time. In contrast, Channels 162 and 163 have different

time-frequency characteristics. Channel 162 has a consistent energy distribution with minimal interference, indicating a stable and clean signal. The frequency components in this channel remain consistent over time, suggesting minimal noise impact. Similarly, Channel 163 exhibits a relatively stable time-frequency distribution, though slightly more variation than Channel 162.

Channel 167 exhibits significant noise across a wide range of frequencies, with chaotic and unstable energy distribution throughout the signal duration, describing continuous interference. Channel 169 also shows less severe noise, with some structured patterns but still noticeable disruptions. In contrast, Channel 166 displays a much more stable and consistent time-frequency distribution, with energy concentrated in specific frequency bands and minimal noise, confirming it as the channel with the cleanest signal.

Channel 207 exhibits the most significant noise, with an unstable and widespread energy distribution across a wide range of frequencies over time, indicating continuous interference. Channel 204 also shows noise but is slightly less than Channel 207. Channel 206 displays a more stable and structured energy distribution, with less noise interference than Channels 204 and 207, though some instability remains. Channel 205 has the cleanest time-frequency representation, with concentrated energy in specific frequency bands and minimal noise.

4.4.3.5 Conclusion and Correlation of Unshielded Low Threshold Dataset with Spectrum Analyzer Findings

In summary, Channel 167 has the highest noise level and instability compared to all other channels, making it the noisiest channel. Channel 169 has significant noise but is slightly less severe than Channel 167. Channel 158 has noticeable noise but is more stable than Channels 167 and 169. Channels 161 and 158 are both noisy, but 161 is less noisy. However, Channels 205, 163, 160, and 162 consistently have the clearest and most stable signals. Channel 205 has the least noise and the most reliable signal. This analysis indicates that Channels 205, 163, 160, and 162 are best for further processing and analysis. Channels 167, 169, and 158 have higher noise levels, so you should be cautious when dealing with them.

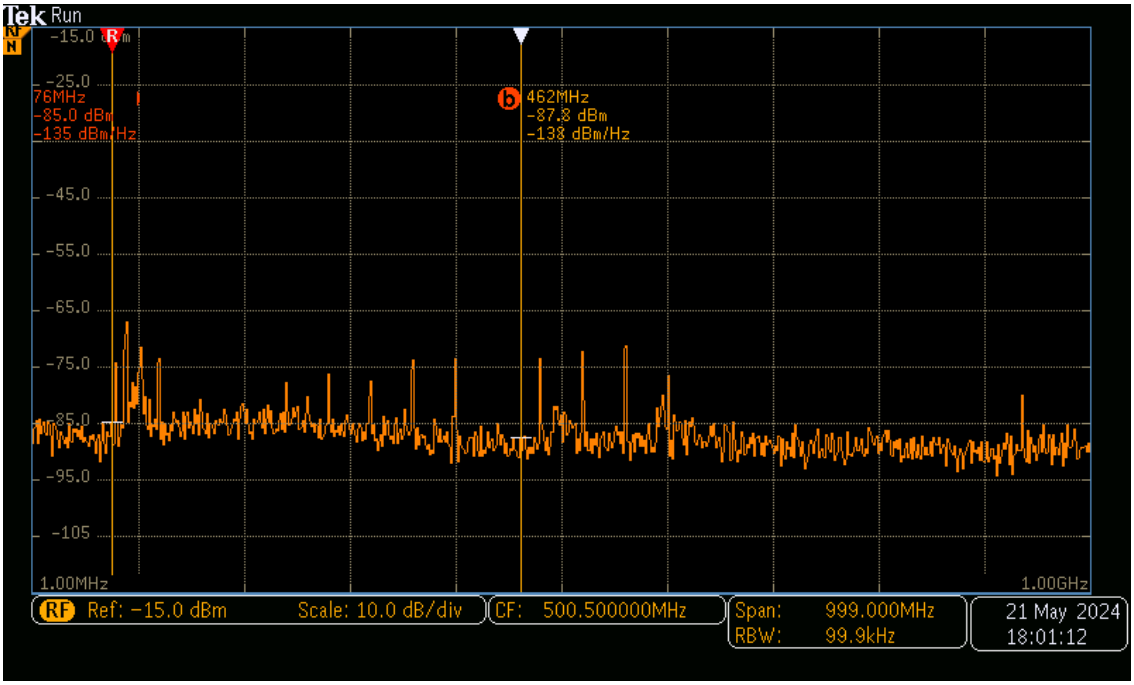


Figure 4.28: spectrom analyzer for Unshielded Low Threshold Analysis.

The spectrum analyzer, figure 4.28, shows a frequency spectrum ranging from 1 MHz to 1 GHz, with significant peaks around 76 MHz and 462 MHz, indicating specific frequencies with considerable energy or interference. These peaks likely correlate with the high noise levels detected in Channels 167 and 169 from the unshielded low threshold dataset. The broad distribution of noise across the spectrum, as seen in the analyzer, matches the chaotic and widespread energy observed in these channels, suggesting to persistent interference. Channel 158, while also noisy, shows slightly more stability than Channels 167 and 169. The extensive noise across the analyzer's spectrum could explain the variations and noise identified in Channel 158.

In contrast, Channels 205, 163, 160, and 162, which were identified as having the most minor noise in the dataset, align with the spectrum analyzer's observation that, despite the noise peaks, The background noise level stays fairly consistent and not too high, especially in areas without strong peaks. The stability in the noise level probably happens because the signals are clearer in these channels. This means the results from the spectrum analyzer support the dataset analysis. The frequencies have significant energy in the noisiest channels, while the cleaner channels have more stable parts of the frequency spectrum.

4.4.4 Dataset4:Unshielded High Threshold Analysis

4.4.4.1 FFT Analysis For Unshielded High Threshold

In figure4.29, FFT plots for Channels 158 to 161 reveal the frequency spectrum of these signals. While Channel 159 shows a slightly noisier signal than the others, the channels exhibit similar spectral characteristics overall. The frequency distributions are relatively flat, with no significant peaks or anomalies that would clearly distinguish one channel from the others. The similarity across all channels suggests that they all have similar noise levels, making it hard to identify one as noticeably noisier than the others. However, Channel 159 shows slightly more variation in volume, indicating a slightly higher noise level.

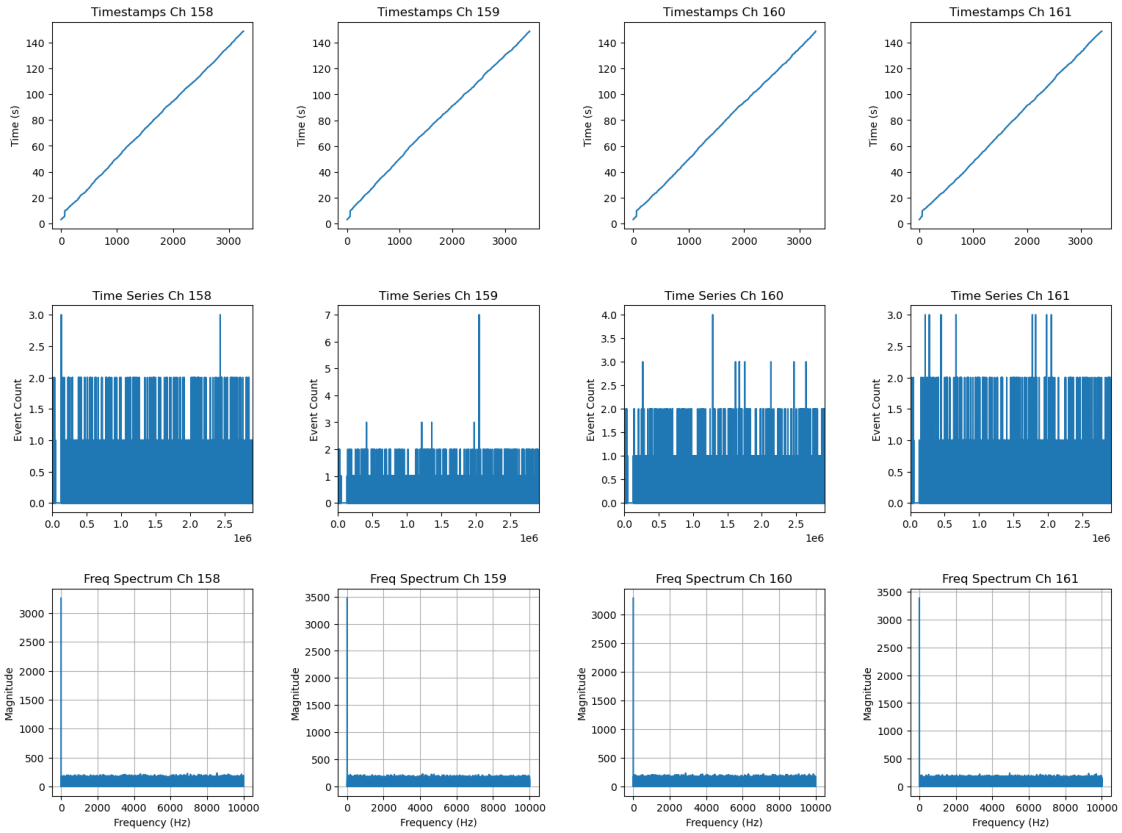


Figure 4.29: Time Series and Frequency Spectrum Analysis for Channels 158 to 161 (Unshielded High Threshold).

In the figure4.30, Channel 162, in particular, shows a noticeable step in the data, which could indicate a minor but distinct shift or anomaly in the signal's frequency content. However, all four channels show similar frequency distributions with no significant noise or anoma-

lies distinguishing one channel from the others. The spectra are relatively flat and consistent, representing that the signals are not heavily influenced by noise and maintain a stable frequency.

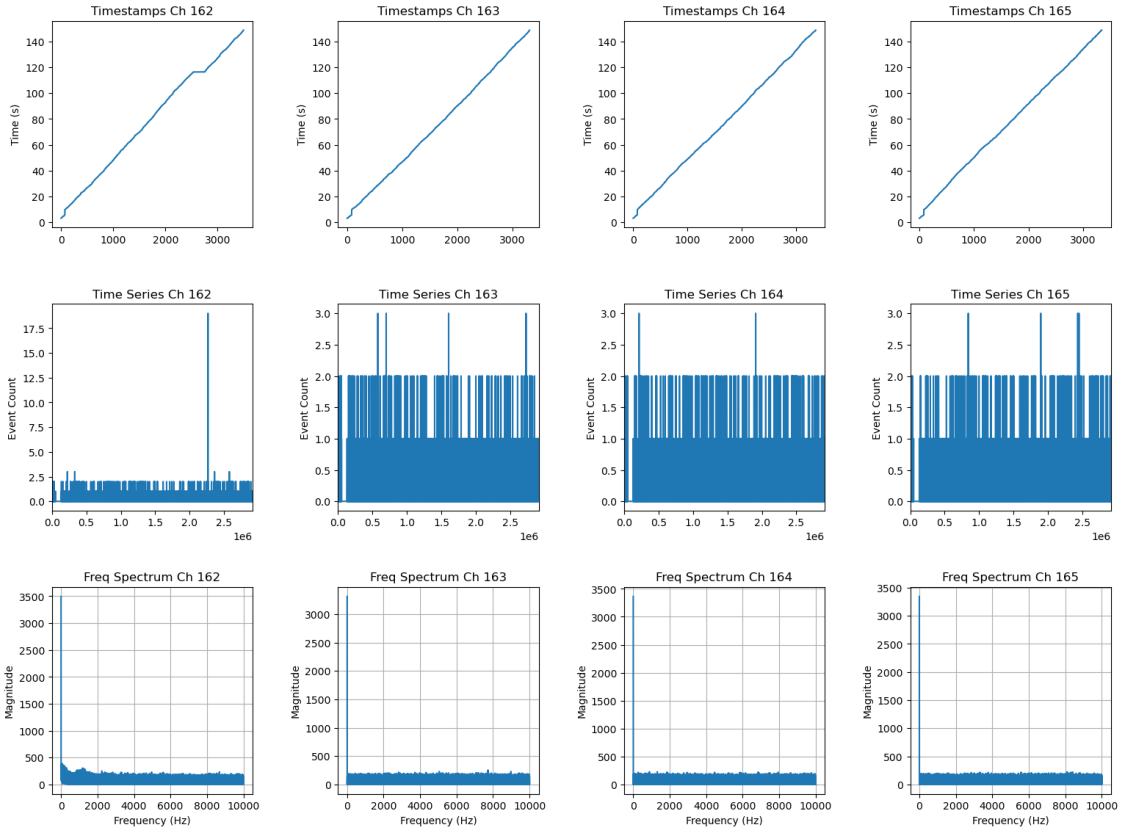


Figure 4.30: Time Series and Frequency Spectrum Analysis for Channels 162 to 165(Unshielded High Threshold).

Moreover, FFT analyzing between 167,168, and 169, Channel 169 shows a subtle step in the data, suggesting a slight deviation or anomaly in the frequency content. Overall, the spectral characteristics of all three channels are very similar. The frequency distributions are mainly flat and consistent across these channels, with no significant peaks or noise that would clearly distinguish one channel as noisier than the others. The minor differences, such as the slight step in Channel 169, represent very subtle variations, but all channels show low noise levels and stable frequency content. figure 4.31 shows these analyses.

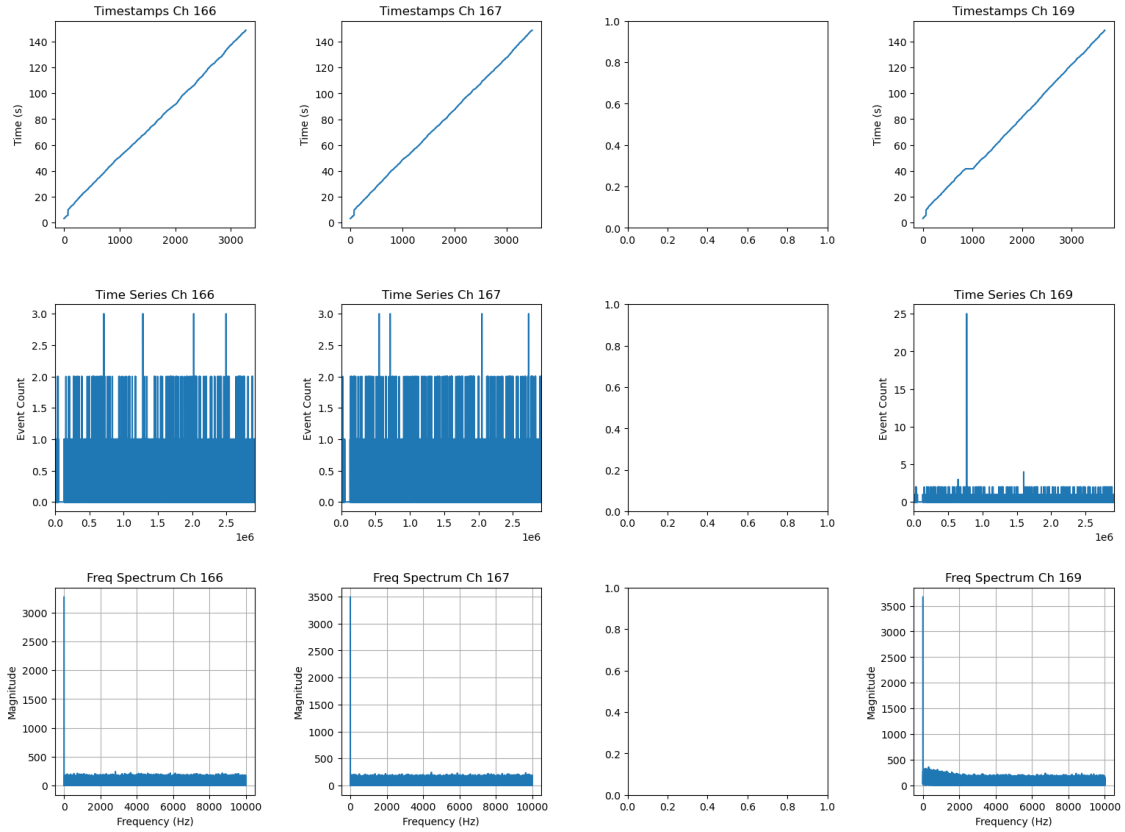


Figure 4.31: Time Series and Frequency Spectrum Analysis for Channels 167 to 169(Unshielded high Threshold Dataset).

The FFT plots, figure 4.32 reveal that the frequency spectra of Channels 204 to 207 are very similar, with no noticeable peaks or anomalies that could offer remarkable noise. Channel 207, initially slightly noisier, shows minimal variations from the others, suggesting that all channels have a similar frequency distribution and low noise levels.

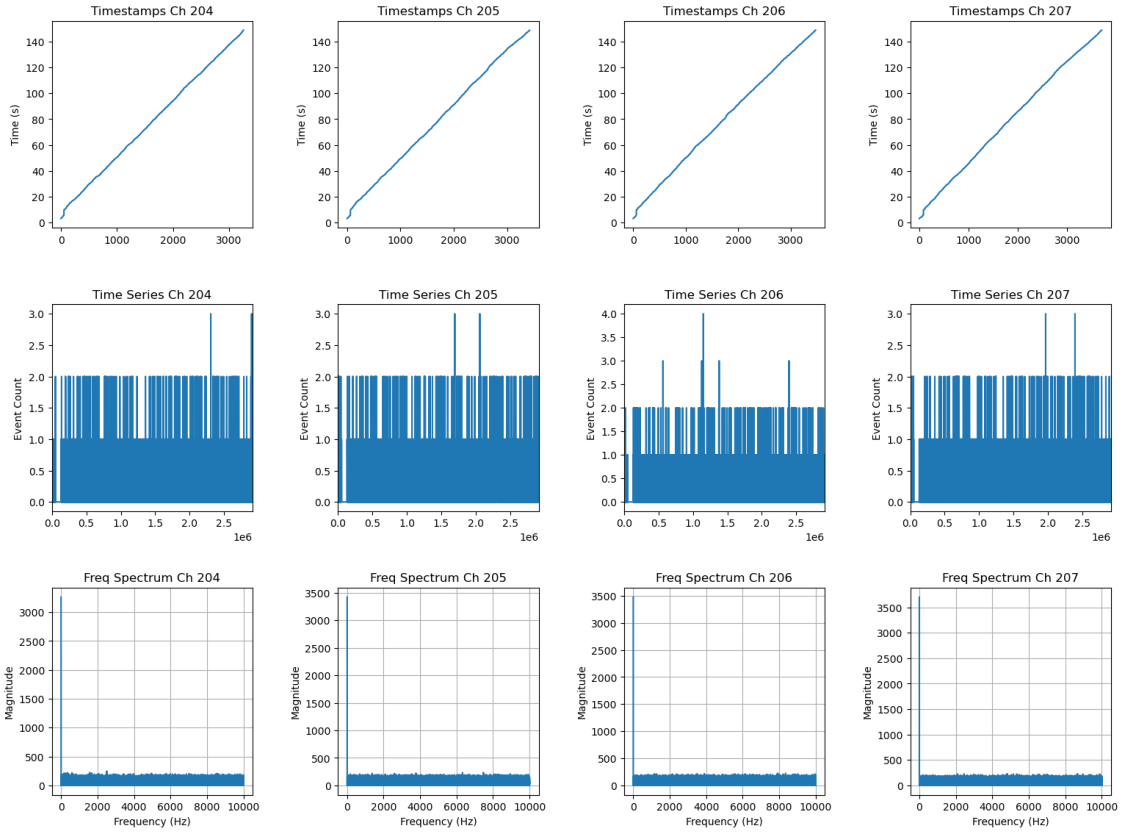


Figure 4.32: Time Series and Frequency Spectrum Analysis for Channels 204 to 207 (Unshielded High Threshold Dataset).

4.4.4.2 Inter-Arrival Times Analysis for Unshielded High Threshold Dataset

The histograms,figure4.33 show that most events occur within a very short time interval, with all channels displaying a similar pattern. The data shows that there are no significant unusual data points or irregularities. This suggests that the timing of the event is consistent across all channels. Channel 159 may have slightly more irregular event intervals, but all channels show similar regularity overall.

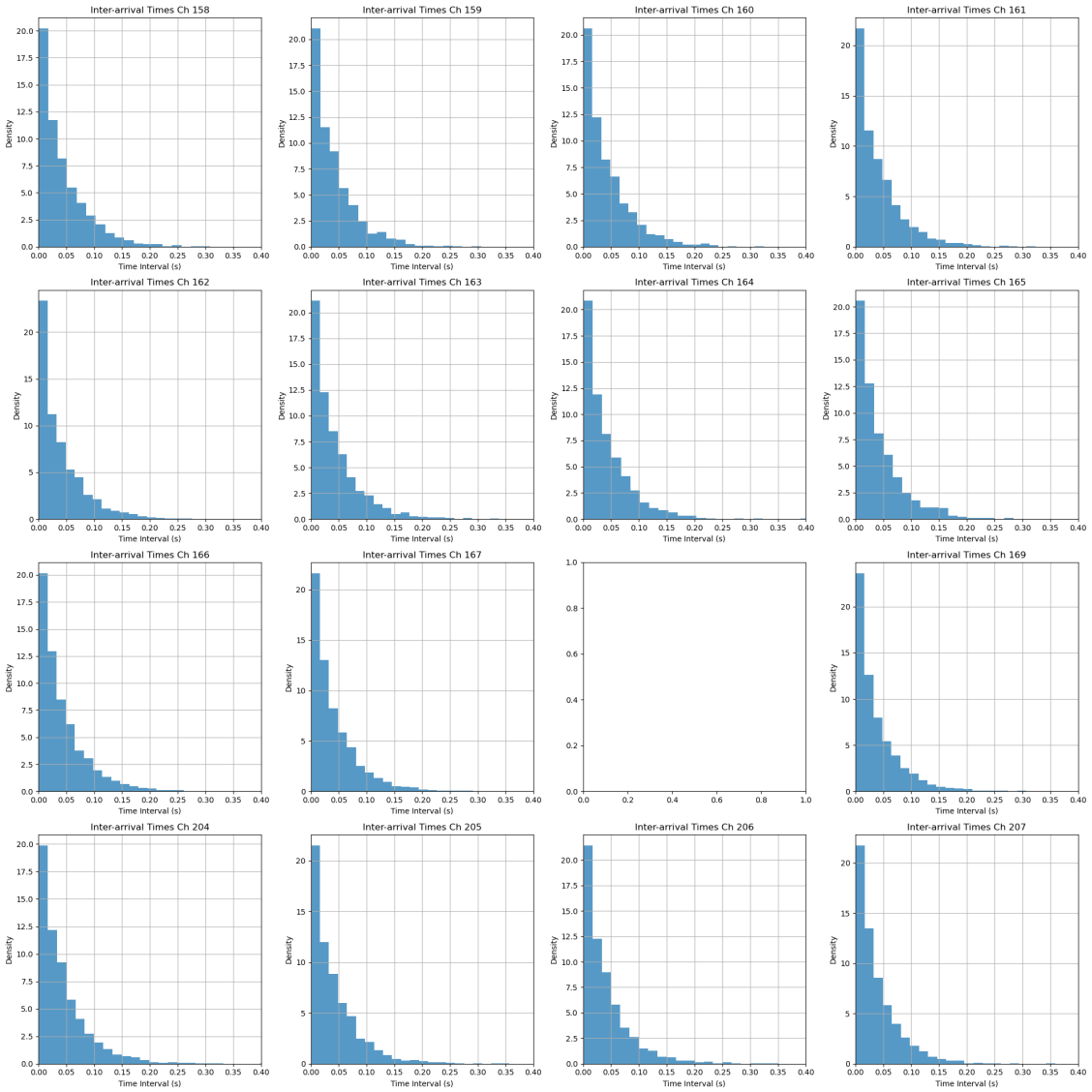


Figure 4.33: Inter-arrival Times Analysis for Unshielded High Threshold Dataset.

The inter-arrival time histograms for Channels 162 to 165 show the distribution of time intervals between consecutive events. The histograms show that most events occur within very short intervals, with all channels displaying similar patterns. There are no remarkable outliers, which shows that the event timing is consistent across the channels. Even though Channel 162 shows slight deviations in other analyses, its inter-arrival time distribution remains regular, showing that these deviations do not significantly impact the timing of events in the signal.

The histograms for the time between events on Channels 166, 167, and 169 show that most

events happen close together. Although Channel 169 might have slightly more uneven timing, the overall differences between the channels are slight. This consistent timing suggests that the signals are stable and not affected much by noise.

The inter-arrival time histograms for Channels 204 to 207 all show that events occur within very short intervals, with a highly consistent distribution across channels. There are no significant outliers or variations in the tail of the distributions, offering that the timing of events is regular in all channels.

4.4.4.3 Auto Correlation Analysis For Unshielded High Threshold Dataset

The autocorrelation plots, figure 4.34 All four channels exhibit a sharp peak at zero lag; the data shows smooth decays, indicating relatively stable signals as expected. However, subtle fluctuations are present in the autocorrelation functions, particularly in Channel 159, which means a slightly higher level of temporal variations and noise. Despite this, the overall similarity in the autocorrelation plots across all channels builds up the notion that they are all affected by similar noise characteristics, with only minor differences in stability.

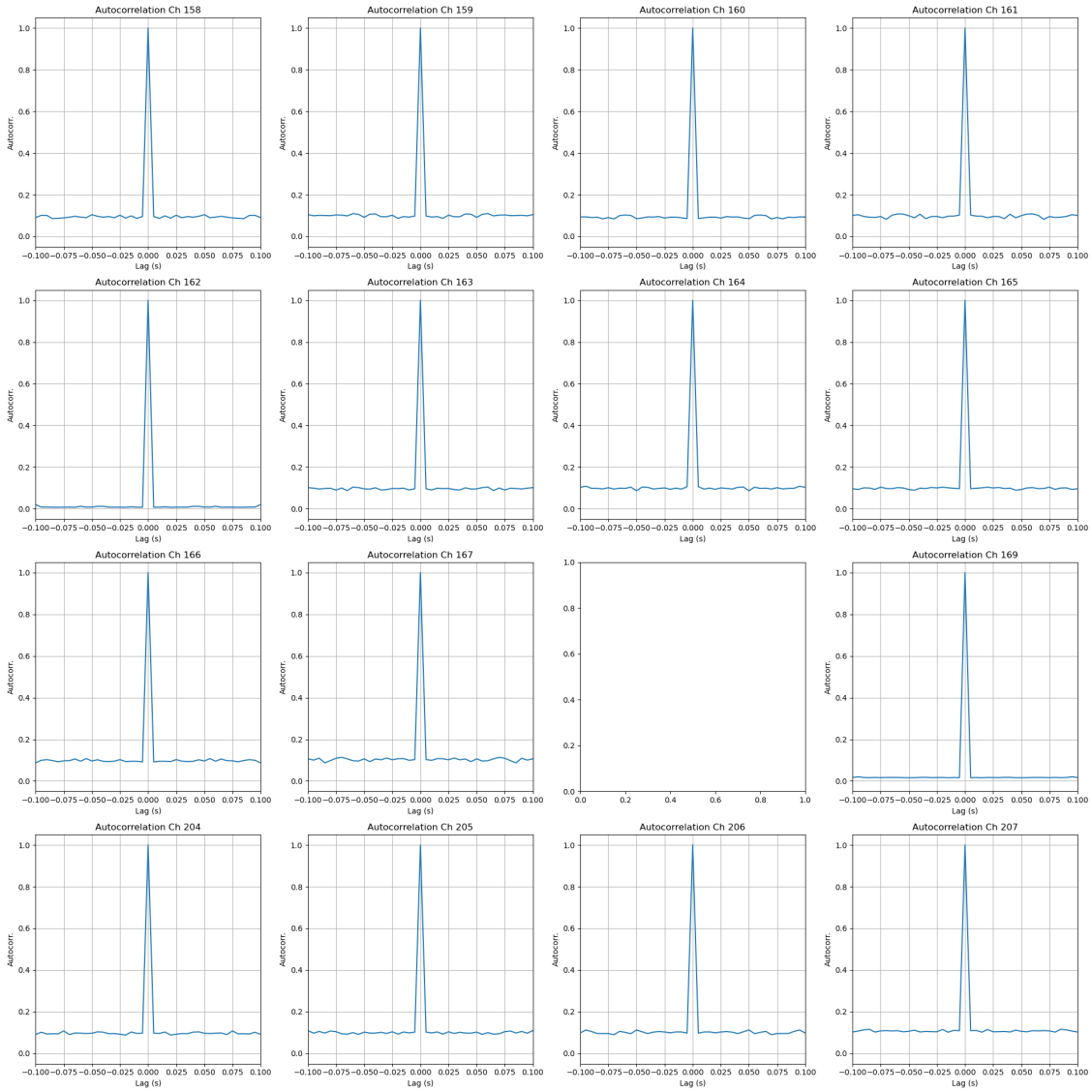


Figure 4.34: auto correlation Analysis for unshielded high threshold Dataset.

As expected, the autocorrelation plots for Channels 162 to 165 display a sharp peak at zero lag, indicating strong self-correlation. However, the autocorrelation plot for Channel 162 shows slight irregularities, with the level of the autocorrelation curve near zero lag being lower and flat than the other channels. This could correlate with the step observed in the FFT and the subtle variations seen in the STFT. Despite these relatively minimal differences, the overall temporal stability across the channels remains similar, indicating that these irregularities do not significantly impact the consistency of the signals.

As expected, the autocorrelation plots for Channels 166, 167, and 169 demonstrate strong temporal consistency, with sharp peaks at zero lag. However, Channel 169 shows very slight irregularities, with the level of the autocorrelation curve near zero lag being slightly lower and flat compared to the other channels. The autocorrelation functions for all channels show that the signals maintain high stability, with no significant temporal noise that would differentiate one channel as noisier than the others.

The channels between 204 and 207 exhibit a sharp peak at zero lag with minimal fluctuations, representing high temporal stability. No significant irregularities in the autocorrelation functions show that these channels behave similarly.

4.4.4.4 Short-Time Fourier Transform (STFT) Analysis For Unshielded High Threshold Dataset

The STFT plots,figure4.35 show that all four channels have a consistent energy distribution across time and frequency, with no significant deviations or unique noise patterns. The energy spread out across the entire range of frequencies, and no prominent fluctuations or shifts would indicate the presence of substantial interference in any specific channel. This also shows that the noise levels are similar across all these channels, and no one channel has significantly higher noise levels than the others.

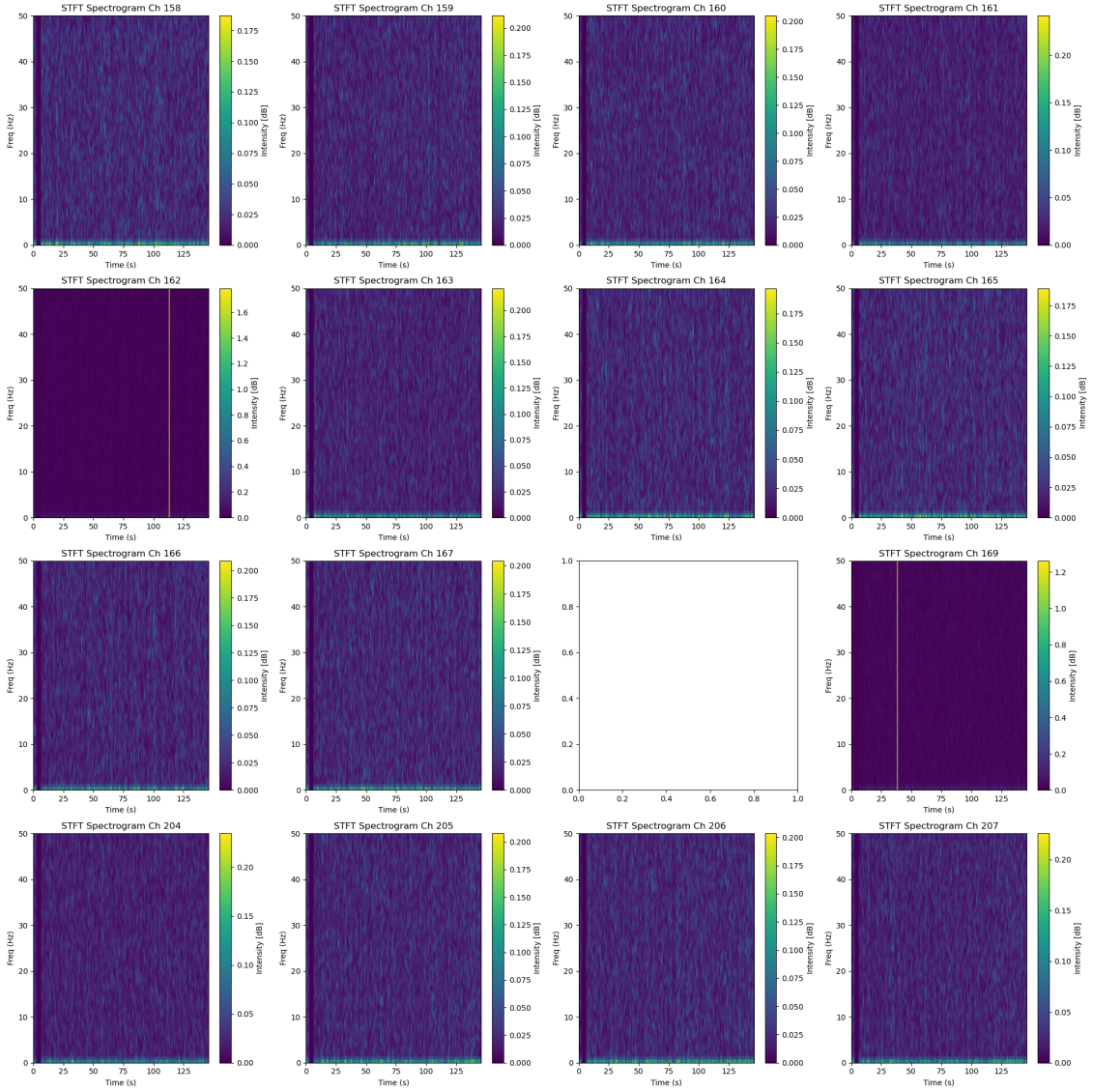


Figure 4.35: STFT Analysis for Unshielded High Threshold Dataset.

The STFT plots for Channels 162 to 165 show that Channel 162 has a distinct vertical line and a more visible background, indicating a recurring event or interference. This aligns with the step observed in the FFT data, offering that while Channel 162 remains generally stable, it experiences slight, consistent fluctuations not seen in the other channels. Despite these differences, the overall signal integrity across all channels remains strong, with minimal noise.

Channel 169 shows a slightly more visible background with a subtle vertical line, which might indicate an event or interference. However, like the FFT results, the overall time-frequency

distribution remains consistent across Channels 166, 167, and 169, representing that the signals are stable over time with minimal fluctuations.

The STFT plots further confirm the consistency across the channels between 204 to 207. The time-frequency distributions show uniform patterns with no significant variations or disruptions. The slight vertical lines observed in some channels, particularly in Channel 207, are subtle and do not show any major noise issues. The overall background remains consistent in all channels, fixing the stability of the signals.

4.4.4.5 Conclusion And Correlation Of Unshielded High Threshold Dataset With Spectrum Analyzer Findings

In conclusion, the channel analyses show remarkably similar signal stability, noise levels, and temporal consistency characteristics. However, Channels 162 and 169 stand out slightly due to minor anomalies. Channel 162 and 169 show a step in the FFT data and subtle irregularities in the autocorrelation. At the same time, Channel 169 exhibits a very slight increase in noise, particularly in the STFT and autocorrelation plots. Despite these small variations, these differences are minimal, and overall, all channels maintain stable signals with low noise levels, making them equally reliable for further processing and analysis.

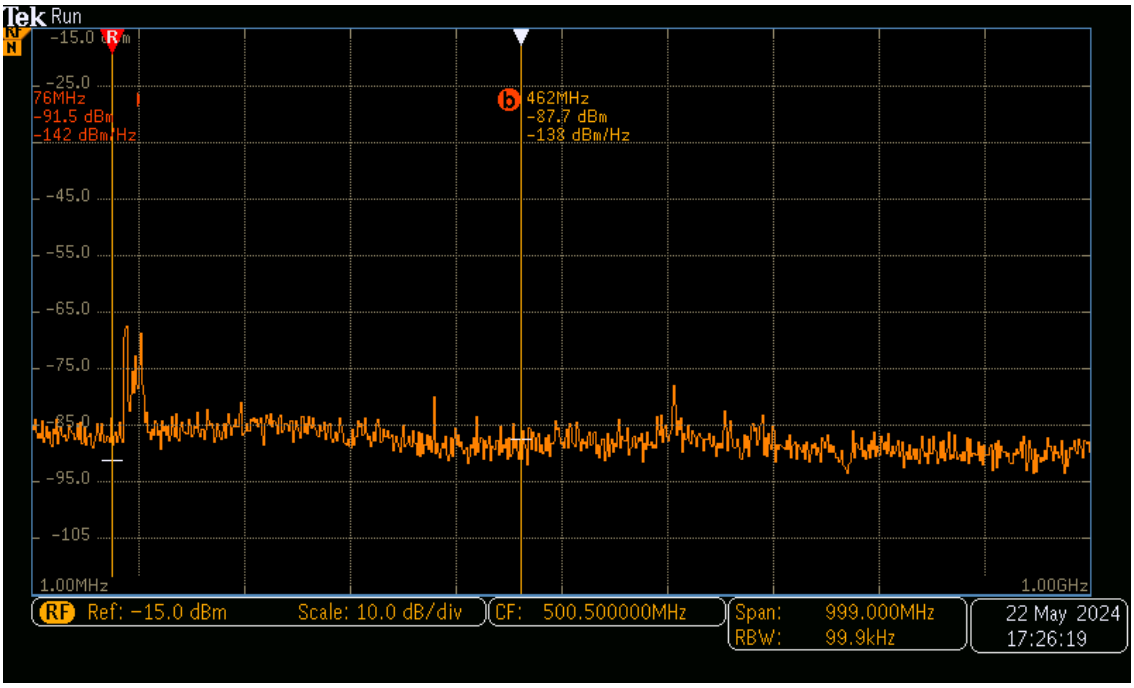


Figure 4.36: Unshielded High Threshold Dataset with Spectrum Analyzer.

With the legacy electronic equipment turned off, the spectrum analysis figure4.36 revealed significantly reduced noise levels across all channels. The previously prominent peaks associated with the legacy electronics have largely disappeared, confirming that much of the earlier detected interference was directly caused by these devices. This noise reduction directly correlates with the minor anomalies and slightly elevated noise levels observed in Channels 162 and 169 during the earlier signal analyses (FFT, STFT, autocorrelation, and inter-arrival time). These channels were particularly sensitive to the noise from the legacy electronics, which led to the observed deviations when the electronics were active. By eliminating this source of interference, the signals across all channels became more uniform, stable, and reliable. This result emphasizes the importance of minimizing environmental noise to maintain the integrity of signal processing and analysis.

4.5 Anomaly Detection Using Machine Learning

Using machine learning algorithms to detect anomalies in event counts is a powerful and meaningful way to ensure the accuracy of time-sensitive systems, such as Drift Tubes or other experimental setups that depend on precise event timing. Each recorded event, or hit, is important for analyzing overall performance in these systems. These events are logged with specific timestamps, and the frequency or count of events over time provides detailed insight into the system's behavior. Under normal conditions, the event count at each time step should follow a consistent pattern, reflecting the system's expected performance.

However, any significant variation from this expected pattern, whether it's an unusually high or low number of events, signals abnormal behavior or an anomaly. Such anomalies can arise from various factors, including noise, external interference, or internal system issues. Identifying these anomalies is critical because they can serve as early warnings of potential problems, allowing us to take action before more serious issues occur.

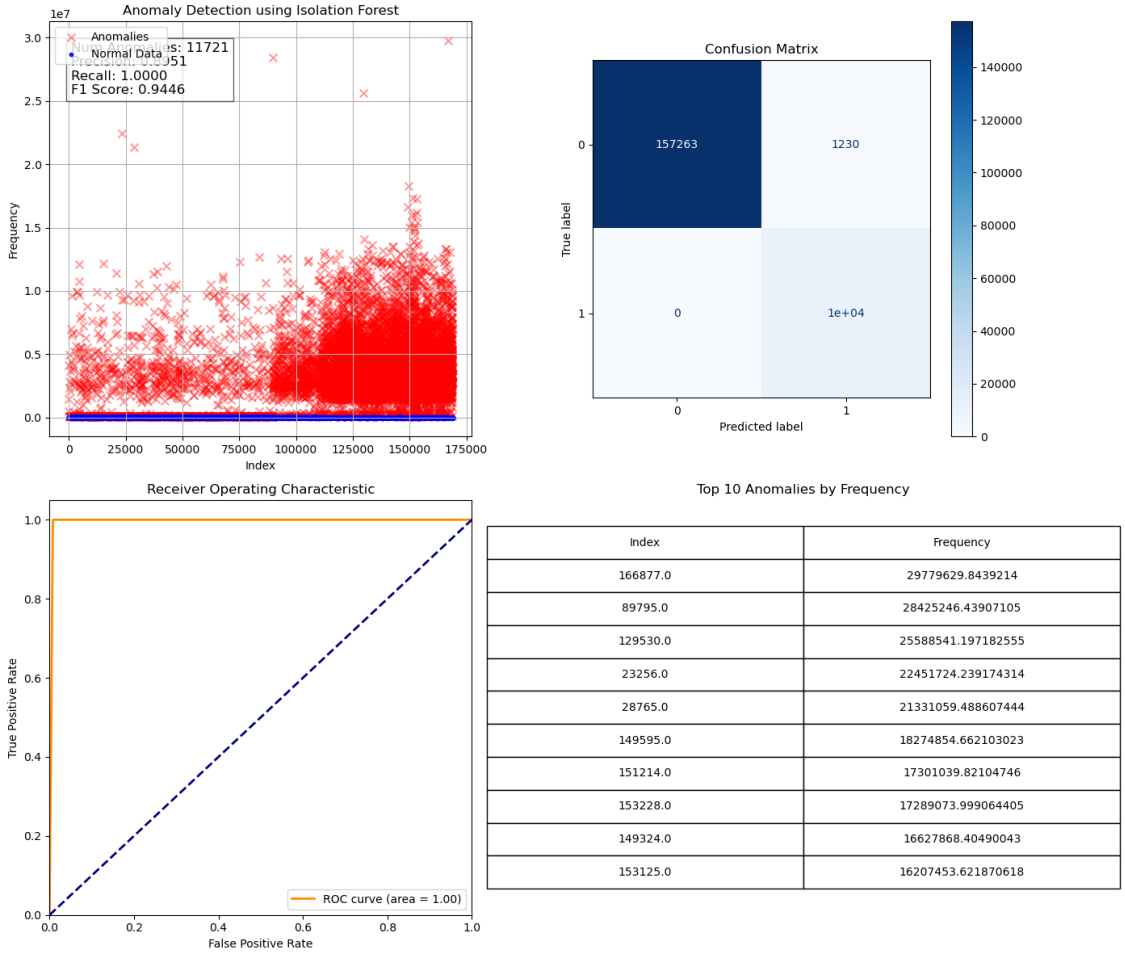
Anomaly detection algorithms are particularly suitable for identifying outliers, as data points differ significantly from most datasets. In event counts, an outlier represents a time step where the number of events deviates substantially from the norm. One of the key strengths of these algorithms is their ability to detect anomalies without requiring predefined notions of what normal or abnormal event counts are. Instead, they automatically identify these anomalies based on how isolated or unusual the events are within the overall data distribution. This capability makes them highly effective tools for uncovering hidden patterns and issues in complex datasets.

To achieve this, we employed advanced machine-learning techniques to detect critical noise events that could affect the functionality of Drift Tubes. We utilized models such as Isolation Forest, Local Outlier Factor (LOF), One-Class SVM, and Gaussian Mixture Model (GMM) to identify abnormal patterns in time-series data. The data was thoroughly prepared and processed to address missing values and standardize frequency data, ensuring consistency and accuracy. Hyperparameters for each model were fine-tuned to optimize performance, and the models were evaluated using precision, recall, F1 score, and ROC-AUC metrics. Our findings demonstrate that these models effectively detected multiple high-frequency noise anomalies, underscoring the critical role of machine learning in improving the robustness of the system.

4.5.1 Machine Learning Approaches for Analyzing Shielded Data

4.5.1.1 Isolation Forest Technique

In this analysis, machine learning techniques were utilized to identify anomalies in the frequency data using the Isolation Forest algorithm. The model detected 11721 anomalies, represented as red points in the scatter plot. These irregularities show significant differences from the expected frequency patterns. The model is evaluated using a few different metrics. The model's precision is 0.8951, showing the percentage of correctly identified anomalies. Its recall is 1.0000, indicating that it successfully detected all true anomalies. The F1 Score, which balances precision and recall, is 0.9446, representing the overall accuracy of the anomaly detection. The Confusion Matrix shows the model's classification results, confirming that all true anomalies were accurately identified without missing any. The ROC curve confirms the model's performance, with an AUC of 1.00, indicating excellent discriminatory ability. In addition, the top 10 anomalies are listed based on their frequency values, which may indicate particularly severe or impactful deviations in the data.the result demonstrated in the 4.37.



Index	Frequency
166877.0	29779629.8439214
89795.0	28425246.43907105
129530.0	25588541.197182555
23256.0	22451724.239174314
28765.0	21331059.488607444
149595.0	18274854.662103023
151214.0	17301039.82104746
153228.0	17289073.999064405
149324.0	16627868.40490043
153125.0	16207453.621870618

Figure 4.37: Result Of Isolation Forest for Shielded Dataset

We used a more advanced Isolation Forest model to thoroughly investigate anomalies in the protected dataset. Our goal was to uncover unusual patterns in absolute timing changes and closely examine how the number of events varied.

The results of this extended analysis, illustrated in Figure 4.38, highlight a new set of anomalies in the channels. The initial Isolation Forest analysis found 902 anomalies based on absolute timing deviations. The refined model focuses on detecting anomalies in the time series by examining the variations in event counts. The anomalies, shown as red points on the scatter plot, indicate time steps where event counts significantly differ from the norm, suggesting potential noise, interference, or other disturbances.

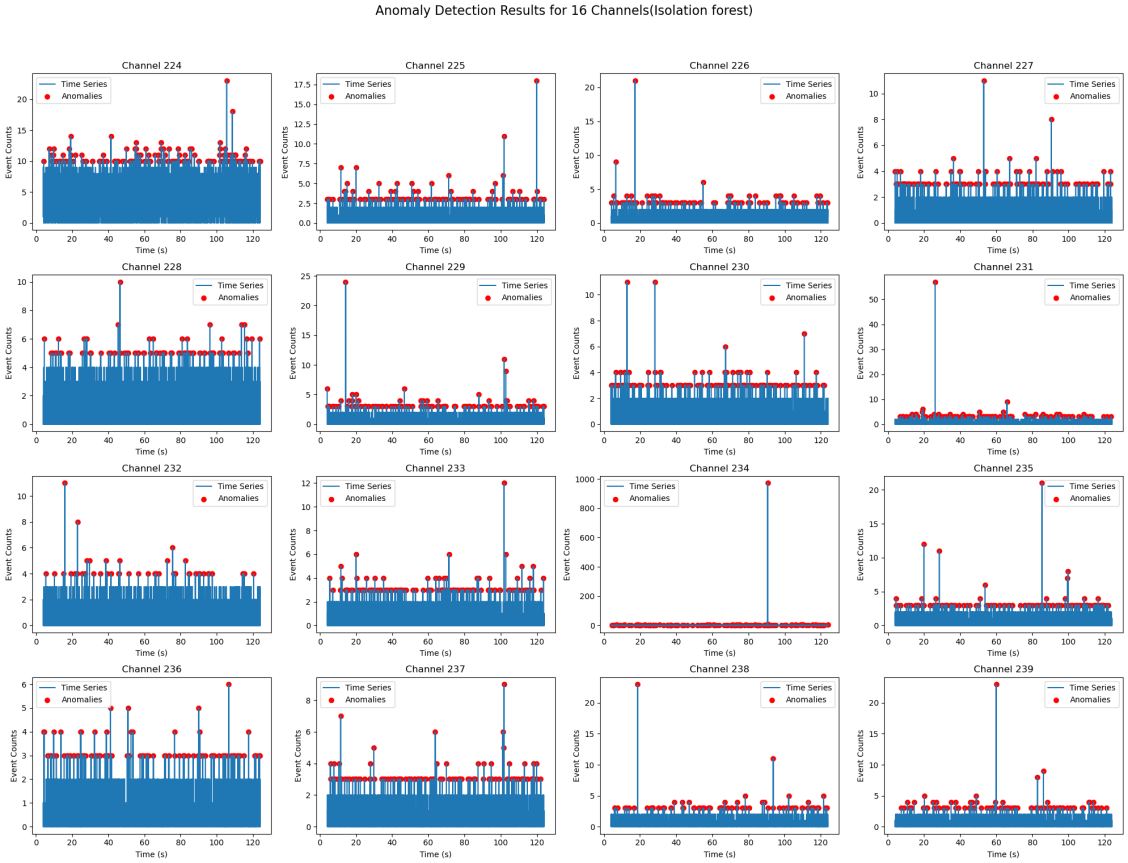


Figure 4.38: Anomaly detection results for 16 channels using the extended Isolation Forest model(shielded Dataset)

After checking the results, the model found unusual patterns in the number of events detected in each channel. some channels showed more frequent unusual patterns than others. These findings align with the patterns observed in the FFT, STFT, autocorrelation, and inter-arrival time analyses. This supports the idea that the unusual patterns have been correctly identified as anomalies.

The table below highlights each channel and the number of anomalies detected by the Isolation Forest model. The results from the Isolation Forest analysis confirm the findings from other analytical methods, showing that the detected anomalies are valid. For example, channels 224 and 228 in the shielded datasets, which displayed slight differences in the FFT and STFT analyses, also showed more anomalies in the event counts. These channels seem more likely to be affected by noise and interference. The anomalies identified by the Isolation Forest and the signal characteristics observed in the FFT, STFT, and other analyses show that this method effectively finds and understands noise patterns in the dataset.

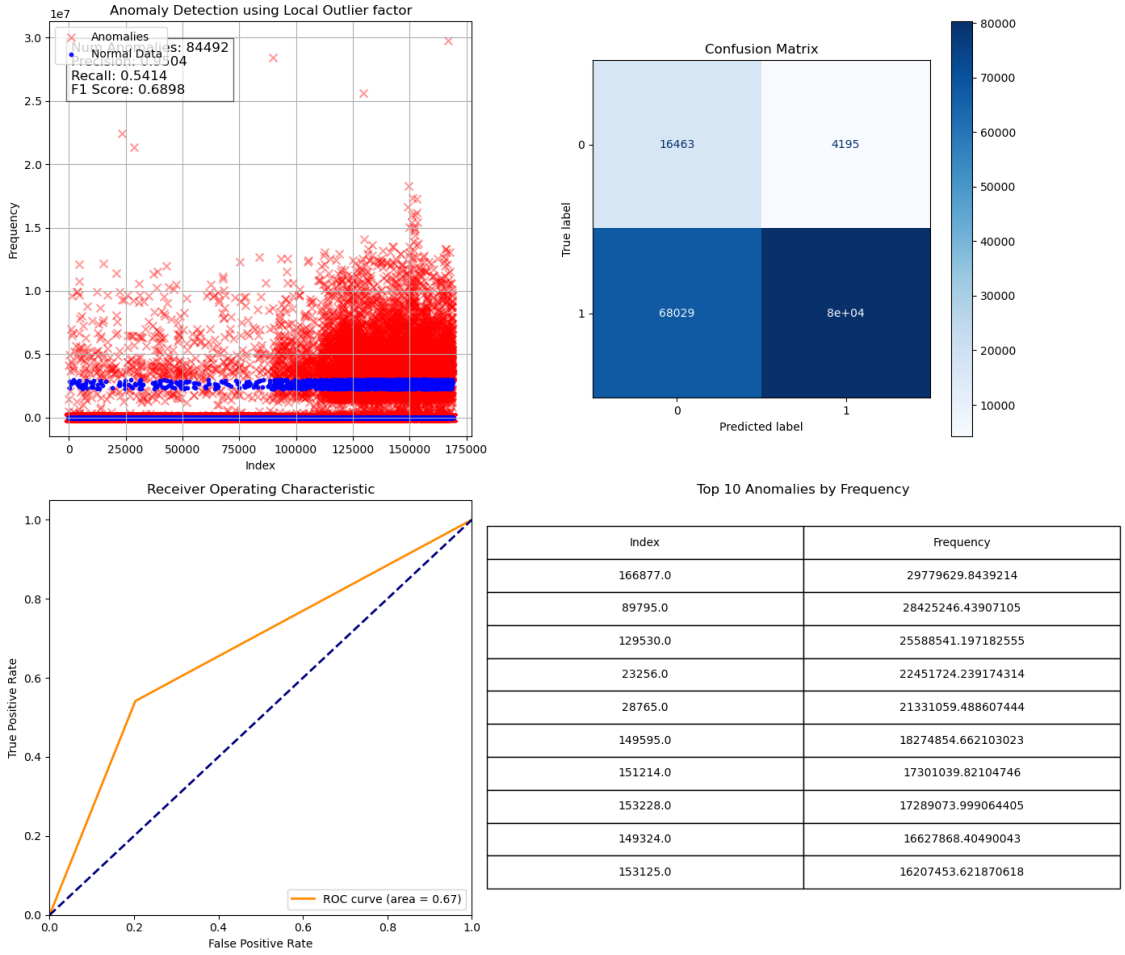
Channel	Number of Anomalies
224	78
225	68
226	61
227	64
228	69
229	56
230	67
231	70
232	25
233	60
234	65
235	62
236	55
237	57
238	34
239	44

Table 4.1: Number of Anomalies Detected with Iforest in Each Channel for shielded dataset

4.5.1.2 Local Outlier Factor(LOF) Technique

The LOF method spotted 84492 anomalies in the dataset, highlighted in red on the scatter plot, and assessed using precision, recall, and F1 Score. The precision of 0.9504 means only a small portion of the discovered anomalies were valid with a recall of 0.5414. The F1 Score of 0.6898 indicates the model's low effectiveness in balancing precision and recall.

The confusion matrix shows many false negatives (68,029) and a considerable number of false positives (4,195), resulting in a moderate precision score of 0.5414. The ROC curve, with an AUC of 0.67, indicates that the Local Outlier Factor (LOF) model performs slightly better than random guessing but has challenges effectively distinguishing anomalies from normal data. The table lists the top 10 anomalies by frequency, highlighting the most significant outliers. The result is demonstrated in the **??**.



Index	Frequency
166877.0	29779629.8439214
89795.0	28425246.43907105
129530.0	25588541.197182555
23256.0	22451724.239174314
28765.0	21331059.488607444
149595.0	18274854.662103023
151214.0	17301039.82104746
153228.0	17289073.999064405
149324.0	16627868.40490043
153125.0	16207453.621870618

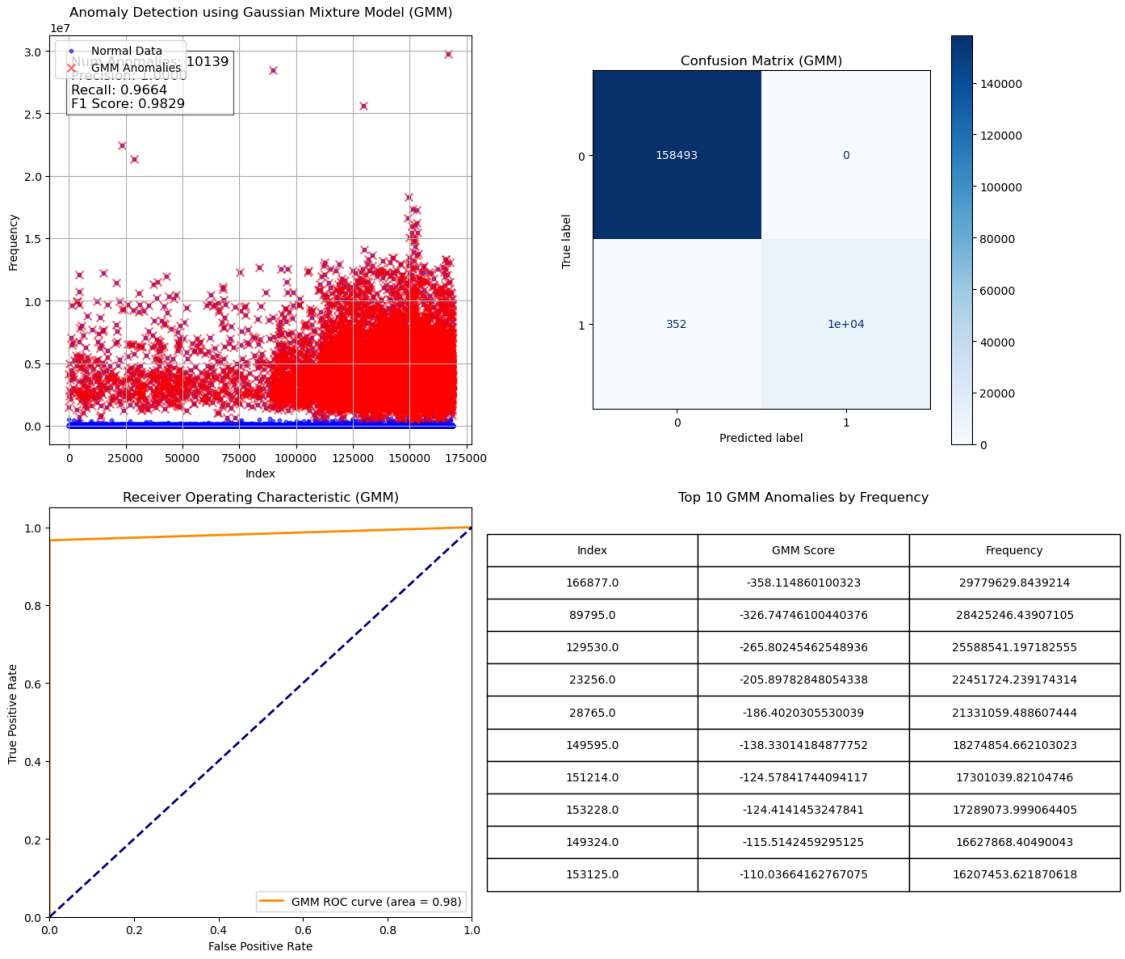
Figure 4.39: result of LOF for shielded dataset

4.5.1.3 Gaussian Mixture Model (GMM) for Shielded Dataset

The Gaussian Mixture Model (GMM) identified 10,391 anomalies in the frequency data, represented as red points in the scatter plot, indicating significant outliers. The model demonstrated high performance with a recall of 0.9664, meaning it correctly identified 96.64% of the actual anomalies. The F1 score was 0.9829, reflecting a strong balance between precision and recall, highlighting the model's overall effectiveness.

The confusion matrix shows that the model accurately classified all normal instances with no false positives, while it correctly identified most of the anomalies, with only 352 false negatives. The ROC curve, with an AUC of 0.98, further confirms the model's ability to distinguish between normal data and anomalies. The table lists the top 10 anomalies by frequency, show-

casing the most significant deviations identified by the GMM. It highlights the most significant differences found by the model. The result is demonstrated in the 4.40.



Index	GMM Score	Frequency
166877.0	-358.114860100323	29779629.8439214
89795.0	-326.74746100440376	28425246.43907105
129530.0	-265.80245462548936	25588541.197182555
23256.0	-205.89782848054338	22451724.239174314
28765.0	-186.4020305530039	21331059.488607444
149595.0	-138.33014184877752	18274854.662103023
151214.0	-124.57841744094117	17301039.82104746
153228.0	-124.4141453247841	17289073.999064405
149324.0	-115.5142459295125	16627868.40490043
153125.0	-110.03664162767075	16207453.621870618

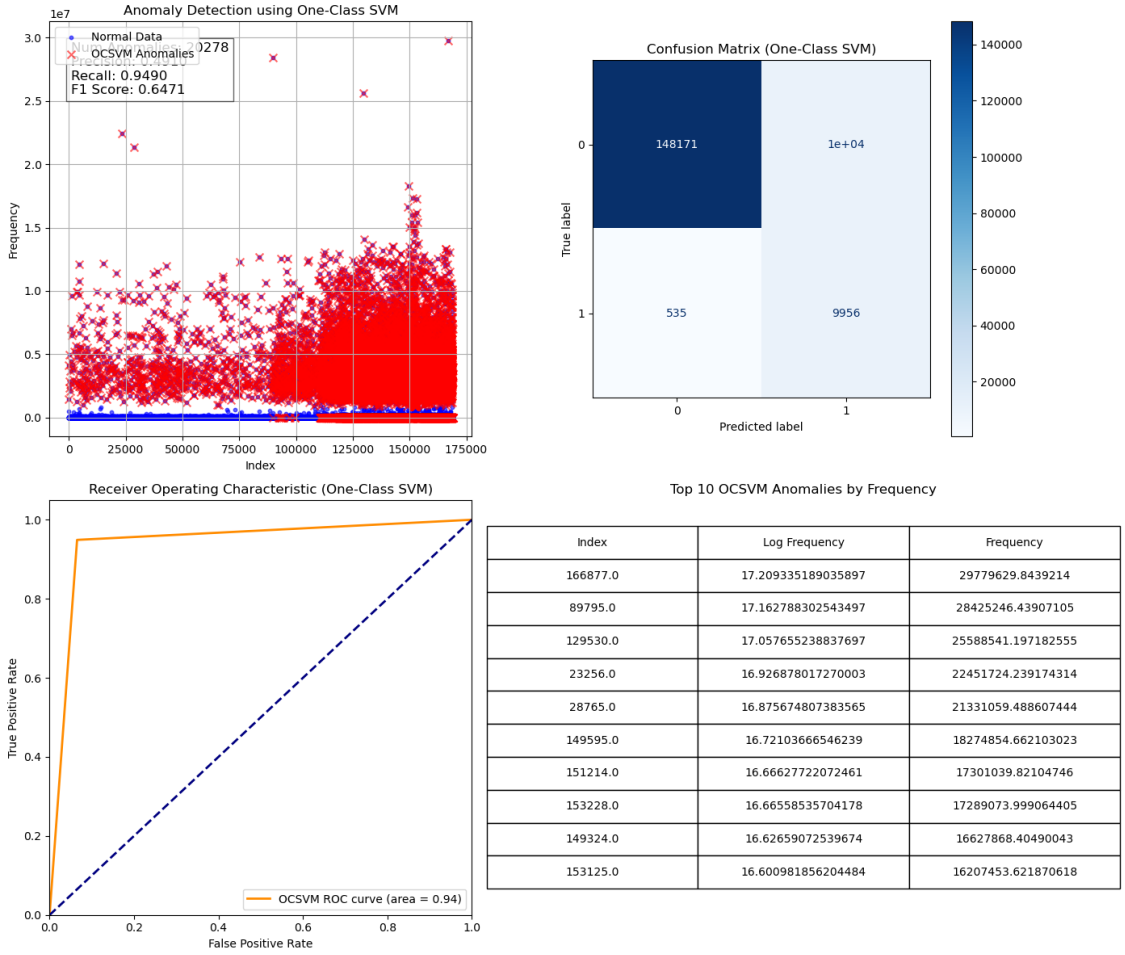
Figure 4.40: result of GMM for shielded dataset

4.5.1.4 One-Class SVM With Cross-Validation for Shielded Dataset

The One-Class SVM model was applied to detect anomalies in the dataset, identifying 20,278 anomalies, indicated by red points in the scatter plot. The model achieved a high recall of 0.9490, meaning it correctly identified 94.90% of the actual anomalies. However, the F1 score was 0.6471, reflecting a moderate balance between precision and recall due to some false positives and false negatives.

The confusion matrix shows that the model accurately classified 148,171 normal instances but also resulted in 10,004 false positives and 535 false negatives, indicating areas for improve-

ment. The ROC curve, with an AUC of 0.94, demonstrates strong performance in distinguishing between normal data and anomalies. The table lists the top 10 anomalies by frequency, highlighting the most significant deviations detected by the One-Class SVM model. The result is demonstrated in the 4.41.



Index	Log Frequency	Frequency
166877.0	17.209335189035897	29779629.8439214
89795.0	17.162788302543497	28425246.43907105
129530.0	17.057655238837697	25588541.197182555
23256.0	16.926878017270003	22451724.239174314
28765.0	16.875674807383565	21331059.488607444
149595.0	16.72103666546239	18274854.662103023
151214.0	16.66627722072461	17301039.82104746
153228.0	16.66558535704178	17289073.999064405
149324.0	16.62659072539674	16627868.40490043
153125.0	16.600981856204484	16207453.621870618

Figure 4.41: result of One-Class SVM with cross-validation for shielded dataset

4.5.1.5 Conclusion

In conclusion, the analysis using various anomaly detection techniques demonstrated varying degrees of success in identifying anomalies within the shielded dataset. The Isolation Forest and GMM methods exhibited high precision, recall, and F1 scores, suggesting their robust ability to detect anomalies accurately with minimal false positives and negatives. On the other hand, the LOF and One-Class SVM models showed moderate effectiveness, with a notable number

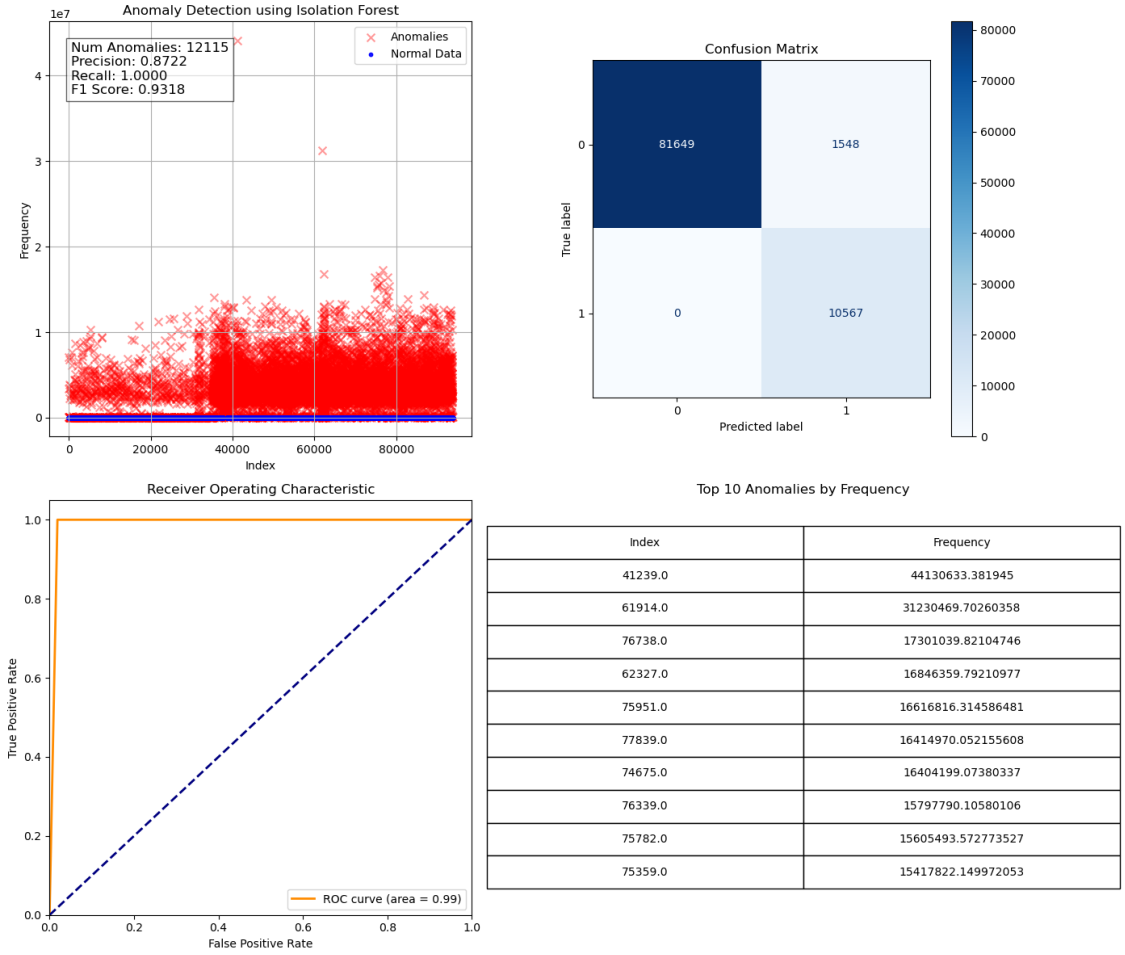
of false positives and false negatives, indicating challenges in distinguishing between normal data and anomalies.

4.5.2 Machine Learning Approaches for Analyzing Short Cable Data

4.5.2.1 Isolation Forest for Short Cable Dataset

The Isolation Forest algorithm was applied to detect anomalies in a dataset of frequency values, identifying 12,115 anomalies. The model achieved a precision of 87.22% of the detected anomalies were true anomalies, with a minimal number of false positives. The recall was perfect at 1.0000, meaning all anomalies in the dataset were successfully detected without any misses. The F1 score was 0.9318.

The confusion matrix provides further insight, showing that out of 93,364 instances, 81,649 normal data points were correctly identified, while 1,548 normal instances were incorrectly labeled as anomalies. The ROC curve was 0.99, which highlights the model's excellent ability. figure 4.47 show these results.



Index	Frequency
41239.0	44130633.381945
61914.0	31230469.70260358
76738.0	17301039.82104746
62327.0	16846359.79210977
75951.0	16616816.314586481
77839.0	16414970.052155608
74675.0	16404199.07380337
76339.0	15797790.10580106
75782.0	15605493.572773527
75359.0	15417822.149972053

Figure 4.42: Result Of Isolation Forest for Short Cable Dataset

Initially, I applied the Isolation Forest model to detect anomalies in the frequency domain, identifying significant deviations that suggested potential noise interference. Following this, I again applied the Isolation Forest model, focusing on detecting anomalies in the event counts and time series data. The results of this second analysis, figure 4.43, 4.2are presented below. The red points in the figure represent the anomalies identified by the Isolation Forest model, which are instances where the event count deviated significantly from the expected pattern.

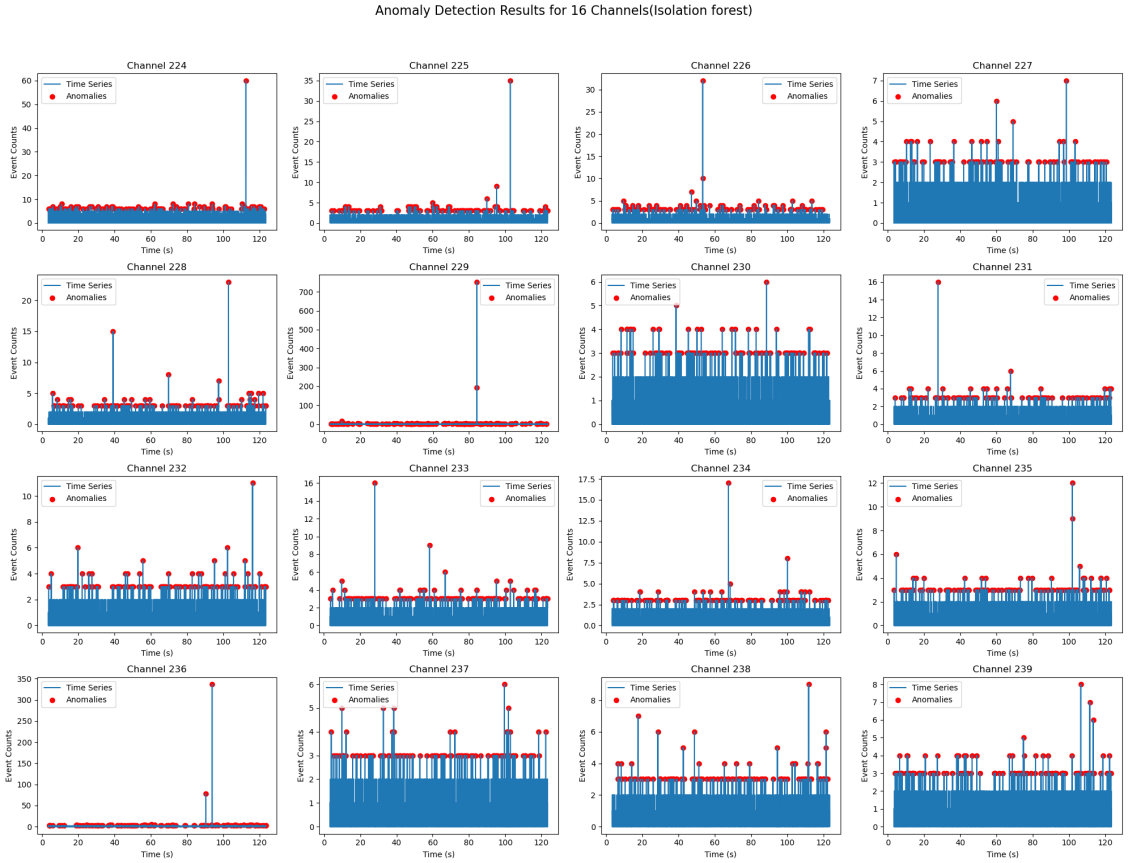


Figure 4.43: Anomaly Detection Results for 16 Channels Using The Isolation Forest Model(Short Cable Dataset)

Channel	Number of Anomalies
224	66
225	49
226	70
227	67
228	63
229	65
230	63
231	52
232	51
233	64
234	60
235	58
236	56
237	51
238	51
239	54

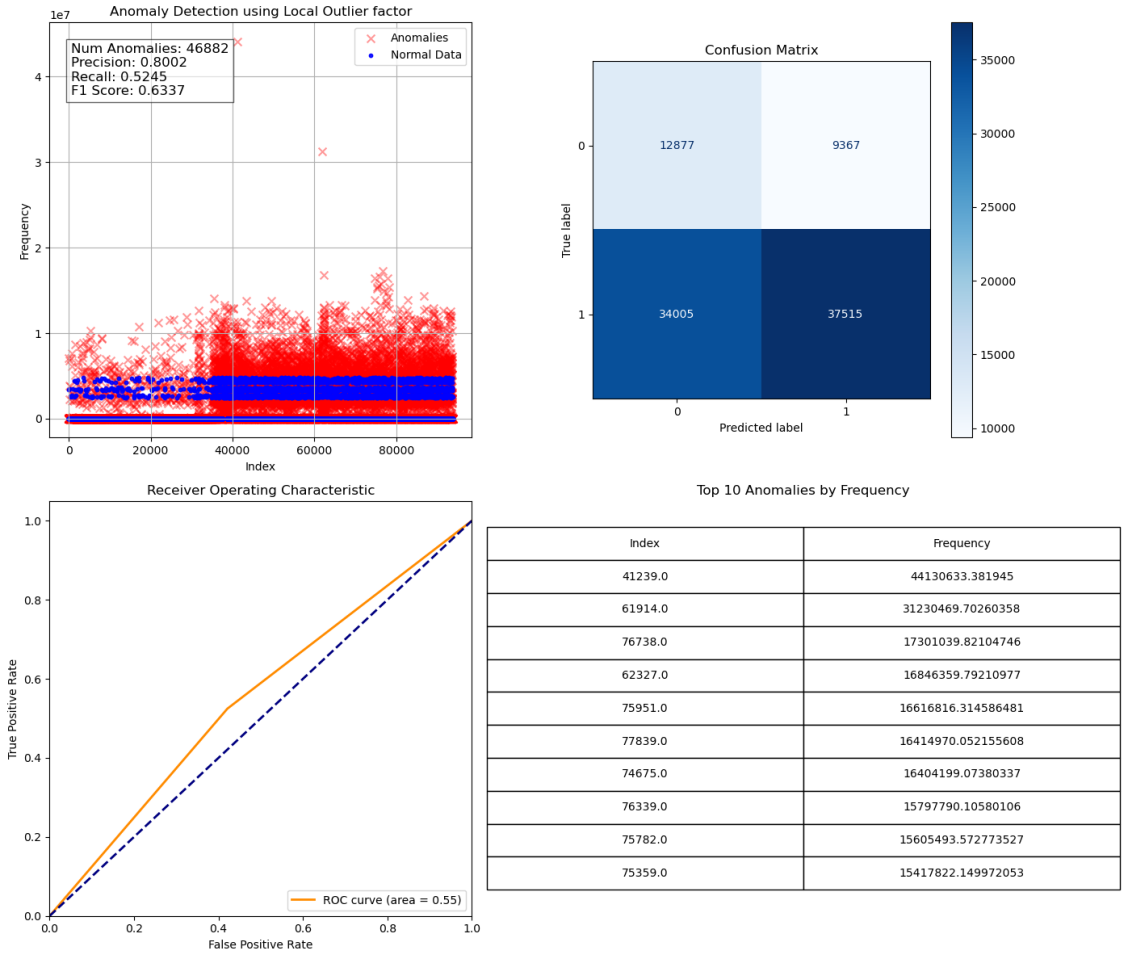
Table 4.2: Number of Anomalies Detected in Each Channel

These detected anomalies are consistent with the last section's results from other time series analysis methods. This alignment suggests that the channels showing more anomalies are likely more affected by noise and interference. The table highlights the number of anomalies detected in each channel, confirming that the patterns identified by the Isolation Forest model agree with the results of multi-faceted time series analysis.

4.5.2.2 Local Outlier Factor(LOF) for Short Cable Dataset

Using the Local Outlier Factor (LOF) algorithm, 46,882 anomalies were identified in the frequency dataset. The model achieved a precision of 0.8002, showing that 80.02% of detected anomalies were correct. However, the recall was lower at 0.5245, meaning it only identified 52.45% of all true anomalies, resulting in an F1 score of 0.6337. The confusion matrix shows many false negatives (34,005) and false positives (9,367). The ROC curve, with an AUC of 0.55, suggests limited effectiveness in distinguishing between normal data and anomalies. The top 10 anomalies by frequency highlight the most significant outliers detected. figure4.44

shows the consequence of this algorithm.



Index	Frequency
41239.0	44130633.381945
61914.0	31230469.70260358
76738.0	17301039.82104746
62327.0	16846359.79210977
75951.0	16616816.314586481
77839.0	16414970.052155608
74675.0	16404199.07380337
76339.0	15797790.10580106
75782.0	15605493.572773527
75359.0	15417822.149972053

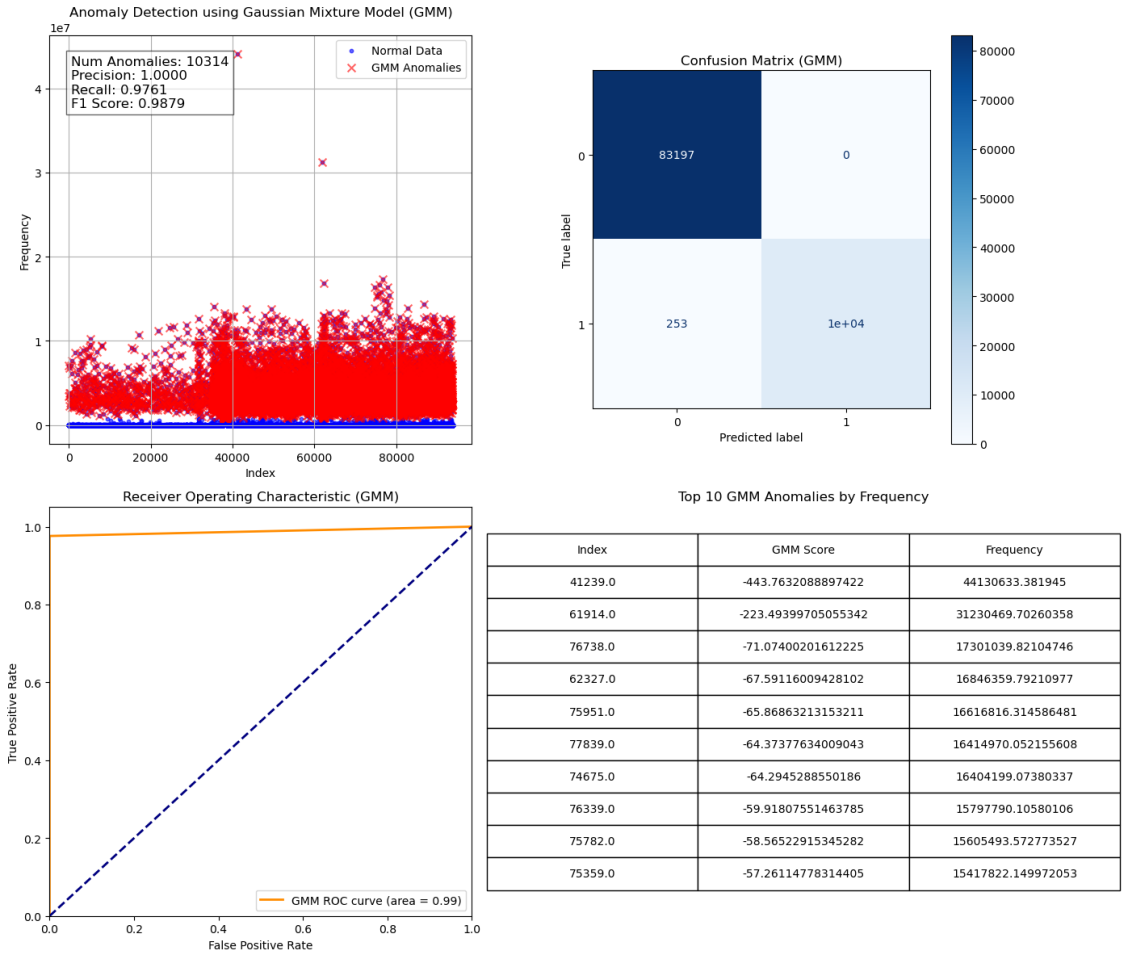
Figure 4.44: Local Outlier Factor(LOF)for Short Cable Dataset

4.5.2.3 Gaussian Mixture Model (GMM) for Short Cable Dataset

The Gaussian Mixture Model (GMM) identified 10,314 anomalies in the dataset, as shown by the red points in the scatter plot. The model achieved a perfect precision of 1.0000, meaning all anomalies detected were true positives, with no false positives. It also had a high recall of 0.9761, indicating it successfully identified 97.61% of all actual anomalies. The F1 score of 0.9879 reflects a strong balance between precision and recall, showcasing the model's overall effectiveness.

The confusion matrix reveals that 83,197 normal data points were correctly identified out of all instances, and no normal instances were incorrectly classified as anomalies. However, 253

anomalies were missed. The ROC curve, with an AUC of 0.99, demonstrates the model's ability to distinguish between normal data and anomalies. This performance shows the GMM's robustness in accurately detecting anomalies in the dataset. Figure 4.45 demonstrates these results.



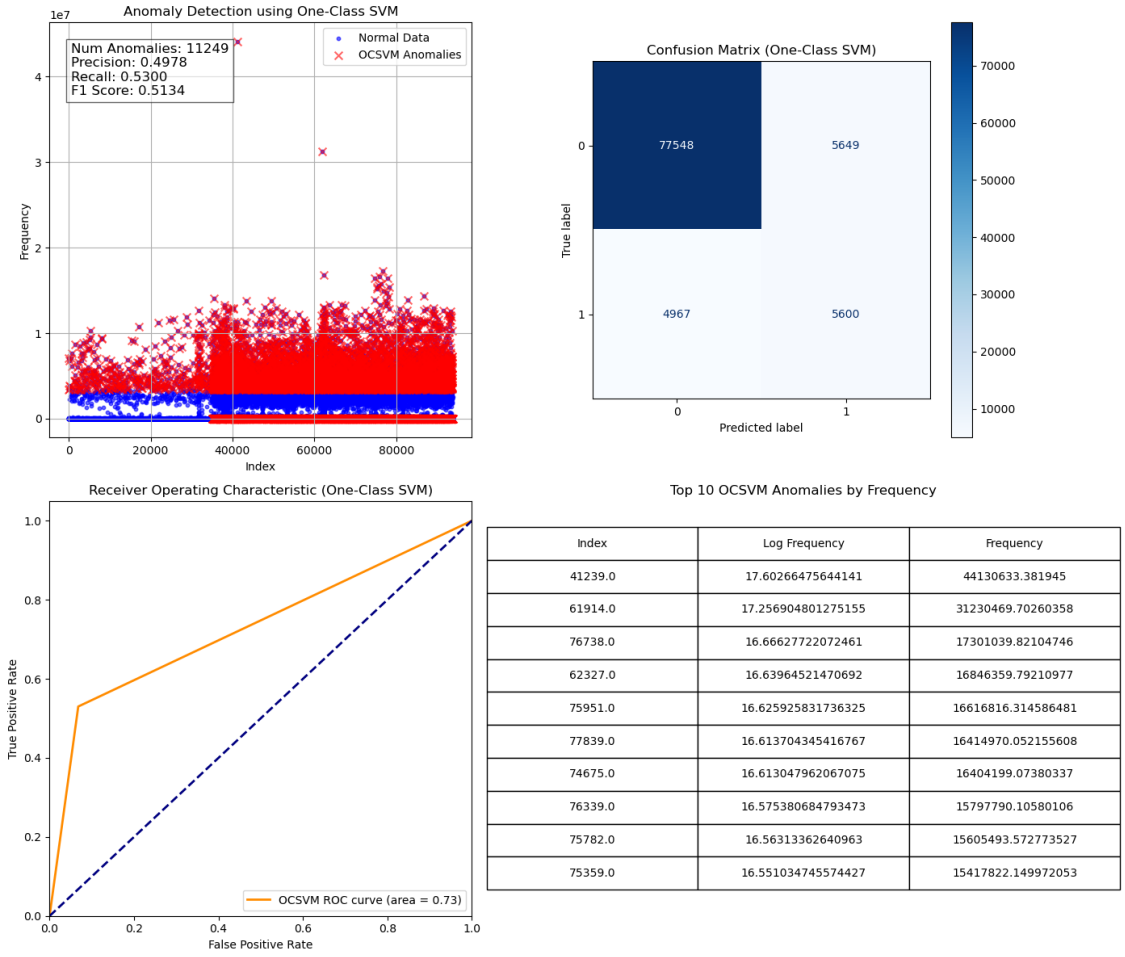
Index	GMM Score	Frequency
41239.0	-443.7632088897422	44130633.381945
61914.0	-223.49399705055342	31230469.70260358
76738.0	-71.07400201612225	17301039.82104746
62327.0	-67.59116009428102	16846359.79210977
75951.0	-65.86863213153211	16616816.314586481
77839.0	-64.37377634009043	16414970.052155608
74675.0	-64.2945288550186	16404199.07380337
76339.0	-59.91807551463785	15797790.10580106
75782.0	-58.56522915345282	15605493.572773527
75359.0	-57.26114778314405	15417822.149972053

Figure 4.45: Result of Gaussian Mixture Model (GMM) for Short Cable Dataset

4.5.2.4 One-Class SVM for short Cable Dataset

The One-Class SVM model identified 11,249 anomalies in the dataset, as indicated by the red points in the scatter plot. The model achieved a precision of 0.4978, meaning that about 49.78% of the detected anomalies were true positives. The recall was 0.5300, indicating that the model correctly identified 53.00% of the actual anomalies present in the data. This led to an F1 score of 0.5134, reflecting a moderate balance between precision and recall.

The confusion matrix shows that 77,548 normal data points were correctly classified out of the total instances, but 5,649 normal instances were mistakenly labeled as anomalies. Additionally, the model missed 4,967 true anomalies. The ROC curve, with an AUC of 0.73, suggests the model has a fair ability to distinguish between normal data and anomalies. Figure 4.46demonstrates these results.



Index	Log Frequency	Frequency
41239.0	17.60266475644141	44130633.381945
61914.0	17.256904801275155	31230469.70260358
76738.0	16.66627722072461	17301039.82104746
62327.0	16.63964521470692	16846359.79210977
75951.0	16.625925831736325	16616816.314586481
77839.0	16.613704345416767	16414970.052155608
74675.0	16.613047962067075	16404199.07380337
76339.0	16.575380684793473	15797790.10580106
75782.0	16.56313362640963	15605493.572773527
75359.0	16.551034745574427	15417822.149972053

Figure 4.46: result of One-Class SVM for short cable dataset

4.5.2.5 Conclusion

In conclusion, applying various anomaly detection algorithms showed that the Isolation Forest algorithm demonstrated robust performance with high precision and perfect recall, effectively identifying all anomalies with minimal false positives. This highlights its suitability for

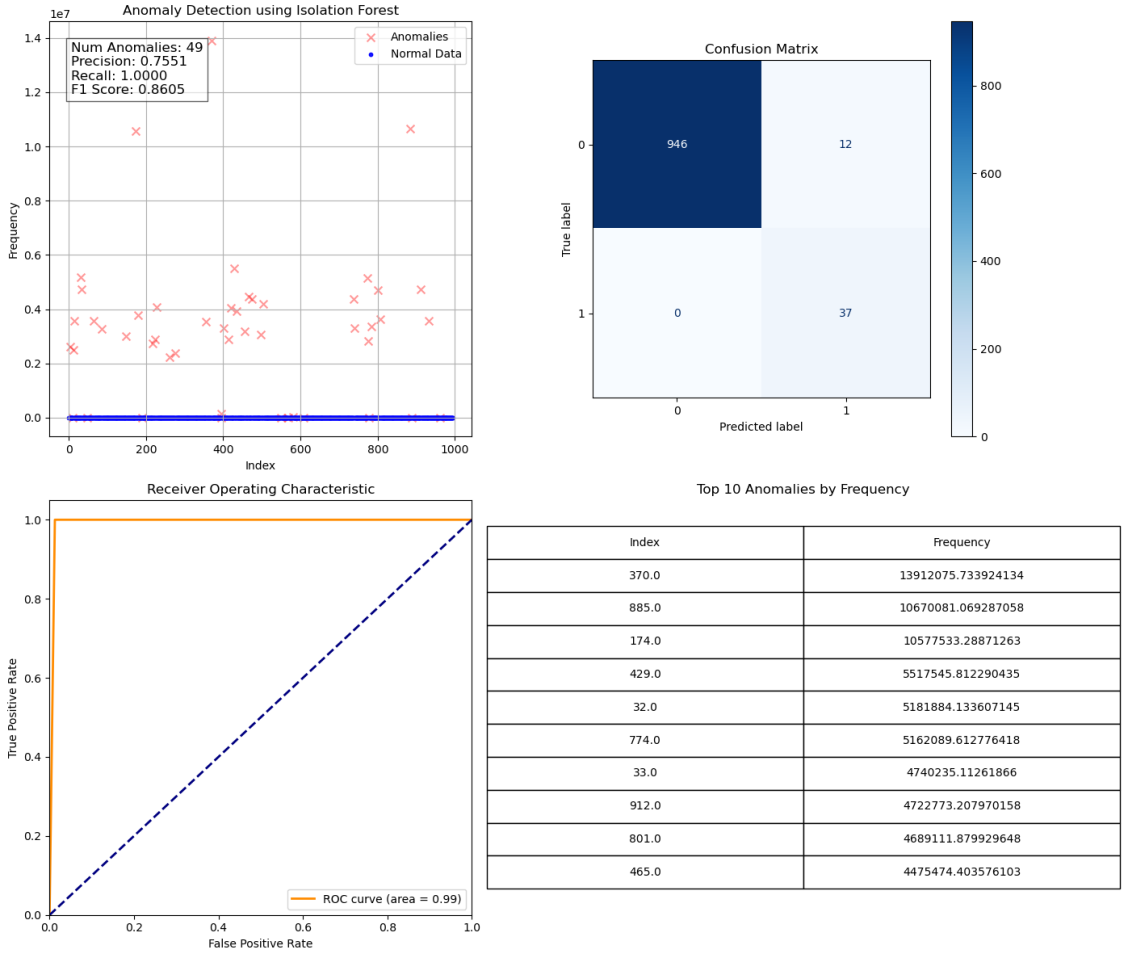
detecting significant deviations in frequency data. The Gaussian Mixture Model (GMM) also performed well, achieving perfect precision and high recall, making it effective for distinguishing normal data from anomalies. In contrast, the Local Outlier Factor (LOF) and One-Class SVM models showed moderate effectiveness, with LOF struggling with many false negatives and One-Class SVM presenting a balanced but lower accuracy. These results suggest that while Isolation Forest and GMM are reliable for anomaly detection in such datasets, LOF and One-Class SVM may not be the best fit for this particular dataset.

4.5.3 Machine Learning Approaches for Analyzing Low Threshold Unshielded Data

4.5.3.1 Result Of Isolation Forest for Low Threshold Unshielded Dataset

The results show the use of the Isolation Forest algorithm for finding anomalies in a dataset. The algorithm detected 49 anomalies with a precision of 0.7551 and a recall of 1.0000, resulting in an F1 score of 0.8605. This means the model was very effective at identifying true anomalies, with only one false positive, as seen in the confusion matrix.

The ROC curve has an area under the curve (AUC) of 0.99, indicating excellent performance with no compromise between detecting true positives and avoiding false positives. The top 10 most frequent anomalies are listed, with the highest frequency being 1282051905.9, showing these points are significantly different from normal data. Figure 4.47 illustrates these results.



Index	Frequency
370.0	13912075.733924134
885.0	10670081.069287058
174.0	10577533.28871263
429.0	5517545.812290435
32.0	5181884.133607145
774.0	5162089.612776418
33.0	4740235.11261866
912.0	4722773.207970158
801.0	4689111.879929648
465.0	4475474.403576103

Figure 4.47: Result Of Isolation Forest for Low Threshold Unshielded Dataset

I reapplied the Isolation Forest model to detect event counts and time series data anomalies. The results of this secondary analysis are shown in Figure 4.48 and Table 4.3. The red points in the figure show the anomalies detected by the Isolation Forest model. They indicate instances where the event counts were significantly different from expected patterns.

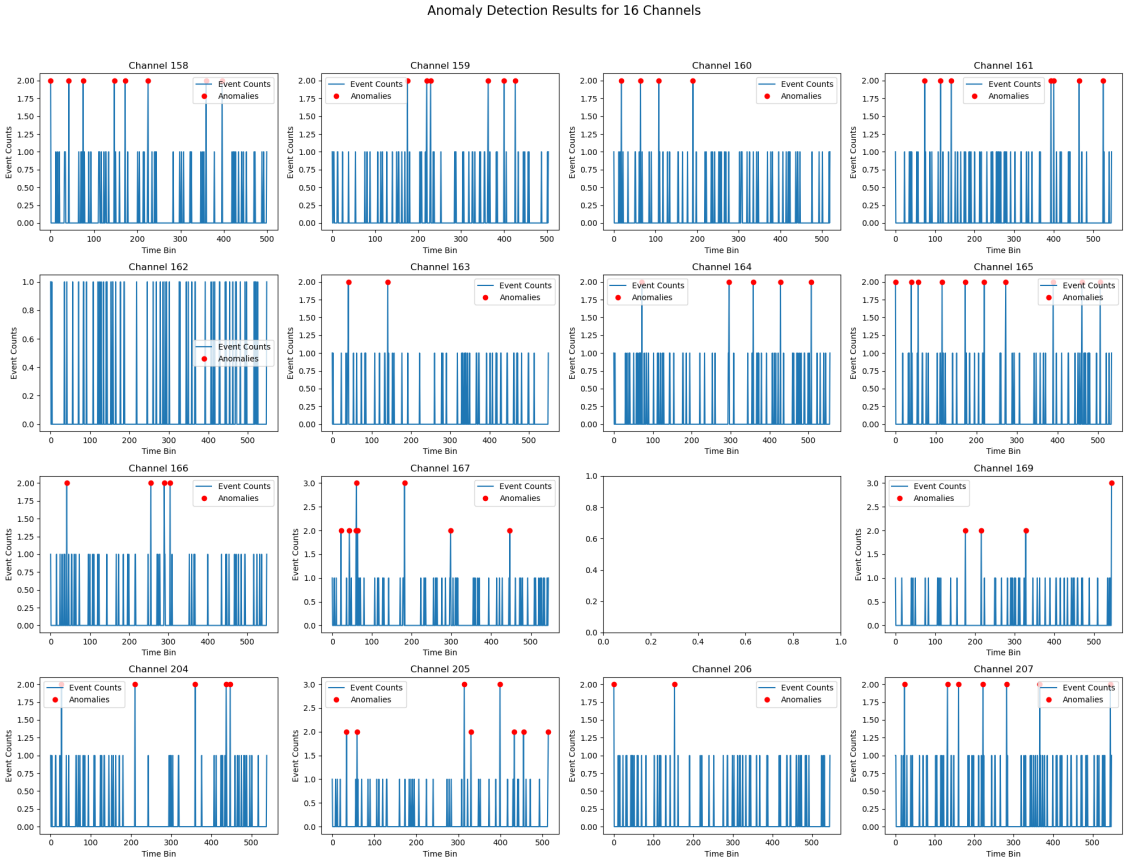


Figure 4.48: Anomaly Detection Results for 16 Channels Using The Isolation Forest Model(Unshielded Low Threshold Dataset)

Channel	Number of Anomalies
158	8
159	8
160	4
161	9
162	0
163	2
164	5
165	10
166	4
167	15
168	
169	12
204	5
205	6
206	2
207	7

Table 4.3: Number of Anomalies Detected (Isolation Forest On The Time Series Of The Low Threshold Unshielded Dataset)in Each Channel

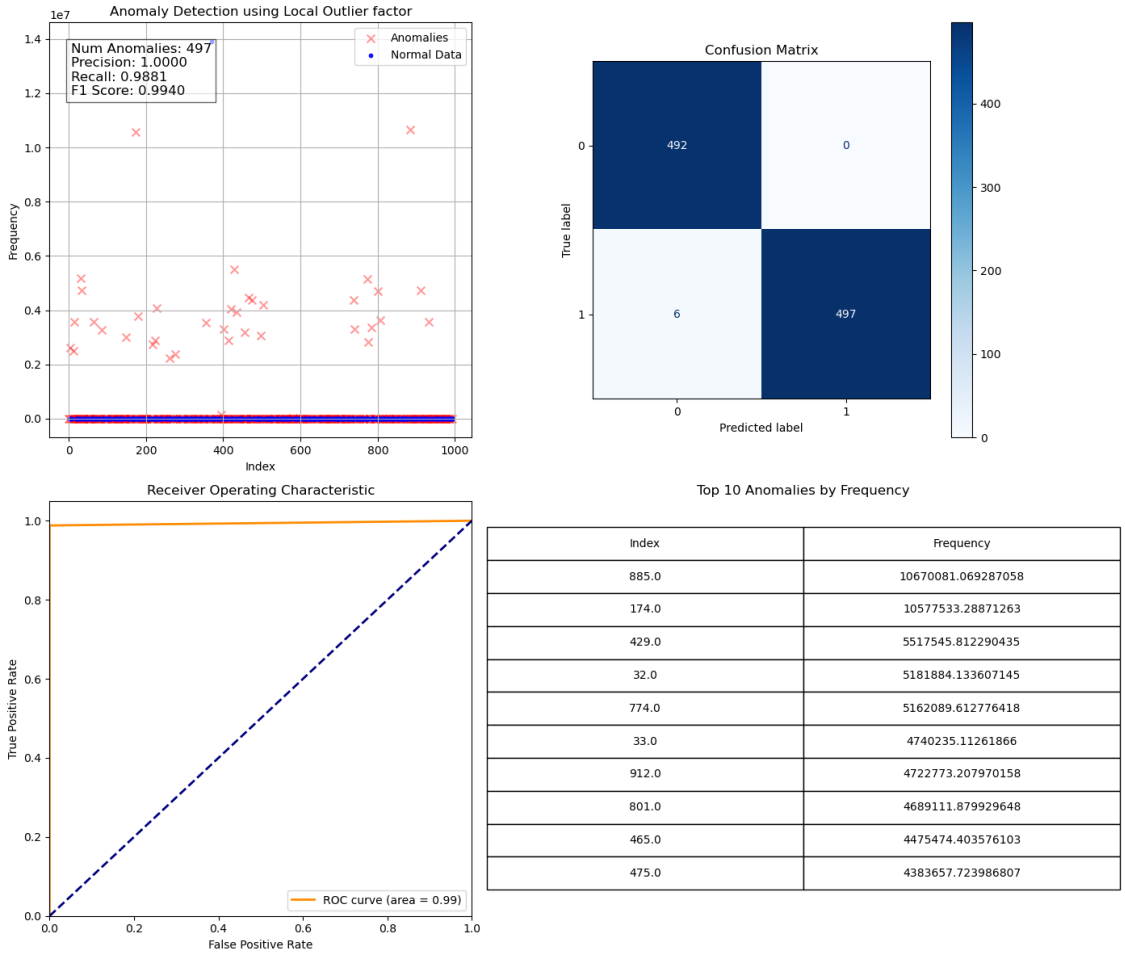
These detected anomalies are consistent with the last section's results from other time series analysis methods. This alignment suggests that the channels showing more anomalies are likely more affected by noise and interference. The table highlights the number of anomalies detected in each channel, confirming that the patterns identified by the Isolation Forest model agree with the multi-faceted time series analysis results, thereby validating the accuracy of these findings.

4.5.3.2 Local Outlier Factor(LOF) for Low Threshold Unshielded Dataset

The Local Outlier Factor (LOF) algorithm detected 497 anomalies in the dataset. The model achieved a perfect precision of 1.0000, indicating that all identified anomalies were true positives with no false positives. The recall was 0.9881, meaning the model correctly identified 98.81% of the actual anomalies present in the data, resulting in a high F1 score of 0.9940.

Figure 4.49 describes that the confusion matrix reveals that the model accurately classified 492 normal data points and 497 anomalies, with only six false negatives, demonstrating its

strong ability to detect anomalies effectively. The Receiver Operating Characteristic (ROC) curve, with an area under the curve (AUC) of 0.99, further confirms the model's excellent performance distinguishing between normal and abnormal data points.



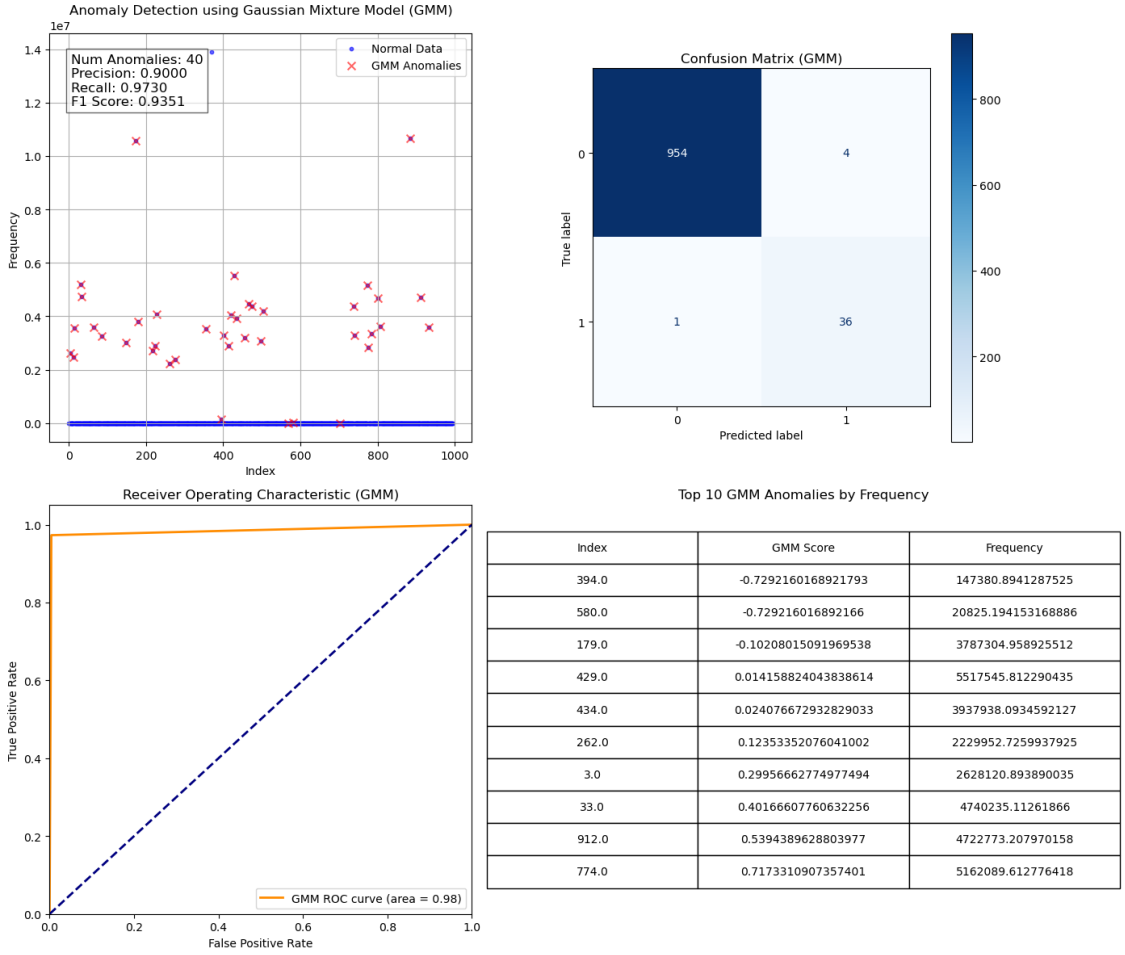
Index	Frequency
885.0	10670081.069287058
174.0	10577533.28871263
429.0	5517545.812290435
32.0	5181884.133607145
774.0	5162089.612776418
33.0	4740235.11261866
912.0	4722773.207970158
801.0	4689111.879929648
465.0	4475474.403576103
475.0	4383657.723986807

Figure 4.49: Result of Local Outlier Factor(LOF) for Low Threshold Unshielded Dataset

4.5.3.3 Gaussian Mixture Model (GMM) for low threshold unshielded dataset

Figure 4.50 describes The Gaussian Mixture Model (GMM), which identified 40 anomalies in the dataset with a precision of 0.9000 and a recall of 0.9730, leading to an F1 score of 0.9351. The confusion matrix shows that the model accurately detected 36 true anomalies with only four false positives and one false negative. The ROC curve, with an AUC of 0.98, highlights the model's ability to differentiate between normal and abnormal data. The table lists the top 10

anomalies by frequency, showcasing significant deviations, with the highest frequency anomaly being 147380.8941287525.

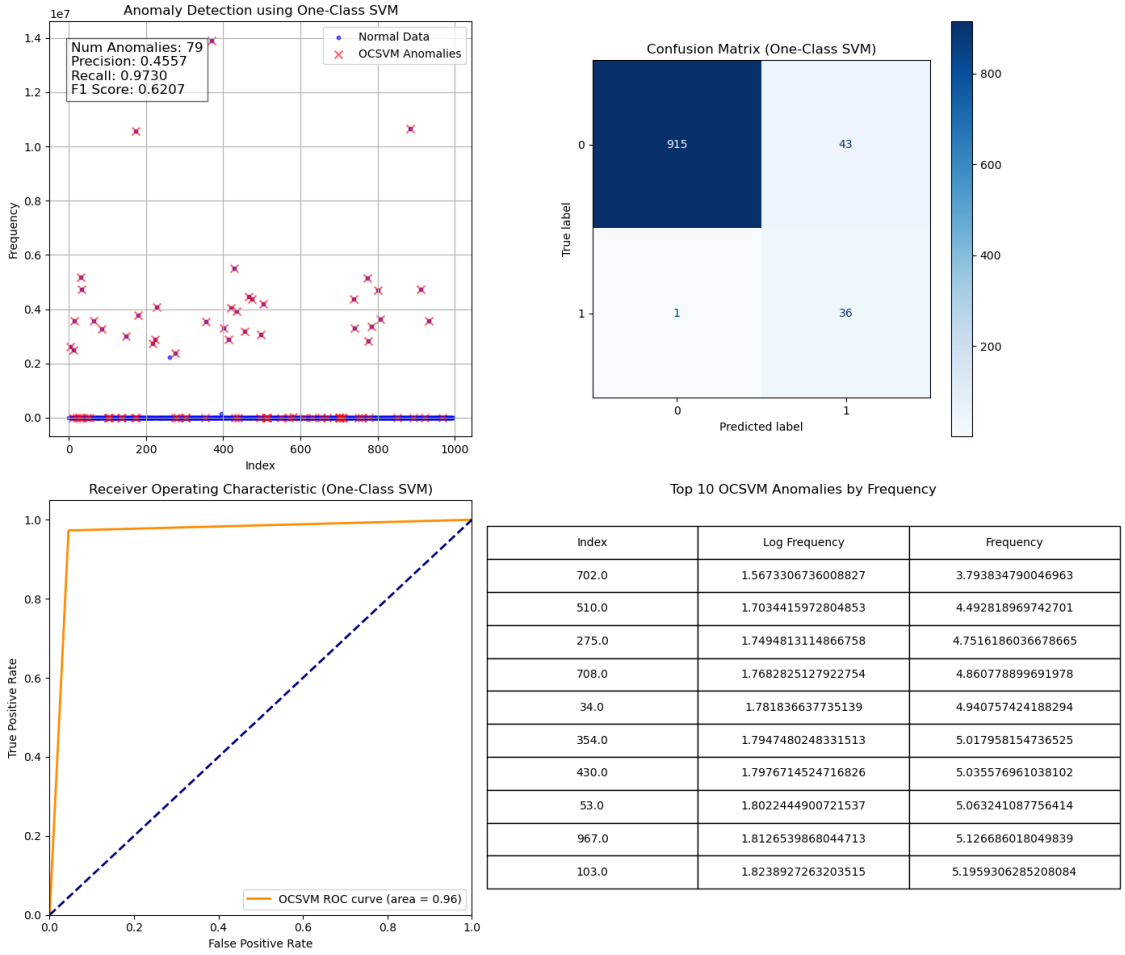


Index	GMM Score	Frequency
394.0	-0.7292160168921793	147380.8941287525
580.0	-0.729216016892166	20825.194153168886
179.0	-0.10208015091969538	3787304.958925512
429.0	0.014158824043838614	5517545.812290435
434.0	0.024076672932829033	3937938.0934592127
262.0	0.12353352076041002	2229952.7259937925
3.0	0.29956662774977494	2628120.893890035
33.0	0.40166607760632256	4740235.11261866
912.0	0.5394389628803977	4722773.207970158
774.0	0.7173310907357401	5162089.612776418

Figure 4.50: Result Of Gaussian Mixture Model (GMM) Low Threshold Unshielded Dataset

4.5.3.4 One-Class SVM for Low Threshold Unshielded Dataset

Based on the results,figure4.51, The One-Class SVM algorithm detected 79 anomalies in the dataset, achieving a precision of 0.4557 and a recall of 0.9730, which results in an F1 score of 0.6207. The confusion matrix reveals that the model correctly identified 36 anomalies and flagged 43 normal instances as anomalies, indicating a moderate rate of false positives. The ROC curve, with an AUC of 0.96, demonstrates the model's strong performance distinguishing between normal and abnormal data points.



Index	Log Frequency	Frequency
702.0	1.5673306736008827	3.793834790046963
510.0	1.7034415972804853	4.492818969742701
275.0	1.7494813114866758	4.7516186036678665
708.0	1.7682825127922754	4.860778899691978
34.0	1.781836637735139	4.940757424188294
354.0	1.7947480248331513	5.017958154736525
430.0	1.7976714524716826	5.035576961038102
53.0	1.8022444900721537	5.063241087756414
967.0	1.8126539868044713	5.126686018049839
103.0	1.8238927263203515	5.1959306285208084

Figure 4.51: Result Of One-Class SVM for Low Threshold Unshielded Dataset

4.5.3.5 Conclusion

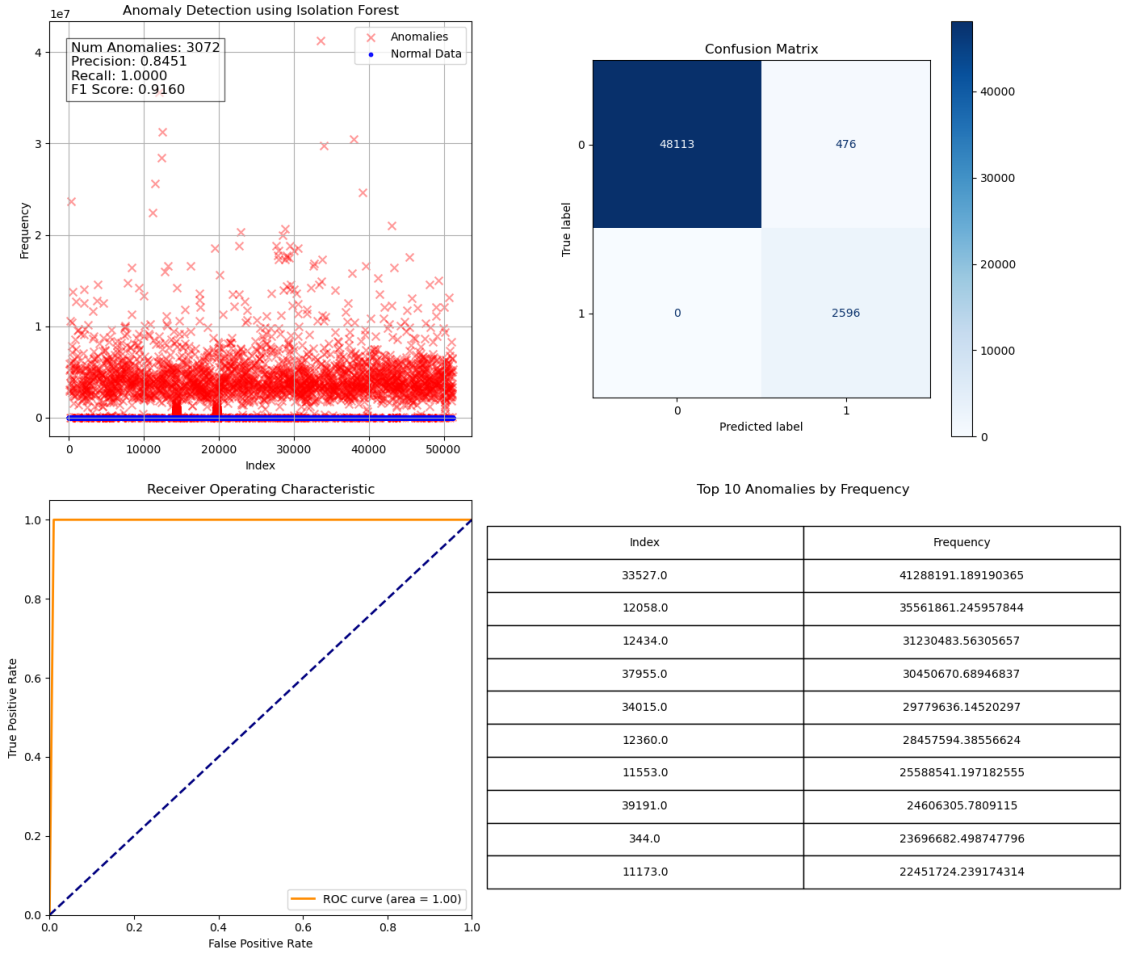
In conclusion, the application of various anomaly detection algorithms—Isolation Forest, Local Outlier Factor (LOF), Gaussian Mixture Model (GMM), and One-Class SVM—on the low threshold unshielded dataset revealed differing levels of effectiveness in identifying anomalies. The Isolation Forest and GMM models showed high precision and recall, with F1 scores of 0.8605 and 0.9351, respectively, indicating their robust performance in detecting true anomalies with minimal false positives. While achieving a high recall of 0.9730, the One-Class SVM had a lower precision, resulting in a moderate F1 score of 0.6207, suggesting a higher rate of false positives. Although consistent with other time series analysis methods, the LOF model confirmed that certain channels are more prone to noise and interference. Overall, these mod-

els collectively validate the presence of important anomalies in the dataset.

4.5.4 Machine Learning Approaches for Analyzing High Threshold Unshielded

4.5.4.1 Isolation Forest for High Threshold Unshielded Dataset

The Isolation Forest model was used to detect anomalies, successfully identifying 3,072 anomalies with a precision of 0.8451, a perfect recall of 1.0000, and an F1 score of 0.9160. These results indicate that the model effectively detected all true anomalies without missing any (high recall) while maintaining high accuracy (precision) in minimizing false positives. The confusion matrix supports these findings, showing 2,596 true anomalies correctly identified without any false negatives and 476 normal data points incorrectly classified as anomalies. The ROC curve, with an AUC of 1.00, demonstrates the model's exceptional ability to distinguish between normal and abnormal data points. .figure 4.52 showS the results.



Index	Frequency
33527.0	41288191.189190365
12058.0	35561861.245957844
12434.0	31230483.56305657
37955.0	30450670.68946837
34015.0	29779636.14520297
12360.0	28457594.38556624
11553.0	25588541.197182555
39191.0	24606305.7809115
344.0	23696682.498747796
11173.0	22451724.239174314

Figure 4.52: Result Of Isolation Forest for High Threshold Unshielded Dataset

I applied the (Iforest) to the dataset to detect anomalies in the frequency domain and time series data. Initially, the Iforest identified significant deviations that suggested potential noise interference in the frequency domain. Following this, the model was used again to detect anomalies in the event counts over time. The results of this secondary analysis are shown in Figure4.53 and Table 4.4. The red points in the figure show the anomalies detected by the Isolation Forest model, representing instances where the event counts significantly deviated from the expected patterns.

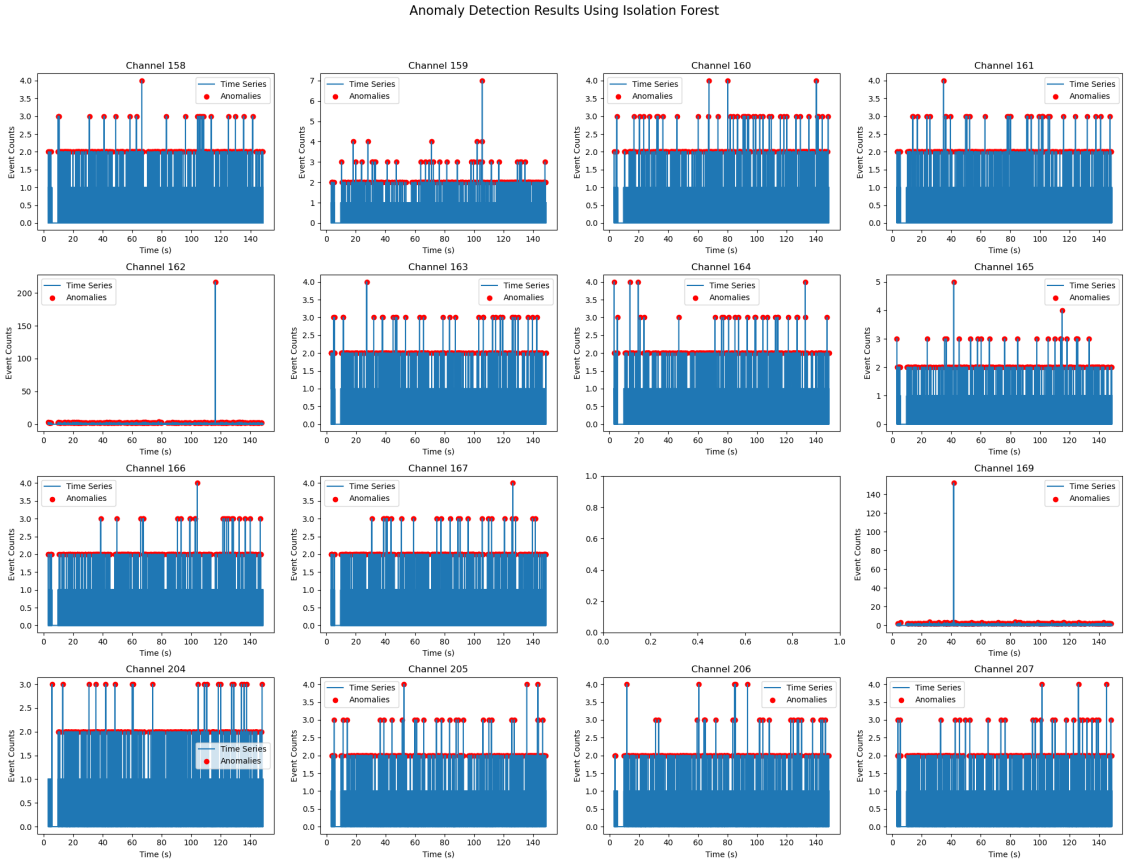


Figure 4.53: Anomaly detection results for 16 channels using the Isolation Forest Model(Unshielded High Threshold Dataset)

Channel	Number of Anomalies
158	170
159	169
160	170
161	169
162	164
163	165
164	166
165	168
166	167
167	169
168	
169	172
204	166
205	168
206	167
207	172

Table 4.4: Number of Anomalies Detected (Isolation forest on the time series of the low threshold unshielded dataset)in Each Channel

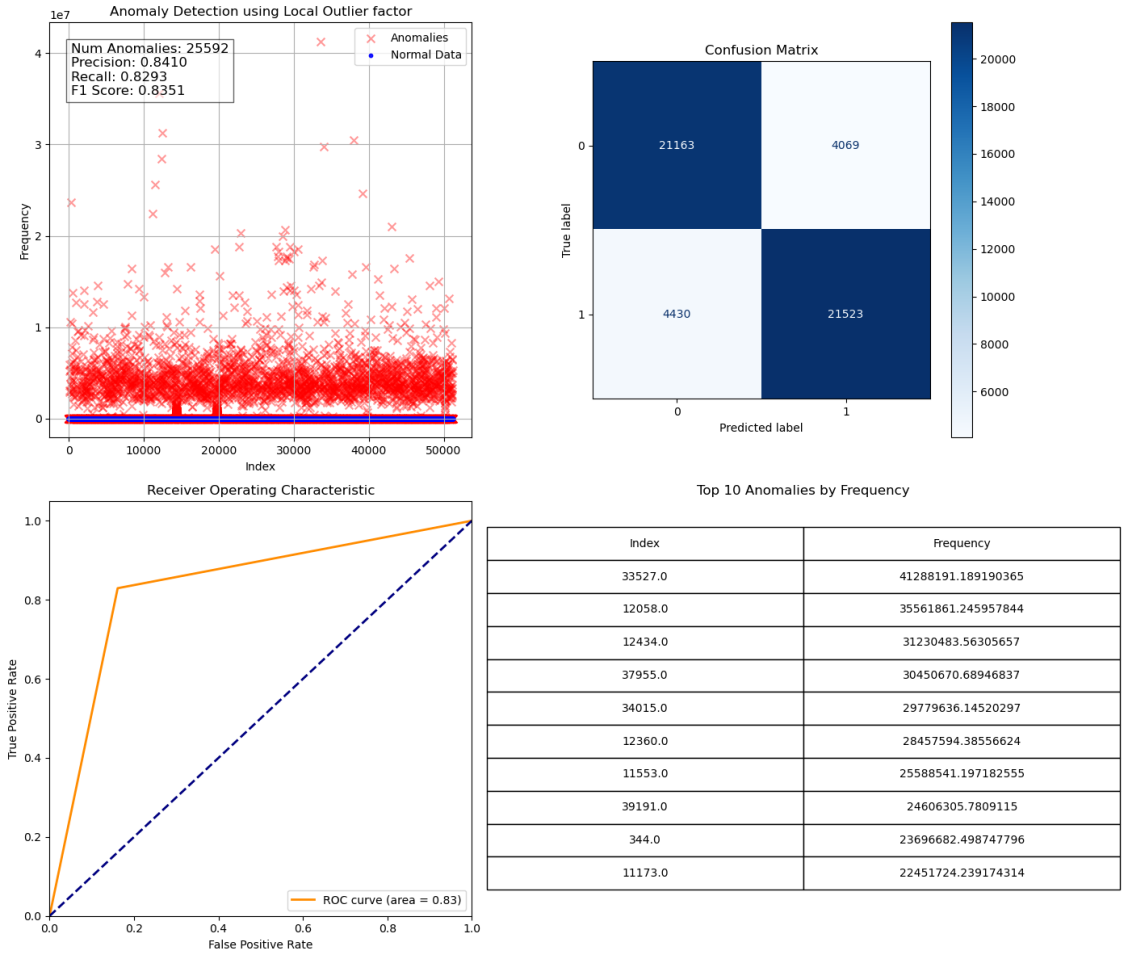
The anomalies we found match the results of other methods we used to analyze time series data. Shows that most of the channels are in a specific frequency range and without noise.

4.5.4.2 Local Outlier Factor(LOF) for High Threshold Unshielded Dataset

In the top left plot, figure 4.54 anomalies identified by the model are marked in red, showing significant deviations from the normal data patterns. The model achieved a precision of 0.8410 and a recall of 0.8293, indicating good detection accuracy and a substantial ability to identify actual anomalies correctly. The F1 score of 0.8351 reflects a strong balance between precision and recall, emphasizing the model's overall effectiveness.

The confusion matrix provides a detailed breakdown of the model's performance, showing that it correctly classified 21,163 normal data points while mistakenly labeling 4,069 normal instances as anomalies. Additionally, it accurately identified 21,523 anomalies, with 4,430 actual anomalies missed. The ROC curve, with an AUC of 0.83, suggests the model has a good

ability to distinguish between normal and abnormal data points.



Index	Frequency
33527.0	41288191.189190365
12058.0	35561861.245957844
12434.0	31230483.56305657
37955.0	30450670.68946837
34015.0	29779636.14520297
12360.0	28457594.38556624
11553.0	25588541.197182555
39191.0	24606305.7809115
344.0	23696682.498747796
11173.0	22451724.239174314

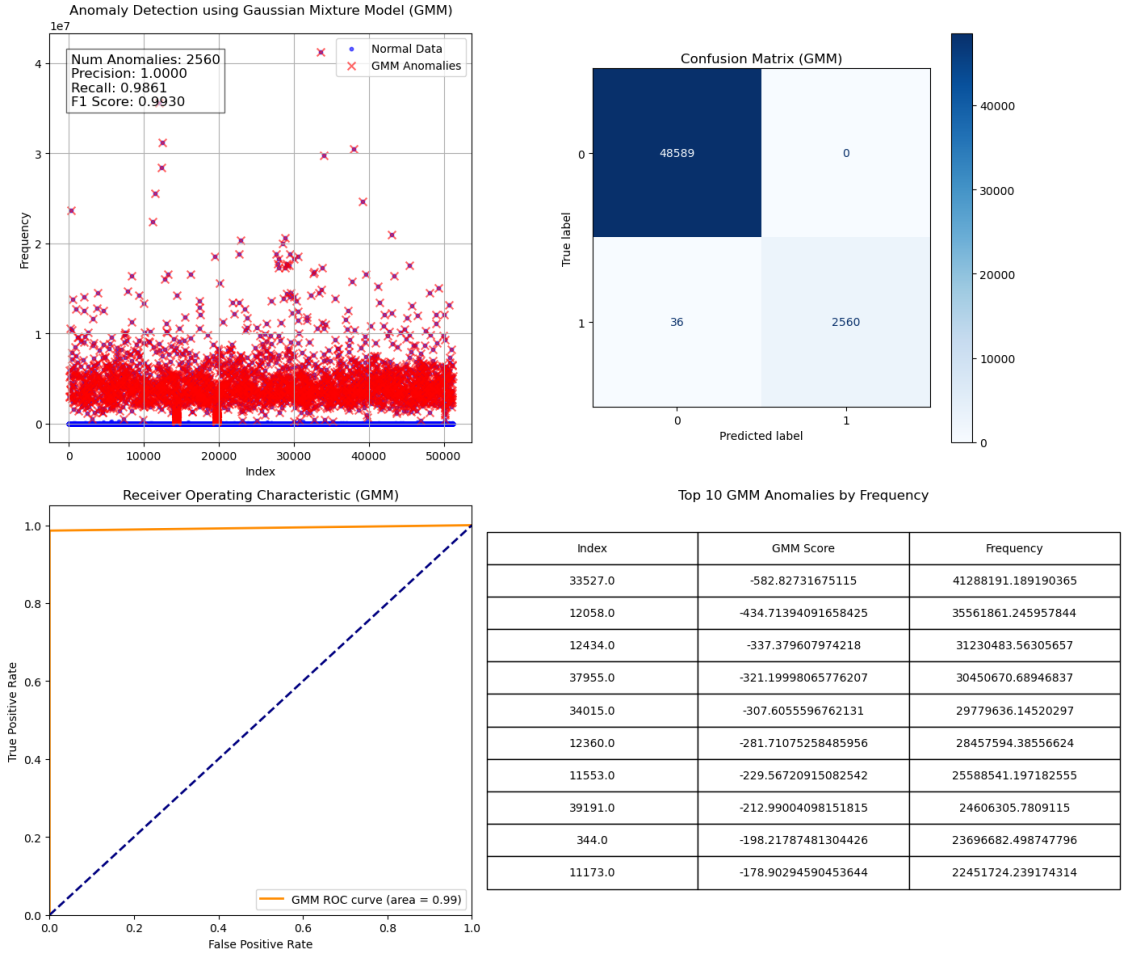
Figure 4.54: Result Of Local Outlier Factor for High Threshold Unshielded Dataset

4.5.4.3 Gaussian Mixture Model(GMM)

In the scatter plot, figure4.55, the anomalies detected by the model are marked in red, indicating deviations from normal data patterns. It demonstrated perfect precision at 1.0000, meaning all detected anomalies were true positives. It also achieved a high recall of 0.9861, capturing 98.61% of all actual anomalies. The F1 score, a balance between precision and recall, was notably high at 0.9930, emphasizing the model's strong performance in accurately identifying anomalies.

The confusion matrix offers a detailed view of the model's accuracy, showing that 48,589 normal data points were correctly identified, with no false positives, and that 2,560 anomalies

were correctly detected. However, 36 anomalies were missed, reflecting a minimal rate of false negatives. The ROC curve, featuring an AUC of 0.99, further validates the model's exceptional ability to distinguish between normal and abnormal instances. The table on the bottom right lists the top 10 anomalies by frequency, showing the most significant outliers in the dataset as identified by the GMM.



Index	GMM Score	Frequency
33527.0	-582.82731675115	41288191.189190365
12058.0	-434.71394091658425	35561861.245957844
12434.0	-337.379607974218	31230483.56305657
37955.0	-321.19998065776207	30450670.68946837
34015.0	-307.6055596762131	29779636.14520297
12360.0	-281.71075258485956	28457594.38556624
11553.0	-229.56720915082542	25588541.197182555
39191.0	-212.99004098151815	24606305.7809115
344.0	-198.21787481304426	23696682.498747796
11173.0	-178.90294590453644	22451724.239174314

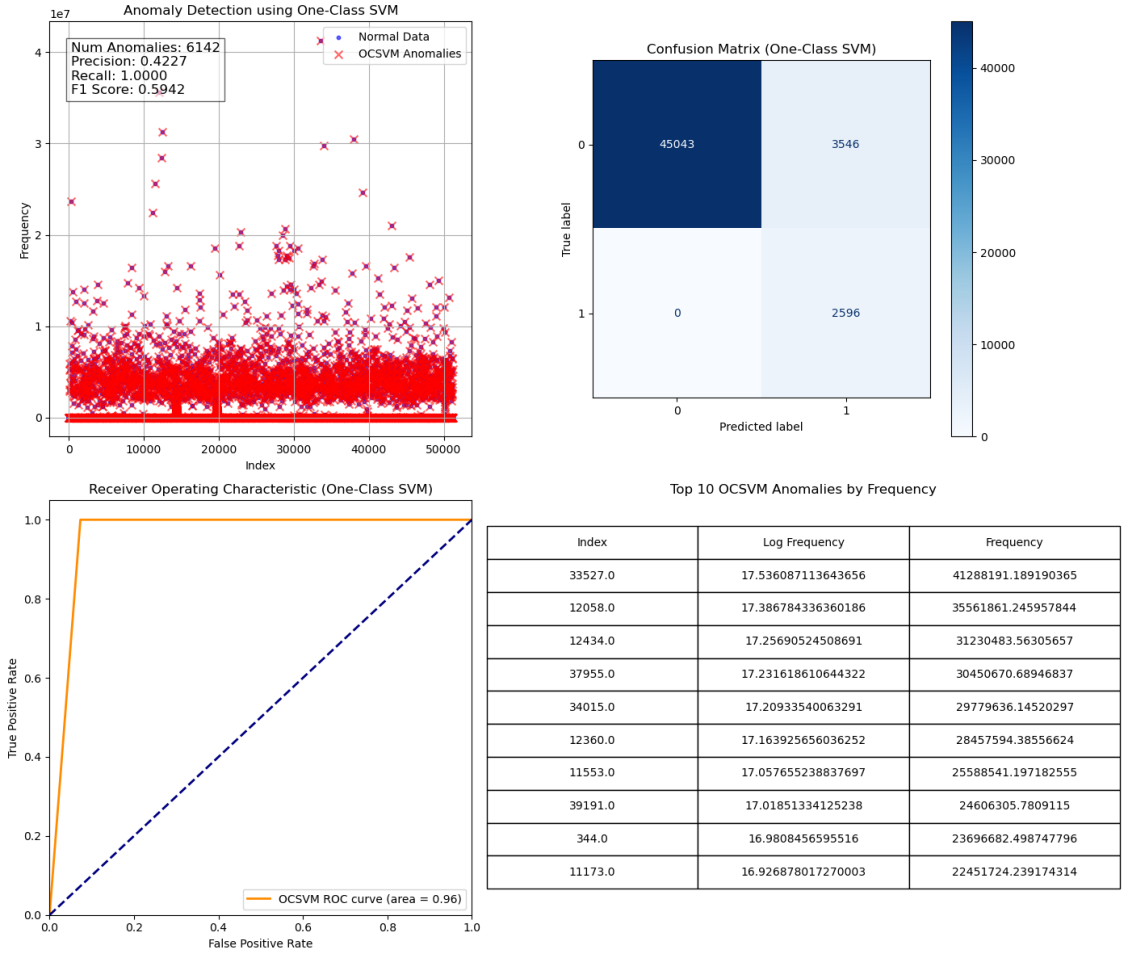
Figure 4.55: Result of Gaussian Mixture Model (GMM) High Threshold Unshielded Dataset

4.5.4.4 One-Class SVM for High Threshold Unshielded Dataset

The One-Class SVM model was used for anomaly detection, identifying 6,142 anomalies highlighted in red in the scatter plot 4.56. The model achieved a precision of 0.4227, indicating that approximately 42.27% of the detected anomalies were true positives. The recall was perfect at 1.0000, meaning the model successfully identified all actual anomalies. However, the preci-

sion score suggests many false positives, leading to an F1 score of 0.5942, reflecting a moderate balance between precision and recall.

The confusion matrix shows that 45,043 normal data points were correctly identified, while 3,546 normal instances were incorrectly classified as anomalies. The ROC curve, with an AUC of 0.96, demonstrates the model's ability to distinguish between normal and abnormal data points. The table on the bottom right lists the top 10 anomalies by frequency, emphasizing the most significant deviations detected by the One-Class SVM model.



Index	Log Frequency	Frequency
33527.0	17.536087113643656	41288191.189190365
12058.0	17.386784336360186	35561861.245957844
12434.0	17.25690524508691	31230483.56305657
37955.0	17.231618610644322	30450670.68946837
34015.0	17.20933540063291	29779636.14520297
12360.0	17.163925656036252	28457594.38556624
11553.0	17.057655238837697	25588541.197182555
39191.0	17.01851334125238	24606305.7809115
344.0	16.9808456595516	23696682.498747796
11173.0	16.926878017270003	22451724.239174314

Figure 4.56: result of One-Class SVM for high threshold unshielded dataset

4.5.4.5 Conclusion

In summary, the anomaly detection methods applied to the high threshold unshielded dataset yielded varying levels of effectiveness. The Isolation Forest and Gaussian Mixture Model (GMM)

performed exceptionally well, achieving high precision and recall and effectively distinguishing between normal and abnormal data points. The Local Outlier Factor (LOF) showed good performance but with a higher rate of false positives. While identifying all true anomalies, the One-Class SVM model had lower precision, indicating more false positives. Overall, the Isolation Forest and GMM were the most effective for accurate anomaly detection in this dataset.

4.6 Advanced Prediction with Deep Learning

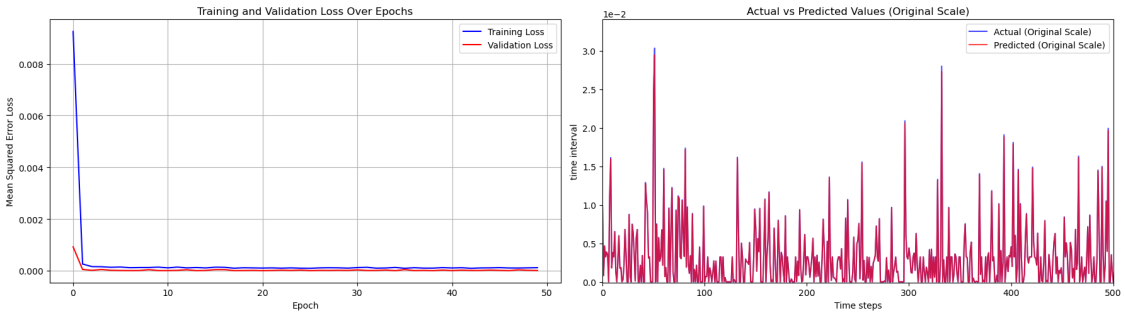
In this section, we describe implementing and evaluating a Long Short-Term Memory (LSTM) model used to predict time series data related to frequency measurements.

4.6.1 Deep Learning Model for Unshielded High Threshold Dataset

The LSTM model was built using the TensorFlow Keras library. The architecture consists of two LSTM layers with 128 and 64 units, respectively. Each LSTM layer is followed by a dropout layer with a dropout rate of 0.2 to prevent overfitting. The final output layer is a dense layer with a single neuron, which is suitable for the regression task. The model was compiled using the Adam optimizer and mean squared error as the loss function.

4.6.1.1 Model Training and Evaluation on Unshielded High Threshold Dataset

The LSTM model was trained for 50 epochs with a batch size of 32, using 80% of the training data for training and 20% for validation. The training and validation loss over epochs is depicted in Figure 4.57, showing a rapid convergence of both losses, indicating a well-trained model.



Test Loss: 0.000006, Test MAE: 0.000854, Test RMSE: 0.002360, Mean Absolute Error (Original Scale): 0.000033, Root Mean Square Error (Original Scale): 0.000092

Figure 4.57: Training and Validation Loss Over Epochs and Actual vs Predicted Value (Unshielded High Threshold Dataset)

The model was evaluated on the test dataset, achieving a test loss of 6×10^{-6}, mean absolute error (MAE) of 0.000854, and root mean square error (RMSE) of 0.002360 on the scaled data. These metrics demonstrate the model's high accuracy in predicting the time series data on a normalized scale.

When the predictions were inverse-transformed to the original scale, the model achieved a mean absolute error of 0.000033 and a root mean square error of 0.000092. These results are summarized in Table 4.5.

Metric	**Value**
Test Loss	6×10^{-6}
Test MAE (Scaled)	0.000854
Test RMSE (Scaled)	0.002360
Mean Absolute Error (Original Scale)	0.000033
Root Mean Square Error (Original Scale)	0.000092

Table 4.5: Evaluation Metrics for the LSTM Model (Unshielded High Threshold Dataset)

The actual versus predicted values on the original scale are shown in Figure 4.57, illustrating that the model's predictions closely follow the actual data points, further validating the model's effectiveness.

4.6.2 Deep Learning Model for Unshielded Low Threshold Dataset

The architecture comprises two LSTM layers, with 128 and 64 units, respectively. Each LSTM layer is followed by a dropout layer with a dropout rate of 0.2 to mitigate overfitting. The

final output layer is a dense layer containing a single neuron, suitable for the regression task of predicting a continuous value. The model was compiled with the Adam optimizer and mean squared error as the loss function.

4.6.2.1 Model Training and Evaluation on Unshielded Low Threshold Dataset

The LSTM model was trained for 50 epochs with a batch size of 32, using 80% of the data for training and 20% for validation. Figure 4.58 illustrates the training and validation loss over the epochs. Notably, the training loss decreases steadily, while the validation loss initially decreases but shows some divergence as training progresses, indicating potential overfitting.

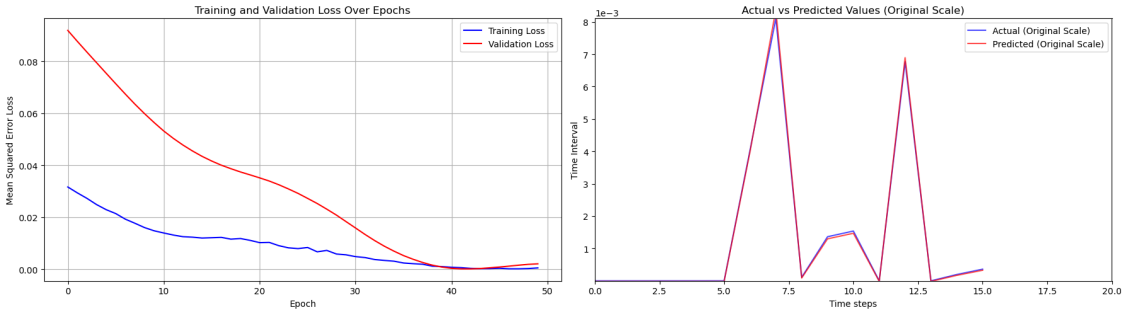


Figure 4.58: Training and Validation Loss Over Epochs and Actual vs Predicted Values (Unshielded low Threshold Dataset)

The model was evaluated on the test dataset, achieving a test loss of 1.3×10^{-5}, mean absolute error (MAE) of 0.002218, and root mean square error (RMSE) of 0.003596 on the scaled data. These metrics suggest the model performs well in predicting the time series data on a normalized scale.

When the predictions were inverse-transformed to the original scale, the model achieved a mean absolute error of 5.0×10^{-5} and a root mean square error of 8.0×10^{-5}. The results are summarized in Table 4.6.

Metric	Value
Test Loss	1.3×10^{-5}
Test MAE (Scaled)	0.002218
Test RMSE (Scaled)	0.003596
Mean Absolute Error (Original Scale)	5.0×10^{-5}
Root Mean Square Error (Original Scale)	8.0×10^{-5}

Table 4.6: Evaluation Metrics for the LSTM Model (Unshielded Low Threshold Dataset)

The actual versus predicted values on the original scale are depicted in Figure 4.58. The figure shows that the model's predictions closely follow the actual data points, further validating the model's effectiveness in capturing the temporal dependencies inherent in the data.

4.6.3 Deep Learning Model for Shielded Dataset

The LSTM model was trained using a dataset that focused on a single channel (column '224') from the dataset. The model used two LSTM layers with 128 and 64 units, each followed by a dropout layer with a rate of 0.2 to prevent overfitting. The final dense layer aimed to predict the sequence, aligning with the input shape. The model used the Adam optimizer and mean squared error (MSE) as the loss function.

4.6.3.1 Model Training and Evaluation on Shielded Dataset

I used two LSTM layers with 128 and 64 units and dropout layers to prevent overfitting. The model was compiled using the Adam optimizer and mean squared error as the loss function. Callbacks for early stopping and learning rate reduction were implemented. The model was trained for a maximum of 50 epochs with a batch size of 32, using an 80/20 split for training and validation data. The training and validation loss over epochs are depicted in Figure **??**, showing effective model convergence with minimal overfitting.

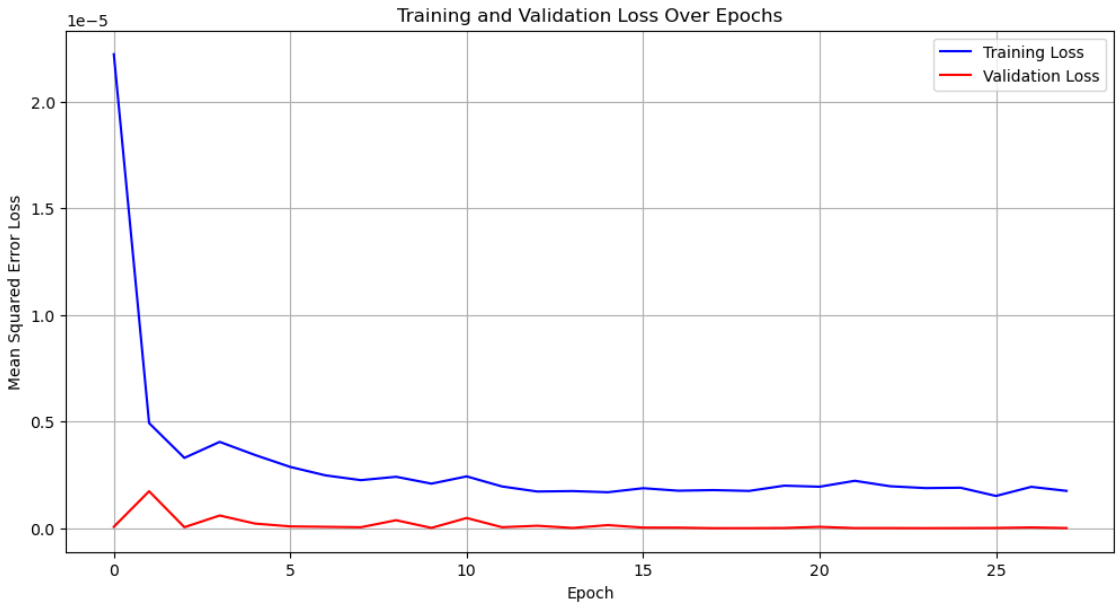


Figure 4.59: Training and Validation Loss Over Epochs for Shielded Dataset)

The model was tested using the test dataset; on the original scale, the model's predictions resulted in a mean absolute error (MAE) of 827.62 and a root mean square error (RMSE) of 9041.15. These results show that the model accurately predicts the time series data and effectively captures the underlying patterns.

The model's R-squared value was calculated for the training and testing datasets. The model achieved nearly perfect scores of 1.0000 for the training dataset and 0.9997 for the testing dataset, showing excellent model fit and generalization capability. The actual versus predicted values on the original scale are shown in Figure 4.62 . This diagram shows that the model's predictions closely match the actual data points, confirming that the model effectively captures the time-based patterns in the data.

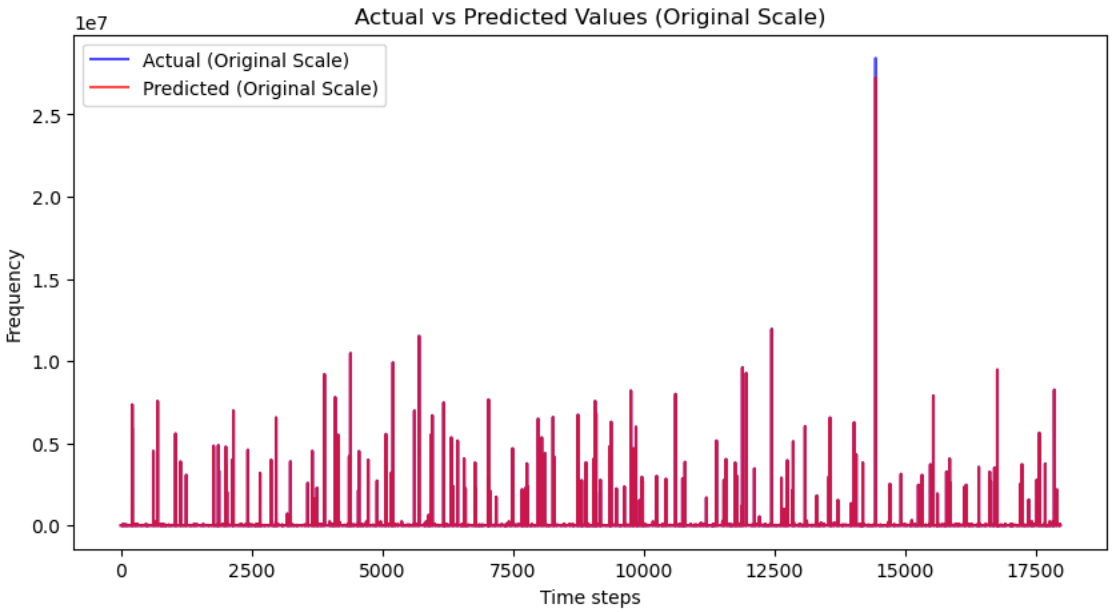


Figure 4.60: Actual vs Predicted Values on Original Scale

4.6.4 Deep Learning Model for Short Cable Dataset

The Long Short-Term Memory (LSTM) model was implemented using the TensorFlow Keras library for time series prediction. The model's architecture includes two LSTM layers, consisting of 128 and 64 units, respectively. Each LSTM layer is followed by a dropout layer with a dropout rate of 0.2 to mitigate overfitting by randomly dropping a fraction of the neurons during training. The final output layer has a single neuron and is suitable for the regression task of predicting continuous values. The model used the Adam optimizer, known for efficiently handling large datasets and noisy gradients. The mean squared error (MSE) was used as the loss function because it effectively penalizes large errors more severely than small ones.

4.6.4.1 Model Training and Evaluation on Short Cable Dataset

The LSTM model was trained over 50 epochs with a batch size of 32. The dataset was divided into two parts: 80% for training and 20% for validation. Figure 4.61 shows the training and validation loss over the epochs. The training loss quickly decreased in the first epochs, indicating that the model learned the data patterns effectively. The validation loss decreased overall, suggesting that the model generalized well to new data without overfitting.

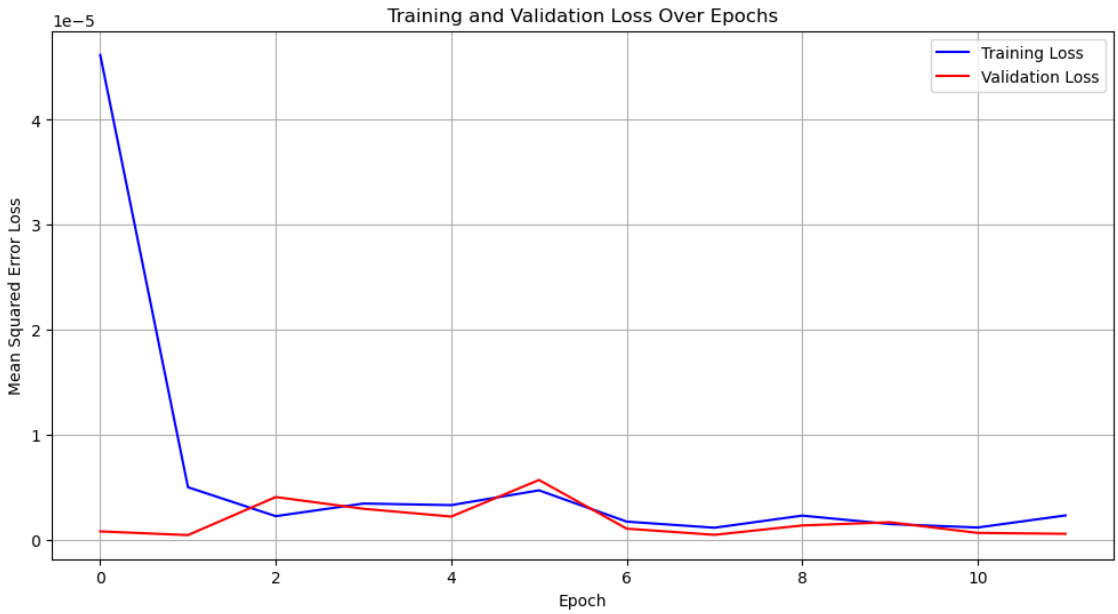


Figure 4.61: Training and Validation Loss Over Epochs

After completing the training, we tested the model using a separate dataset to see how well it could predict. The results showed that the test loss was extremely low at 4.66×10^{-7} on the scaled data. The original scale's mean absolute error (MAE) was 143,971.21, and the root mean square error (RMSE) was 875,234.67. These results indicate that the model is highly accurate in predicting time series data, effectively capturing the patterns and relationships within the dataset.

Metric	**Value**
Test Loss (Scaled)	4.66×10^{-7}
Mean Absolute Error (Original Scale)	143,971.21
Root Mean Square Error (Original Scale)	875,234.67

Table 4.7: Evaluation Metrics for the LSTM Model

Figure 4.62 shows the real values versus the predicted values on the original scale. The model's predictions match closely with the actual data points, which confirms its ability to capture the patterns in the time series data. This shows that the model can handle complex patterns well and predict future data points accurately.

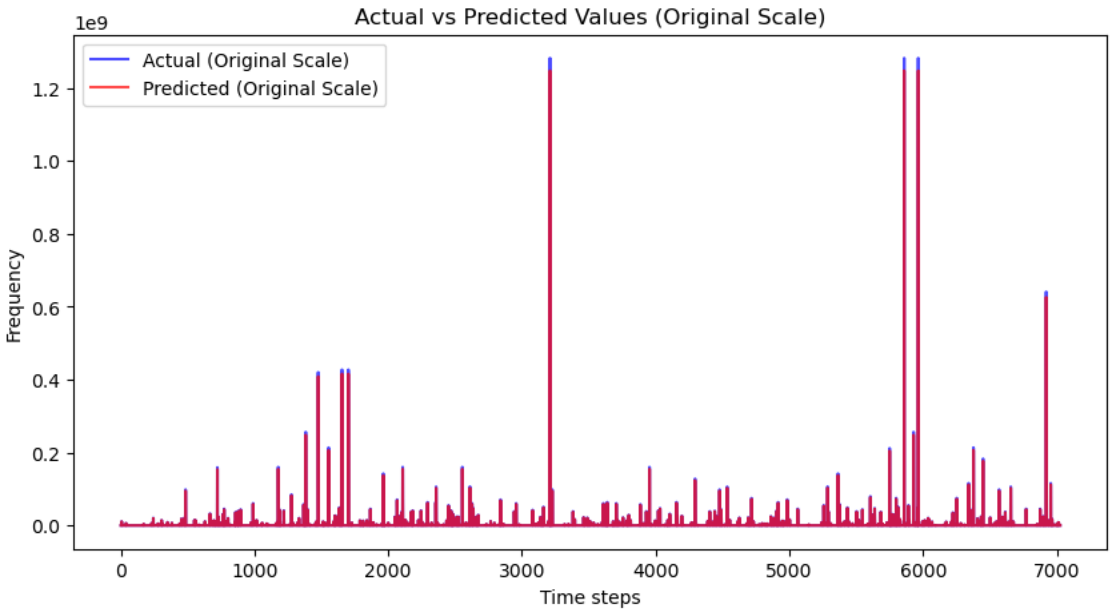


Figure 4.62: Actual vs Predicted Values on Original Scale

I found that the Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) were quite high after evaluating the LSTM model on my dataset. This is likely due to the very noisy nature of the data, which contains a lot of variability and outliers. The high level of noise makes it difficult for the model. I applied the model to two datasets, one with time intervals and another with frequencies.

4.6.5 Conclusion

(LSTM) models developed and evaluated in this study for various datasets show the effectiveness of deep learning in time series prediction tasks. Across all datasets, the models could capture temporal dependencies and accurately predict frequency and time interval, as indicated by low test loss values and strong R-squared scores. However, The datasets showed significant challenges due to inherent noise and variability, especially for the shielded and short cable datasets. In this case, the Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) were notably higher. This indicates that although LSTM models are strong at handling complex temporal patterns.

5 Conclusion

5.1 Summary of Findings

This study looked at the common mode current noise in the CMS Drift Tube cabling in order to better understand and reduce electromagnetic interference (EMI) in high-energy physics experiments. We used experimental measurements, signal processing, machine learning, and deep learning techniques to study different cable configurations and their impact on noise levels in the CMS Drift Tubes, especially in relation to the Large Hadron Collider (LHC) operations.

5.1.1 Cable Configuration Analysis

We examined four different cable setups: Unshielded with High Threshold, Short Cable, Unshielded with Low Threshold, and Shielded. Our findings show the following:

1. Shielded Setup: Shielding reduced low-frequency noise but unexpectedly let in some high-frequency noise, possibly because of outside electromagnetic interference or internal cable resonance. This shows how complex EMI pathways are and how challenging it is to block noise with shielding completely.

2. Short Cable Setup: Short cables generally have lower noise levels across most frequencies, suggesting that shorter transmission paths can effectively reduce high-frequency noise. However, some low-frequency noise is still present, showing that cable length alone is not enough to completely eliminate noise.

3. Unshielded with Low and High Thresholds: The unshielded setups showed different reactions to noise based on the detection threshold. The setup with a low threshold captured a wide range of frequencies, including higher-frequency components. At the same time, the high threshold configuration effectively filtered out most noise but also missed lower amplitude signals. These results show the trade-offs between sensitivity and noise reduction in finding subtle signals in high-energy physics experiments.

5.1.2 Advanced Signal Processing and Machine Learning

We used advanced methods like Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), autocorrelation, and machine learning and deep learning models to study noise characteristics and identify irregularities. These techniques helped us understand noise patterns across different cable configurations.

5.1.2.1 Signal Processing Techniques

Analysis using FFT and STFT showed different frequency components and changes in noise over time with various cable setups. Autocorrelation and inter-arrival time analyses found patterns and groupings in the noise. These findings are important for identifying noise sources and creating specific plans to reduce the noise.

5.1.2.2 Machine Learning for Anomaly Detection

Various machine learning models, including Isolation Forest, Local Outlier Factor (LOF), Gaussian Mixture Model (GMM), and One-Class SVM, were employed to detect anomalies in the frequency data. The Isolation Forest and GMM models were very good at precision and recall, effectively identifying significant deviations indicative of noise or interference. In contrast, LOF and One-Class SVM showed moderate effectiveness, with higher rates of false positives.

5.1.2.3 Deep Learning for Time Series Prediction

Long Short-Term Memory (LSTM) models were used to predict time series data related to frequency and time interval. These models accurately predicted noise patterns, even though the datasets had a lot of variability and inherent noise. They effectively captured time-based relationships.

5.1.3 Experimental Results and Correlations with Spectrum Analyzer Measurements

After many tries, we discovered strong connections between the noise in the collected data and the spectrum analyzer plot obtained using a common-mode current probe. Surprisingly, the data showed that the unshielded cable setup performed slightly better than the shielded one in certain situations. This unexpected outcome suggests that the noise propagation path is complex and hard to predict or control. A significant finding was that turning off the old electronics near the test setup led to a big decrease in noise. This resulted in much lower noise levels in the collected data, confirmed by a big drop in the spectrum analyzer readings. This discovery emphasizes how nearby electronic equipment can greatly affect noise levels. It also highlights the importance of carefully controlling the environment around sensitive experimental setups to reduce electromagnetic interference.

5.2 Implications for Future Research and Upgrades

The findings from this study have important implications for future research and the ongoing Phase 2 Upgrade of the CMS Drift Tubes:

5.2.1 Improving Noise Mitigation Strategies

Analyzing the different types of cable setups and their noise characteristics can help us improve our cable infrastructure and shielding methods. We should improve our shielding designs to block out high-frequency interference in the future. We also need to look into using different materials or setups that are less affected by outside noise sources.

5.2.2 Implementing at CERN

The results offer valuable information on noise behavior and its correlation with spectrum analyzer measurements. These findings were obtained in a controlled environment at INFN Padova. Future experiments should be carried out at CERN, directly on the CMS Drift Tubes, to further validate these results and improve the noise reduction techniques. Conducting these experiments in the actual operating environment of the LHC will provide a more thorough understanding of noise propagation and its effects under real-world conditions. Moreover, applying these techniques at CERN will allow for a more precise evaluation and optimization of

the cabling and shielding configurations, ultimately improving detector performance in high-energy physics experiments.

5.2.3 Expanding Deep Learning Applications

LSTM models have been successful in predicting time series data. This suggests deep learning could be helpful in real-time monitoring and adaptive noise filtering. Future researchers could look into using deep learning models in online systems to adjust noise thresholds and optimize signal quality during live experiments.

References

[1] D. B. et al., "The large hadron collider conceptual design," CERN, Tech. Rep. CERN/AC/95-05, 1995.

[2] L. Evans and P. B. (editors), "Lhc machine," JINST, Tech. Rep. 3 S08001D, 2008.

[3] "Lhc homepage," http://lhc.web.cern.ch/lhc/.

[4] "Cms experiment," https://home.cern/science/experiments/cms.

[5] The CMS Collaboration, "The compact muon solenoid technical proposal," CERN, Tech. Rep. CERN/LHCC 94-34, LHCCp1, 1994.

[6] ——, "The cms experiment at the cern lhc," *JINST*, vol. 3, p. S08004, 2008.

[7] ——, "The tracker project: Technical design report," CERN, Tech. Rep. CERN/LHCC 98-6, CMS TDR 5, 1998.

[8] ——, "Addendum to the tracker technical design report," CERN, Tech. Rep. CERN/LHCC 2000-016, CMS TDR 5 Addendum 1, 2000.

[9] "Nuclear instruments and methods in physics research section a: Accelerators, spectrometers, detectors and associated equipment," vol. 581, no. 1-2, pp. 343–346, 2007.

[10] A. Peisert, "Silicon microstrip detectors," Istituto Nazionale di Fisica Nucleare - Sezione di Padova, Via F. Marzolo 8, I - 31350 Padova, Tech. Rep., January 27 1992.

[11] R. Brown and D. Cockerill, "Electromagnetic calorimetry," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 666, pp. 47–79, 2012, review.

[12] The CMS Collaboration, "The hadronic calorimeter: Technical design report," CERN, Tech. Rep. CERN/LHCC 97-31, CMS TDR 2, 1997.

[13] ——, "The magnet project: Technical design report," CERN, Tech. Rep. CERN/LHCC 97-10, CMS TDR 1, 1997.

[14] ——, "The muon project: Technical design report," CERN, Tech. Rep. CERN/LHCC 97-32, 1997.

[15] CMS Collaboration, "Performance of the cms drift tube chambers with cosmic rays," *Journal of Instrumentation*, vol. 5, p. T03015, March 2010, published under licence by IOP Publishing Ltd.

[16] ——, "Performance of the cms cathode strip chambers with cosmic rays," *Journal of Instrumentation*, vol. 5, p. T03018, March 2010, published under license by IOP Publishing Ltd.

[17] A. Samalan and on behalf of the CMS Collaboration, "Improved resistive plate chambers for the upgrade of the cms muon detector," *Journal of Physics: Conference Series*, vol. 2374, p. 012006, 2022, pAPER • OPEN ACCESS.

[18] D. R. C. F. B. J. P.-P. I. R. J. S. M. B. D. C. R. I. S. V. P. D. R. S. M. F. R. A. Triossi, Á. Navarro-Tobar, "Electronics developments for phase-2 upgrade of cms drift tubes," *PoS TWEPP2018*, 2018.

[19] A. G. M. T. A. T. M. Bellato, A. Bergnoli, "Radiation hardness and quality validation of the on-detector electronics for the cms drift tubes upgrade," *Journal of Instrumentation*, vol. 19, p. C06001, 2024.

[20] M. Cerna and A. F. Harvey, "The fundamentals of fft-based signal analysis and measurement," 2000, application Note 041.

[21] M. Hasegawa-Johnson, "Lecture 5: Short-time fourier transform and filterbanks," http://example.com, 2020, lecture notes for ECE 417: Multimedia Signal Processing, Fall 2020.

[22] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. New York: Springer, 2016.

[23] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Wiley, 2015.

[24] R. J. Cook and J. F. Lawless, *Statistical Methods for the Analysis of Recurrent Events*. Springer, 2007.

[25] J. Sanchez, "Utilizing machine learning for signal classification and noise reduction in amateur radio," *arXiv preprint arXiv:2402.17771*, 2024. [Online]. Available: https://arxiv.org/abs/2402.17771

[26] K. M. T. F. T. Liu and Z.-H. Zhou, "Isolation forest," *Proceedings of the 8th IEEE International Conference on Data Mining*, pp. 413–422, 2008.

[27] A. A. P. et al., "Data quality monitoring of the cms experiment at the lhc," *arXiv preprint arXiv:1808.00911*, 2018. [Online]. Available: https://arxiv.org/abs/1808.00911

[28] A. Oluwasegun and J.-C. Jung, "Anomaly detection in nuclear reactor signals using gaussian mixture models," *Nuclear Engineering and Design*, vol. 402, p. 112098, 2023.

[29] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[30] B. Rubinstein, "A fast noise filtering algorithm for time series prediction using recurrent neural networks," *arXiv preprint arXiv:2009.12345*, 2020. [Online]. Available: https://arxiv.org/abs/2009.12345

[31] G. Tsibouris and M. Zeidenberg, "Testing the efficient markets hypothesis with gradient descent algorithms," *Neural Networks in the Capital Markets*, pp. 101–115, 1995.

[32] S. Z. F. Shan and H. Li, "Recurrent neural networks for predicting tunnel boring machine performance," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 16, pp. 1538–1551, 2024.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] P. Engineering, "Understanding common mode noise," 1999, application note on differential mode and common mode signals, and electromagnetic compatibility in UTP systems.

[35] H. W. Ott, *Electromagnetic Compatibility Engineering*. John Wiley & Sons, Inc., 2009, first published: 12 August 2009.

Acknowledgments

In the name of Allah, the Most Gracious and the Most Merciful. I begin by expressing my profound gratitude to Almighty ALLAH for granting me the strength, knowledge, and perseverance to complete this study and see it to its completion.

I express my deepest gratitude to my supervisor, Prof. Andrea Triossi of the Department of Physics at the University of Padova. His insightful guidance, invaluable support, and professional mentorship were instrumental in completing this dissertation. I am also sincerely grateful to Dr. Antonio Bergnoli of the INFN group at the University of Padova for providing the essential facilities, insightful guidance, and support necessary for this research.

Finally, I deeply appreciate my husband and parents' unwavering support, encouragement, and keen interest in my academic achievements.