UNIVERSITY OF ÇUKUROVA
INSTITUTE OF NATURAL AND APPLIED SCIENCES

**Ph.D. THESIS**

**Furkan DÖLEK**

# SEARCH FOR DARK MATTER IN THE MONO-JET EVENTS AT THE CMS EXPERIMENT AND DETECTOR R&D FOR THE DUNE EXPERIMENT

**DEPARTMENT OF PHYSICS**

**ADANA, 2023**

**UNIVERSITY OF ÇUKUROVA**

**INSTITUTE OF NATURAL AND APPLIED SCIENCES**

**SEARCH FOR DARK MATTER IN THE MONO-JET EVENTS AT THE CMS EXPERIMENT AND DETECTOR R&D FOR THE DUNE EXPERIMENT**

**Furkan DÖLEK**

**Ph.D. THESIS**

**DEPARTMENT OF PHYSICS**

We certify that the thesis titled above was reviewed and approved for the award of degree of the Doctor of Physics by the board of jury on 22/02/2023.

| ............................................ | ...................................................... | .................................. |
|---|---|---|
| Prof. Dr. İsa DUMANOĞLU<br>SUPERVISOR | Prof. Dr. Aysel KAYIŞ TOPAKSU<br>MEMBER | Prof. Dr. Eda EŞKUT<br>MEMBER |
| ............................................ | ...................................................... | |
| Asst. Prof. Dr. Yalçın GÜLER<br>MEMBER | Assoc. Prof. Dr. Evrim Ersin KANGAL<br>MEMBER | |

This Ph.D.Thesis is performed in Department of Physics of Institute of Natural and Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Sadık DİNÇER
Director
The Institute of Natural and
Applied Sciences

This study was supported by Çukurova University Scientific Research Fund and Scientific and Technological Research Council of Turkey (TUBITAK).
Ç.Ü. Project No: FDK-2018-10357
TUBITAK Project No: 1059B141801124

*To my lovely wife*

# ÖZ

## DOKTORA TEZİ

**CMS DENEYİ'NDE MONO-JET OLAYLARINDA KARANLIK MADDE ARAŞTIRMASI VE DUNE DENEYİ İÇİN DEDEKTÖR AR-GE**

**Furkan DÖLEK**

**ÇUKUROVA ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**
**FiZiK ANABİLİM DALI**

Danışman : Prof. Dr. İsa DUMANOĞLU
İkinci Danışman : Doç. Dr. Burak BILKI
Yıl : 2023, Sayfa : 113
Jüri : Prof. Dr. İsa DUMANOĞLU
: Prof. Dr. Aysel KAYIŞ TOPAKSU
: Prof. Dr. Eda EŞKUT
: Dr. Öğr. Üyesi Yalçın GÜLER
: Doç. Dr. Üyesi Evrim Ersin KANGAL

Bu tez, iki çalışmadan oluşmaktadır. İlk çalışma, 13 TeV merkezi kütle-enerjili proton-proton çarpışmalarında gerçekleştirilen Run 2 sırasında CMS tarafından toplanan 137,2 $fb^{-1}$ tam veri setini kullanarak, mono-$H_s$ son durumuna bozunan Karanlık Higgs bozonu sinyalini aramayı amaçlamaktadır. Olası mono-$H_s$ kanıtlarını bulmak için $p_T^{miss}$ ve geri tepen büyük konili jet kütlesi gibi değişkenleri kullanırız. Ancak analiz model bağımlıdır ve $p_T^{miss}$ dağılımı ve büyük koni jetin kütle dağılımı ayrıntılı bir analiz gerektirir.

İkinci çalışma, ProtoDUNE-SP'deki sıvı argon ortamına ksenon katkılanmasının etkilerini araştırmak üzerinedir. Çalışmanın amacı, nitrojen kirliliğinden kaynaklanan ışık kayıplarını geri kazanmak ve zaman projeksiyon odası deneylerinde argonun saflığını korumaktır. Ksenonun sıvı argona enjekte edilmesinin detektör performansı üzerindeki etkilerini gösteriyoruz ve ilk sonuçlarla karşılaştırıyoruz. Bu çalışma, ProtoDUNE-SP'deki ilk ksenon katkılanmasının performansıyla ilgili kapsamlı bir analiz sunar.

Bu çalışmalar, yeni fizik arayışlarındaki süregelen çabaları ve detektör teknolojisindeki gelişmeleri göstermektedir.

**Anahtar Kelimeler:** Karanlık Madde,Ksenon, CMS, ProtoDUNE-SP

**ABSTRACT**

**Ph.D THESIS**

**SEARCH FOR DARK MATTER IN THE MONO-JET EVENTS AT THE CMS EXPERIMENT AND DETECTOR R&D FOR THE DUNE EXPERIMENT**

**Furkan DÖLEK**

**DEPARTMENT OF PHYSICS**
**INSTITUTE OF NATURAL AND APPLIED SCIENCES**
**UNIVERSITY OF ÇUKUROVA**

Supervisor : Prof. Dr. İsa DUMANOĞLU
Co-Supervisor : Assoc. Prof. Dr Burak BILKI
Year : 2023, Pages : 113
Jury : Prof. Dr. İsa DUMANOĞLU
: Prof. Dr. Aysel KAYIŞ TOPAKSU
: Prof. Dr. Eda EŞKUT
: Asst. Prof. Dr. Yalçın GÜLER
: Assoc. Prof. Dr. Evrim Ersin KANGAL

This thesis consists of two studies. The first study aims to search for evidence of a Dark Higgs boson in the mono-$H_s$ final state using the entire dataset of 137.2 $\text{fb}^{-1}$ collected by CMS during Run 2 using proton-proton collisions with a center-of-mass energy of 13 TeV. We employ variables such as $p_{\text{T}}^{\text{miss}}$ and the mass of the recoiling large-cone jet to identify a possible mono-$H_s$ signature. However, the analysis is model-dependent, and the $p_{\text{T}}^{\text{miss}}$ distribution and mass spectrum of the large-cone jet require a detailed analysis to be generalized.

The second study discusses the performance of Xenon doping in the liquid argon environment at ProtoDUNE-SP. The study aims to recover light losses due to nitrogen contamination and maintain the purity of argon in time projection chamber experiments. We demonstrate the effects of injecting Xenon into liquid argon on the detector's performance and compare them to the initial results. The study presents a comprehensive analysis of the performance of the first Xenon doping at ProtoDUNE-SP.

These studies demonstrate the ongoing efforts in the search for new physics and the advancements in detector technology.

**Key Words:** Dark Matter,Xenon ,CMS, ProtoDUNE-SP

## GENİŞLETİLMİŞ ÖZET

Karanlık Madde (KM) araştırma çalışmaları, Kayıp Dik Enerji (KDE) olaylarına odaklanarak yürütülmektedir. Yüksek enerjili parçacık çarpışmalarında enerji ve momentum korunur. Büyük Hadron Çarpıştırıcısı (BHÇ)'ında gerçekleştirilen doğrudan veya dolaylı KM araştırmaları arasında, mono-jet, mono-Z, mono-Higgs, mono-top ve vektör bozon sinyalleri bulunmaktadır. Bu tez çalışması süresince, literatür taraması sonucu mono-top ile ilgili bilgiler elde edilmiş ve devam eden çalışmalara altyapı oluşturulmuştur.

KM, evrenin işleyişinin anlaşılması açısından fizikçiler için büyük önem taşır ve kozmolojinin en büyük sorunlarından biridir. KM henüz tespit edilememiş olmasına rağmen, kozmolojik gözlemler ve ölçümlerle varlığı desteklenmektedir. Galaksilerin etrafındaki yoğun kütleli halo, galaksi döngüsel hızlarındaki anormallikler, kütle çekimsel merceklenme ve kozmik mikrodalga arka plan ışıması üzerindeki etkiler, KM'nin varlığına ilişkin güçlü kanıtlar sunmaktadır.

BHÇ, KM araştırmaları için önemlidir ve teorik olarak KM üretme ve tespit etme potansiyeline sahip olduğu düşünülmektedir. Bu tezde bahsedilen çalışmalar, 2016, 2017 ve 2018 yıllarında CMS verileri kullanılarak yapılan Karanlık Higgs (KH) araştırmalarını içermektedir. Bu çalışma, LPC-DM (LHC Physics Center - Dark Matter) grubunun bir üyesi olarak gerçekleştirilmiş ve bu alanda yazılım geliştirici ve fizik araştırmacısı olarak katkı sağlanmıştır.

KM parçacıklarının doğrudan dedektörde gözlemlenememesi nedeniyle, KM'nin varlığını tespit etmek için Standart Model (SM) parçacıklarının gözlemlenmesi gerekmektedir. Çarpıştırıcılarda gerçekleşen reaksiyonlar sonucu ortaya çıkan SM parçacıkları ve bu parçacıkların geri tepme momentum gözlemlenir. Bu süreçte, KM sinyali, mono-X veya MET+X (MET, KDE) şeklinde ortaya çıkar.

KM araştırmaları için kullanılan önemli kanallardan biri, başlangıç durumunda bir kuark veya gluon ile birlikte KM üretildiği olaylardır. CMS'de mono-jet olaylarında KM araştırmaları, KM'nin SM son durumlarına direkt bozunumları evrendeki kalıntı bolluğuna ulaştığında, parametre uzayını güçlü bir şekilde sınırlar (Akchurin et al., 2016). Eğer KM parçacıkları karanlık sektördeki en hafif parçacıklar değilse, bu sınırlama bir miktar azalmış olur ve yeni bozunum kanallarına yol açar. KM parçacıklarının ardışık olarak SM durumlarına dönüştüğü bir çift KH bozonuna bozunması gibi yeni bir bozunum kanalı mümkün hale gelir. Spin-1 $Z'$ taşıyıcısı, SM'in Z bozonuna benzer bir karanlık sektör taşıyıcı parçacığıdır. Bu taşıyıcı, karanlık sektör parçacıkları arasındaki etkileşimleri sağlar ve onların üretiminde ve bozunumlarında rol oynar. Eğer karanlık sektörde böyle bir taşıyıcı mevcutsa, bu durum KH yayılımının ve $Z'$ üretiminin olasılığını artırabilir.

KH üretimi ve arka plan süreçleri arasındaki ayırım, kayıp dik momentum sinyali ve büyük yarıçaplı tek jetler gibi belirli sinyallerin ölçümüne dayanmaktadır. Bu sinyaller, potansiyel olarak zayıf etkileşimli parçacıkların varlığına işaret edebilir ve bu parçacıkların tespiti için önemli bir araç sağlar. Ayrıca, KH bozonunun son durumu, mono-jet ve mono-V (V burada Z veya W bozonunu temsil eder) son durumlarına benzer diyagramlara yol açar. Bu nedenle, bu çalışmada tüm mono-jet, mono-Higgs ve mono-V durumları araştırılmaktadır. 2016, 2017 ve 2018 yıllarındaki Run2 verilerinin tamamı 137,2 $\mathrm{fb}^{-1}$ ve CMS 13 TeV merkez kütle enerjisi ile sunulacaktır.

KH analizinde, LPC-DM grubu işbirliği ile geliştirilen DECAF adlı yazılım kullanılarak erken sonuçlar elde edilmiştir. Bu yazılımın temel amacı NanoAOD benzeri bir format olan yeni bir ntuple formatının üretilmesidir. DECAF yazılımının temel kütüphanesi, FNAL Analaysis Tool'un yeni adı olan COFFEA'ya referans vermektedir. Bu kütüphane, yeni nesil makine öğrenimi, veri analizi ve veri görselleştirme gibi kütüphaneleri içermektedir. Bu yazılımlar kullanılarak genel olarak mono-jet adımları takip edilmiştir. Mono-jet çalışmalarına ek olarak mono-dark-Higgs ve mono-V sinyalleri de dahil edilerek hassasiyet artırılmaya çalışılmıştır.

Bu tezde, CMS çalışmalarına paralel olarak protoDUNE-SP deneyinin sıvı argon zaman projeksiyon odası (LArTPC) dedektörünün foton algılama performansı üzerine çalışmalar yapılmıştır. Bu çalışma kapsamında, sıvı argon zaman projeksiyon odalarının foton algılama performansı incelenmiştir. Bu teknoloji, nötrino osilasyonları ve karanlık madde etkileşimleri gibi alanlarda lider konumda olup, nadir olayların araştırılması için kullanılmaktadır. Sıvı argon zaman projeksiyon odaları, 40 yılı aşkın süredir geliştirilmekte olup, DUNE (Deep Underground Neutrino Experiment) deneyi için son derece karmaşik hale gelmiştir.

DUNE, nötrinoların özelliklerinin daha ayrıntılı bir şekilde araştırılması için yapılan bir deneydir. Bu deney, Fermilab'da yer alan bir hızlandırıcıdan nötrino ışınları üretilerek, bu nötrinoların 1300 km uzaklıktaki Güney Dakota'daki Sanford Yeraltı Araştırma Tesisi'nde (SURF) bulunan devasa bir dedektöre yönlendirilmesiyle gerçekleştirilmektedir.

DUNE deneyi, gelecekte nötrino araştırmalarında bir dönüm noktası olması beklenmektedir. Bu deneyin hedefleri, nötrino osilasyonlarının daha hassas ölçümlerinin yapılması, kozmik ışınların arka plan gürültüsünün en aza indirilmesi ve karanlık madde araştırmalarına imkan sağlamasıdır. Ayrıca, DUNE projesi nötrino astronominin geliştirilmesine de yardımcı olacaktır. Bu amaçlar doğrultusunda, uluslararası bir işbirliği ile yürütülen DUNE projesi, çeşitli ülkelerden bilim insanları ve kurumları tarafından desteklenmektedir. Bu hedefler, nötrino araştırmalarının gelecekteki yönünü belirleyecektir.

Sıvı argon zaman projeksiyon odaları (LArTPC), nötrino ve karanlık madde araştırmalarında önemli bir role sahip olan ve son yıllarda giderek daha popüler hale gelen bir dedektör teknolojisidir. Bu teknolojinin geliştirilmesi ve foton algılama performansının artırılması, nadir olayların daha hassas bir şekilde araştırılmasına olanak sağlamaktadır. ProtoDUNE-SP deneyinde yapılan çalışmalar, sıvı argon zaman projeksiyon odalarının foton algılama performansını iyileştirmek için önemli adımlar atmaktadır. Bu çalışmalar, DUNE deneyi için gerekli olan son derece karmaşık LArTPC dedektörlerinin geliştirilmesine de katkı sağlayabilir.

LArTPC'ler, homojen bir elektrik alanı altında argonda biriken iyonizasyon yükünü kullanarak olayların uzaysal konumlarinin ve enerjilerinin yeniden yapılandırılmasını sağlar. Sıvı argon, yüksek performanslı bir sintilatör olup, Vakum UltraViolet (VUV) bölgesinde 128 nm dalga boyunda bir spektrum ve yaklaşık $4 \times 10^4$ ($2.4 \times 10^4$) foton/MeV verimle ışık yaymaktadır. Bu verim, 0 V/cm (500 V/cm) elektrik alanı için geçerlidir. Parıldama ışığı, tekli halinin uyarılmasıyla üretilir.

LArTPC teknolojisinin kullanımı, DUNE deneyinin hedeflerine ulaşmak için çok önemlidir. Bu teknoloji sayesinde, nötrino osilasyonları ve karanlık madde araştırmalarında daha hassas ölçümler yapılabilmekte ve nadir olayların araştırılması mümkün hale gelmektedir. LArTPC dedektörlerinin geliştirilmesi ve foton algılama performansının artırılması, DUNE deneyinin ana hedeflerinden biri olan nötrino astronominin geliştirilmesine de yardımcı olacaktır. Bu bağlamda, ProtoDUNE-SP deneyinde yapılan çalışmaların önemi büyüktür ve sıvı argon zaman projeksiyon odalarının foton algılama performansının artırılması için atılan adımlar, DUNE deneyi için gereken son derece karmaşık LArTPC dedektörlerinin geliştirilmesine de katkı sağlayacaktır.

Sıvı argonda VUV ışığını tespit etmek zor olsa da, fiziksel avantajları oldukça dikkat çekicidir, özellikle DUNE deneyi için. Sintilasyon ışığı, TPC'deki üçüncü uzaysal koordinatı birkaç nanosaniye hassasiyetle yeniden oluşturmak için kullanılan etkileşim süresini (t0) sağlar. Ayrıca, foton ölçümünün en büyük faydası, hızlandırıcılardan ve uzaydan gelen parçacıkların ayırt edilebilmesi olarak görülmektedir.

**ACKNOWLEDGEMENTS**

I am deeply grateful to my supervisor, Prof. Dr. Isa DUMANOGLU, for his guidance and support at the beginning of my High Energy Physics Ph.D. career. His expertise and encouragement have been instrumental in shaping my research and career trajectory. I would also like to thank the Çukurova University CMS team and Project manager Prof. Dr. Aysel KAYIS TOPAKSU for their unwavering support throughout my studies, as well as Doc Dr. Evrim Ersin KANGAL and Eda ESKUT for their role as jury members in evaluating my research.

Special thanks go to Prof. Dr. Jay DITMAN and the HCAL team for their invaluable guidance and insights. I am also grateful to Dr. Yalcin GULER for sharing his expert knowledge of exotica analysis and teaching me about the mainstream software of CERN and CMS.

I extend my gratitude to the CMS Hadronic Barrel and Endcap Upgrade team, as well as my colleagues Alexey KALININ and Grace CUMMINGS, for their contributions to my research. Their insights and assistance were crucial to the success of my project.

I am incredibly grateful to my co-advisor, Dr. Burak BILKI, for introducing me to the DUNE experiment. His advice and support have helped me gain an upper hand. I also had the pleasure of working with Dr. Francesco PIETROPAOLO and Dr. Filippo RESNATI, whose mentorship and guidance were indispensable in my growth as a researcher.

I would like to express my appreciation to my colleagues Serhan TUFANLI and Christopher MACIAS, with whom I worked closely on the production of the anode plane. Their dedication, expertise, and support were essential to the success of this project, and I am grateful for the opportunity to have worked with them.

I would like to express my gratitude to my mother, sister, and father for their unwavering love and support throughout my life and especially during my Ph.D. journey. Their encouragement, understanding, and sacrifice have been critical in enabling me to pursue my academic dreams. I am forever indebted to them for their unwavering commitment to my success.

Lastly, I want to express my deepest gratitude to my wife, Fatma BORAN, who has been my rock throughout my Ph.D. journey. As a fellow physicist, she has been a constant source of inspiration, motivation, and invaluable support. Her unwavering love and encouragement have been instrumental in helping me achieve my goals, and I cannot thank her enough for all she has done for me. I love her so much and am grateful to have her by my side every step of the way.

I am grateful for the support of the TENMAK project (code: 2018TAEKCERN-A5.H6.F2.19) for their contributions to this study.

A part of this thesis work has been carried out within the scope of the TUBITAK Leading Researchers Program under project number 118C224, titled Development of Novel Radiation Detector Technologies and Data Analysis Algorithms.

**CONTENTS** **PAGE**

**LIST OF TABLES** **PAGE**

**LIST OF FIGURES** **PAGE**

**ABBREVATIONS**

| | | |
|---|---|---|
| CERN | : | European Organization for Nuclear Research |
| CMS | : | Compact Muon Solenoid |
| LHC | : | Large Hadron Collider |
| SM | : | Standard Model |
| BHÇ | : | Beyond the Standard Model Higgs |
| KM | : | Dark Matter candidate particle |
| VBF | : | Vector Boson Fusion |
| PDF | : | Parton Distribution Function |
| MC | : | Monte Carlo simulation |
| QCD | : | Quantum Chromodynamics |
| QED | : | Quantum Electrodynamics |
| QFT | : | Quantum Field Theory |
| NNLO | : | Next-to-Next-to-Leading Order |
| NLO | : | Next-to-Leading Order |
| LO | : | Leading Order |
| R & D | : | Research and Development |
| DUNE | : | Deep Underground Neutrino Experiment |
| LArTPC | : | Liquid Argon Time Projection Chamber |
| ProtoDUNE-SP | : | Prototype Double-Phase Liquid Argon Time Projection Chamber |
| TPC | : | Time Projection Chamber |
| VUV | : | Vacuum UltraViolet |
| PD | : | PhotoDetector |
| PMT | : | PhotoMultiplier Tube |

| | |
|---|---|
| DAQ | : Data Acquisition |
| ROI | : Region of Interest |
| DECAF | :Analysis Framework |
| COFFEA | : Columnar Object Framework For Effective Analysis |
| LPC-DM | : LHC Physics Center - Dark Matter |
| MET | : Missing Transverse Energy |

# 1. INTRODUCTION

Two different studies were carried out in this thesis to explore different aspects of particle physics research. The first investigation focuses on the CMS experiment and its physics results, while the second one is centered on the Deep Underground Neutrino Experiment (DUNE) and detector performance studies."

One of the central topics explored in this thesis is dark matter (DM), which comprises most of the matter in the Universe as observed through astrophysical means (Smith et al., 2005). DM is hypothesized to be composed of stable, electrically neutral, massive particles that interact with baryons, at least through gravitational force. If such a DM particle also interacts non-gravitationally with SM particles, then DM could be produced in proton-proton collisions at the Large Hadron Collider (LHC) at CERN. However, since DM particles do not leave detectable signals in the detector, their presence can only be inferred when they are produced in association with a visible SM particle $X(=g, q, \gamma, Z, W$, or $h)$ (Abercrombie et al., 2015). This production generates final states commonly referred to as mono-X or MET+X signatures, where MET (or ptmiss) is the missing transverse momentum.

The mono-jet channel, which typically searches for DM produced in association with a gluon or a quark emitted as initial-state radiation (ISR), is the most sensitive analysis channel for DM production. The CMS mono-jet DM search has strongly limited the parameter space where DM particles can achieve their relic abundance from direct annihilation into SM final states (Akchurin et al., 2016). However, this constraint can be relaxed if DM particles are not in the lightest state in the dark sector, leading to new annihilation channels. If the DM mass is generated via a Higgs mechanism in the dark sector, and the resulting dark Higgs boson is lighter than DM, DM particles can annihilate into a pair of dark Higgs bosons ($h_s$), which then decay into SM states. This scenario quickly sets the observed relic abundance. Assuming a small mixing with the SM Higgs boson, the dark Higgs boson mainly decays into a pair of b quarks.

If the dark sector includes a spin-1 $Z'$ mediator, it is possible for the $Z'$ to be produced and radiate a dark Higgs (known as dark-Higgsstrahlung) with high probability. This gives rise to a distinctive signature, called the mono-$h_s$ signature, which is characterized by large missing transverse momentum due to the decay of the $Z'$ mediator into DM and a double-b-tagged, highly-boosted large-radius jet. This signature can be used to distinguish the dark Higgs production from the background processes (Duerr, Grohsjean, et al., 2017). In this paper, we present an analysis that searches for evidence of DM in the mono-$h_s$ final state. We report the results obtained using the full dataset of $137.2\,\mathrm{fb}^{-1}$ collected by the Compact Muon Solenoid (CMS) experiment at a center-of-mass energy of 13 TeV during Run 2 (2016 to 2018).

In addition to the CMS analysis described in the previous section, this thesis also presents an analysis of data from DUNE and a study of the performance of the DUNE photon detection system and the ProtoDUNE experiments and detector performance studies.

The DUNE experiment is a long-baseline neutrino experiment that aims to measure the parameters that govern neutrino oscillation, which is the phenomenon of neutrinos changing from one type to another as they travel through space. DUNE consists of two large detectors, one located at Fermilab in the United States and one at the Sanford Underground Research Facility in South Dakota, separated by a distance of 1300 kilometers. The detectors are designed to observe neutrinos produced by a beam of protons at Fermilab and measure their properties as they travel to the sensor in South Dakota. The photon detection system in DUNE is an essential component of the experiment, designed to detect light produced by charged particles created when neutrinos interact with the detector material. The system consists of arrays of photodetectors sensitive to the ultraviolet and visible light produced by these particles. The performance of this system is critical for achieving the experiment's scientific goals, and it is essential to understand the system's response to different types of events to optimize its performance.

ProtoDUNE, on the other hand, is a prototype for the DUNE detectors located at The European Organization for Nuclear Research (CERN) in Switzerland. It consists of two detectors, one using liquid argon and the other using liquid argon and gaseous argon mixture. ProtoDUNE aims to test the performance of the detectors and their readout systems before installation in the DUNE experiment. This thesis presents results from the ProtoDUNE experiments and detector performance studies.

This thesis also includes an analysis of data from the X-arapuca experiment, a novel photon detection technology developed to enhance the detection efficiency of scintillation light in liquid argon detectors. The X-arapuca technology utilizes arrays of wavelength-shifting plates and silicon photomultipliers to collect and detect the produced photons. The analysis presented in this thesis focuses on the performance of the X-arapuca system in detecting scintillation light in ProtoDUNE, as well as the system's potential applications in future liquid-argon neutrino detectors.

## 2. CMS DETECTOR

CERN hosts the most significant scientific research facility, which includes the Large Hadron Collider (LHC) and multi-collaborative experiments such as CMS, A Toroidal LHC Apparatus (ATLAS), and many others. In addition to important discoveries in particle physics, it is the birthplace of significant technological developments, such as the World Wide Web and positron emission tomography (PET).

The CMS experiment, shown in Figure 2.1., is one of two large particle physics detectors constructed on the LHC at CERN. It has a vast physics program, from studying the Standard Model (including the Higgs boson) to looking for extra dimensions and particles that could form dark matter. Although it has the same scientific objectives as the ATLAS experiment, it uses distinct technical solutions and a different magnet-system design.

The LHC accelerator delivers protons to the CMS collision center every 50 nanoseconds, with the number of protons reaching a hundred billion orders of magnitude at very close speeds to light. Thanks to the powerful magnets, two opposite beam bunches can be focused to a diameter of fewer than 20 microns, like a pinpoint. However, the recorded number of collisions is only around 20 per bunch.

CMS is layered with subdetectors, one on top of another. The first detector is the Silicon Tracker, which covers an angular range of $|\eta| < 2.5$ and provides the collision's location, timing, and magnitude with the help of strip diodes (Karimäki et al., 1997).

The Electromagnetic Calorimeter (ECAL) surrounds the Silicon Tracker, and incoming electrons and photons are forced to stop at this level and create showers via the electromagnetic interaction. Such calorimeters can measure energies with high

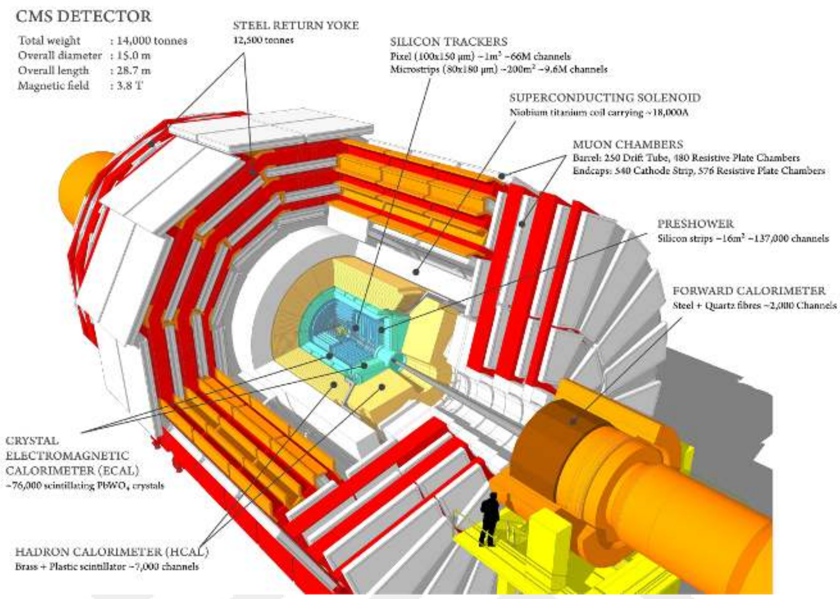



Figure 2.1. The CMS detector at the LHC, showing its sub-detectors and components. The complex system measures the properties of particles produced in proton-proton collisions, including the silicon tracker, calorimeters, muon detectors, and super-conducting solenoid magnet.

resolution, and the deposition of the energy is expected to be proportional to light. ECAL is composed of 68,524 scintillating lead tungstate ($PbWO_4$) crystals.

Since ECAL is transparent to hadronic particles, such as quarks, gluons, protons, neutrons, and pions, it is covered with a Hadronic Calorimeter (HCAL). Thus, hadronic particles are consumed at this stage and help to extract the timing, positioning, and energy information. HCAL is constructed with layers of brass absorbers and plastic scintillators, with the light output carried by optical fibers to the photodetectors. Formerly, it was read with a hybrid photodiode; after phase 1 upgrade, it is now read with SiPMs. HCAL has an angular coverage of $|\eta| < 5.2$ in the pseudorapidity range.

The initial trackers and calorimeters surround the Super-Conducting Solenoid

Magnet, which produces a 4 Tesla magnetic field that bends the charged particle trajectory and allows researchers to calculate the momentum and ratio between charge and mass.

# 3. THE DARK HIGGS

The dark Higgs model used to interpret the results presented in this note predicts Majorana dark matter (DM) particles $\chi$ that obtain their mass from the vacuum expectation value $w$ of a new complex Higgs field $S$, which is a singlet under the SM gauge group (Duerr, Kahlhoefer, et al., 2016; Kahlhoefer et al., 2016b). The field $S$ carries a charge $q_S$ under a new $\mathrm{U}(1)'$ gauge group, so $w$ breaks the gauge symmetry spontaneously and generates the mass of a corresponding $Z'$ gauge boson. The symmetry breaking gives rise to a new physical Higgs boson, defined via $S = \frac{1}{\sqrt{2}}(s+w)$, referred to as the dark Higgs boson $h_s$. In this setup, DM particles acquire an axial coupling to the $Z'$, so that all three particles in the dark sector are coupled to each other.

The renormalizable terms in the Lagrangian are given by:

$$\mathcal{L}\chi = -\frac{1}{2}g' q\chi Z'^{\mu}\bar{\chi}\gamma^5\gamma_\mu\chi - \frac{y_\chi}{2\sqrt{2}}s\bar{\chi}\chi + \frac{1}{2}g'^2 q_S^2 Z'^{\mu}Z'_\mu(s^2+2sw) \qquad (3.1)$$

where $g'$ denotes the $\mathrm{U}(1)'$ gauge coupling, and gauge invariance requires that the charge of the DM particle satisfy $q_\chi = q_S/2$.

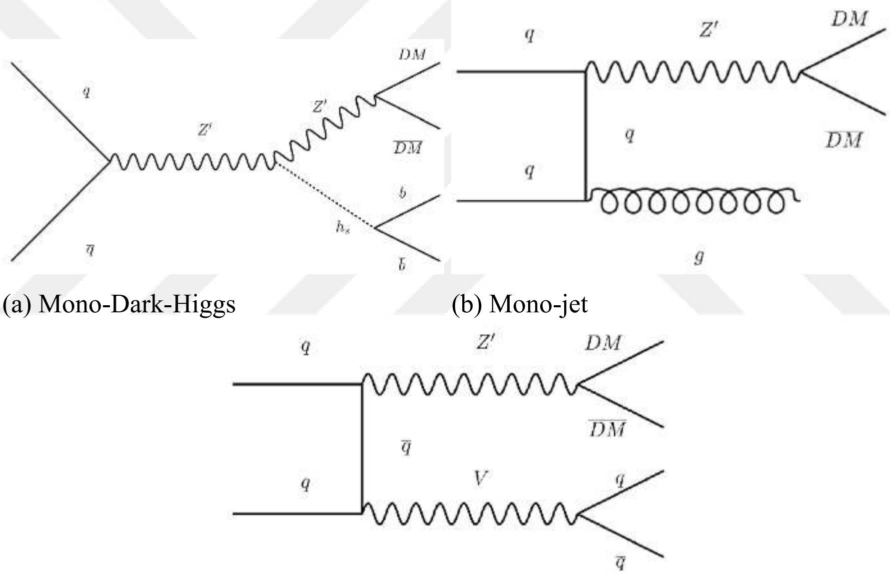
The masses of the DM particle and the $Z'$ are respectively given by $m_\chi = y_\chi w/\sqrt{2}$ and $m_{Z'} = 2g' q_\chi w$, while the mass of the dark Higgs boson $m_s$ is an independent parameter, which should not be much larger than $w$ to preserve perturbative unitarity. The dark sector thus contains four independent parameters, $m_\chi$, $m_{Z'}$, $m_s$, and $g_\chi \equiv g' q_\chi$. In terms of these parameters, the interaction Lagrangian yields:

$$\mathcal{L}\chi = -\frac{1}{2}g\chi Z'^{\mu}\bar{\chi}\gamma^5\gamma_\mu\chi - g_\chi\frac{m_\chi}{m_{Z'}}s\bar{\chi}\chi + 2g_\chi Z'^{\mu}Z'_\mu(g_\chi s^2 + m_{Z'}s). \qquad (3.2)$$

In this scenario, the $h_s$ boson is the lightest state in the dark sector, and DM particles can annihilate into other dark sector states, which subsequently decay into SM states. The non-zero mixing between the $h_s$ boson and the SM Higgs boson (Frandsen et al., 2012)

For this search, only the vector couplings of the $Z'$ boson to SM quarks will be considered:

$$\mathcal{L}_\chi = -g_q Z'^{\mu} \bar{q}\gamma_\mu q \tag{3.3}$$




(a) Mono-Dark-Higgs (b) Mono-jet

(c) Mono-V

Figure 3.1. Feynman diagrams for the mono-$h_s$ production of the dark Higgs model. The diagrams depict the possible ways in which the dark Higgs boson ($h_s$) can be produced in association with SM particles, resulting in missing transverse momentum ($p_T^{miss}$) due to the undetected DM particle. The diagrams represent the Mono-Dark-Higgs, Mono-jet, and Mono-V processes, respectively.

The axial-vector coupling of the $Z'$ boson to SM quarks will not be considered in this study, as it requires significant modifications to ensure SM gauge invariance (Kahlhoefer et al., 2016a). Although additional couplings of the $Z'$ boson to SM leptons are possible, they would impose further constraints (Ekstedt et al., 2016), and hence, they are not taken into account.

Figure 3.1. displays the Feynman diagrams for the mono-$h_s$ production of the dark Higgs model. The figure shows three different models, namely Mono-Dark-Higgs, Mono-jet, and Mono-V. These models represent the production of the dark Higgs boson in association with dark matter particles, a jet, or a vector boson. The search is focused on the production of the dark Higgs boson in association with a single jet or a vector boson, as these final states offer high sensitivity for the signal, while keeping the background at a manageable level.

In addition to the dark Higgs model, this thesis also investigates the i2HDM (Inert Doublet Model Type II) which is another extension of the SM that introduces a second Higgs doublet. Like the dark Higgs model, this model also predicts additional neutral gauge bosons, such as the $Z'$ boson, and can potentially explain the observed excess of events in the dilepton final state. However, the i2HDM also predicts modifications to the couplings of the SM particles, including the top quark, which can lead to different experimental signatures. Therefore, this study explores the potential of the i2HDM model to explain the excess events observed in the dilepton final state.Type 2 refers to the specific way in which the second Higgs doublet couples to fermions in the i2HDM. In type 2 models, one Higgs doublet couples to up-type quarks and the other to down-type quarks and leptons, resulting in modified couplings of the top quark and other fermions. This is in contrast to type 1 models, where both Higgs doublets couple to all fermions, resulting in less distinctive experimental signatures (Belyaev et al., 2018).

### 3.1. Analysis Tool

Trees are widely used in various fields including mathematics, computer science, data sciences, and finance. However, the current file-based data representation does not allow for the addition or removal of branches from the original data, nor for the efficient extraction of branches from ROOT files. This leads to a convoluted approach of handling data, where tuples are called in a loop and ordering skimming or slimming is done. As a result, there is limited interactivity, portability, and unnecessary duplication of immutable branches, which in turn reduces effective use of storage space.

To overcome the constraints of the current file-based data representation, adopting a columnar database notion for input data representation is one solution. In a columnar data representation, caching and indexing can replace skimming and slimming. Each row of the table is assigned to the event number, and each column is assigned to a physics object, a process known as caching. The indexing process orders the data in decreasing physics objects.

Caching and indexing allow for the elimination of duplication of immutable columns, and for the addition or removal of columns to modify the event content without doing a tuple. This also enables array programming where the program can make the same operation to multiple events simultaneously, and eliminate event loops that are inefficient on the CPU. A schematic comparison of traditional and columnar ways of processing can be seen in Figure 3.2.

The Fermilab team has developed a cluster of software packages named Columnar Object Framework For Efficient Analysis (Coffea) to leverage the scientific python ecosystem and big data technologies in order to optimize analysis code development.

The columnar data representation is faster than the traditional event-loop

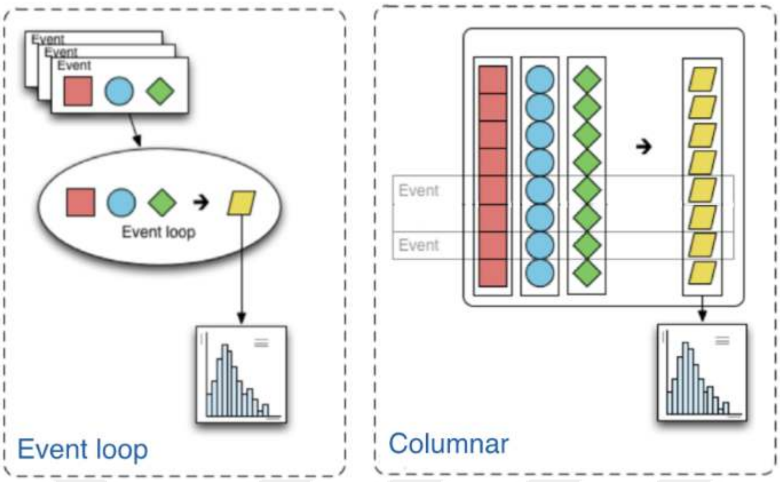



Figure 3.2. Comparison of the event-loop and columnar data processing paradigms (taken from (Smith et al., 2020)).

algorithm for processing data. It eliminates the need for duplicating immutable branches and allows for efficient addition or removal of columns. The program can make the same operation to multiple events simultaneously, eliminating the need for event loops that are inefficient on the CPU. This allows for efficient use of storage space and faster processing times, resulting in optimized analysis code development.

### 3.2. Monte Carlo Sample Generation

The MC samples for the mono-$h_s$ process shown in Figure 3.1. are generated with the `MadGraph5_aMC@NLO v2.6.0` (Alwall, Frederix, et al., 2014) using `NNPDF v3.0` (The NNPDF collaboration et al., 2015) for 2016 production and `NNPDF v3.1` (ibid.) for 2017/2018 production. `Pythia v8.2.4.0` (Sjöstrand et al., 2015) with the CUEP8M1 tune and the CP5 tune were used for hadronization and underlying events generation for 2016 and 2017/2018 productions, respectively. The samples are generated at leading order (LO). The $h_s$ boson mass hypotheses considered for this search are 50, 70, and 90 GeV. The mass grid in Table 3.1. shows the values of the mass of the DM particle and the mass of the $m_{Z'}$ bosons scanned in the signal generation.

Table 3.1. Summary of the $Z'$ boson and DM particle mass hypotheses used in the generation of the MC signal samples. The numbers in parentheses are the shifted $Z'$ masses used to avoid cases where the DM mass is exactly half the $Z'$ mass.

| $Z'$ mass (GeV) / DM mass (GeV) | 200(195) | 300(295) | 500(495) | 1000(995) | 2000(1995) | 2500(2495) | 3000(2995) |
|---|---|---|---|---|---|---|---|
| 100 | off-shell | on-shell | | | | | |
| 150 | off-shell | off-shell | on-shell | on-shell | | | |
| 250 | | | off-shell | | | | |
| 500 | | | off-shell | off-shell | on-shell | | |
| 750 | | | | | | on-shell | |
| 1000 | | | | off-shell | off-shell | | on-shell |
| 1250 | | | | | | off-shell | |
| 1500 | | | | | off-shell | | off-shell |

The `MadGraph5_aMC@NLO` generator, which uses a Matrix Element (ME) description, is less precise at describing partons than they are soft or collinear. Nevertheless, the `Pythia` hadronizer tends to break down when partons are hard and widely spread. For the MLM jet matching (Hoeche et al., 2005) procedure, a distance parameter $k_{\mathrm{T}}$ between partons in the $\eta - \phi$ space is used. It is defined as follows (Catani et al., 1993; Alwall, Höche, et al., 2008):

$$k_T = \sqrt{2 \min\left(p_{Ti}, p_{Tj}\right) \left[\cosh(\eta_i, \eta_j) - \cos(\phi_i, \phi_j)\right]} \tag{3.4}$$

where the indices $i, j$ run over the the partons. A requirement on $k_T$ to be larger than a value called "xqcut" is imposed. The xqcut value is the required parton separation at parton level. Events that satisfy such requirement are passed to `Pythia` for parton showering. After showering, but before hadronization and decays, the final-state partons are clustered into jets. These jets are then compared to the original partons from the matrix element event. A jet is considered to be matched to a parton if the

$k_\mathrm{T}$ between the parton and the jet is smaller then a value called "qcut". The event is rejected unless each jet is matched to a parton. For the dark Higgs model generation, the matching parameters xqcut and qcut have been set to be 15 GeV and 20 GeV respectively.

### 3.3. Summary of the Analysis Strategy

The analysis aims to identify a possible mono-$h_s$ signature by using two discriminating variables: missing transverse momentum ($p_\mathrm{T}^\mathrm{miss}$) and the mass of the recoiling large-cone jet. The $p_\mathrm{T}^\mathrm{miss}$ is expected to have a harder spectrum for the signal processes compared to the SM backgrounds, while the mass of the large-cone jet is expected to show peaks around the dark Higgs mass. Large-cone jets are clustered using the anti-$k_\mathrm{T}$ algorithm with a radius of $R = 1.5$, and the DeepAK15 algorithm is used to identify fat jets from the decay of an $h_s$ boson.

To model the SM backgrounds and constrain their uncertainties, a maximum-likelihood fit is performed simultaneously in a signal-enriched data sample (the signal region, SR) and control samples (control regions, CRs) obtained by inverting requirements on lepton/b-jet multiplicity and the DeepAK15 score. The models for the main backgrounds (i.e., $t\bar{t}$, Z+jets, and W+jets productions) are derived from data using bin-by-bin transfer factors. These factors modify the yields for a given background in a specific bin of the 2D $p_\mathrm{T}^\mathrm{miss}$-vs-mass distribution in SR, based on the data of the respective bin in CRs. The transfer factors are associated with statistical and systematic uncertainties.

Scale factors are derived to correct the efficiency of the DeepAK15 tagger on the different signal and background processes. These factors are obtained through standalone measurements in orthogonal samples or during the signal extraction in situ. The term "in situ" refers to a measurement that is performed within the same detector and under the same conditions as the primary analysis, rather than using an external measurement or simulation (CMS Collaboration, 2015).

### 3.4. Datasets and Triggers

The analysis of experimental data plays a crucial role in the search for new physics at the LHC. In this section, we describe the datasets and triggers used for the analysis, which are necessary for the study of the signal and the background processes. The data used in this analysis are certified events from the 2016, 2017, and 2018 LHC data-taking periods. These datasets correspond to a total integrated luminosity of 137.2 $\text{fb}^{-1}$, which is essential for obtaining accurate results. In addition, we present a list of trigger paths and datasets used for the analysis, which includes the MET and electron/photon datasets. The MET dataset includes muon pass-through triggers, which allow for the selection of events with high missing transverse energy ($p_{\mathrm{T}}^{\mathrm{miss}}$) or high muon recoil. These triggers are essential for the search for new physics beyond the SM. The study carried out uses certified events from the following data-taking periods:

- Run 2016B to 2016H datasets, `17Jul2018` reconstruction, corresponding to $35.9\,\text{fb}^{-1}$.
- Run 2017B to 2017F datasets, `31Mar2018` reconstruction, corresponding to $41.5\,\text{fb}^{-1}$.
- Run 2018A to 2018C datasets, `17Sep2018` reconstruction, corresponding to $59.74\,\text{fb}^{-1}$. ‘

The different reconstruction eras require to achieve a consistent reconstruction performance for the whole Run2 dataset. All those datasets correspond to a total integrated luminosity of $137.2\,\text{fb}^{-1}$.

The analysis uses CRs defined by muons and electrons, so the MET and electron/photon datasets are used. The MET dataset includes muon pass-through triggers, which allow us to select events with high $p_{\mathrm{T}}^{\mathrm{miss}}$ or high muon recoil. A list of all trigger paths and datasets is given in Tables 3.2.

Table 3.2. HLT paths, and the associated primary data used in the analysis, respectively era 2016, era 2017 and era 2018.

| HLT path | Primary dataset |
|---|---|
| **2016** | |
| HLT_PFMET170_* | MET |
| HLT_PFMETNoMu[X]_PFMHTNoMu[X]_IDTight | MET |
| HLT_Ele27_WPTight | SingleElectron |
| HLT_Ele105_CaloIdVT_GsfTrkIdT | SingleElectron |
| HLT_Photon165_HE10 | SinglePhoton |
| HLT_Photon175 | SinglePhoton |
| HLT_Photon200 | SinglePhoton |
| HLT_Photon[Y]_R9Id90_HE10_IsoM | SinglePhoton |
| HLT_Photon300_NoHE | SinglePhoton |
| HLT_ECALHT800 | SinglePhoton |
| HLT_CaloJet500_NoJetID | SinglePhoton |
| **2017** | |
| HLT_PFMETNoMu120_PFMHTNoMu120_IDTight | MET |
| HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60 | MET |
| HLT_Ele35_WPTight_Gsf | SingleElectron |
| HLT_Ele115_CaloIdVT_GsfTrkIdT | SingleElectron |
| HLT_Photon200 | SinglePhoton |
| **2018** | |
| HLT_PFMETNoMu120_PFMHTNoMu120_IDTight | MET |
| HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60 | MET |
| HLT_Ele35_WPTight_Gsf | EGamma |
| HLT_Ele115_CaloIdVT_GsfTrkIdT | EGamma |
| HLT_Photon200 | EGamma |

### 3.4.1. Trigger Efficiency and Scale Factors

The main data-taking trigger for signal selection in the 2017 and 2018 data is the logical OR of two HLT paths:

- HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
- HLT_PFMETNoMu120_PFMHTNoMu120_HT60.

Both triggers are completely unprescaled during those years. The path with the HT60 requirement also has different L1 seeding with respect to the former, with lower $p_{\mathrm{T}}^{\mathrm{miss}}$ thresholds while having an added $H_{\mathrm{T}} > 60\,\mathrm{GeV}$ requirement (i.e., the seeds with the _HTT60er suffix) for noise rate mitigation. Signal events always meet that extra requirement, and this strategy increases the search sensitivity with no negative effects.

The offline cuts are chosen such that the triggers are near 100% efficiency. Triggers are applied to data events, and to simulate the effect of MET and SingleElectron triggers on data, efficiency factors are applied to MC events. The measurement of these efficiencies is further described in Ref (Akchurin et al., 2016).lied to MC events. The measurement of these efficiencies is further described in Ref (ibid.).

As shown in Figure 3.3., it has been noticed by comparing the trigger efficiency measured in $\mathrm{Z} \to \mu^{+}\mu^{-}$selection and $\mathrm{W} \to \mu\nu$ selection (starting from events recorded by single-muon triggers) that these two measurements differed, which implies that application of the efficiency curve derived in the $\mathrm{W} \to \mu\nu$ CR to $\mathrm{Z} \to \mu^{+}\mu^{-}$ events might not be correct.

The difference between the trigger efficiency measured in the $\mathrm{Z} \to \mu^{+}\mu^{-}$ selection and $\mathrm{W} \to \mu\nu$ selection has been observed on both data and MC, likely due to the absence of prescale information in MC causing a difference in the turn-on position compared to data. To correct for this, simulated $\mathrm{Z} \to \mu^{+}\mu^{-}$ events

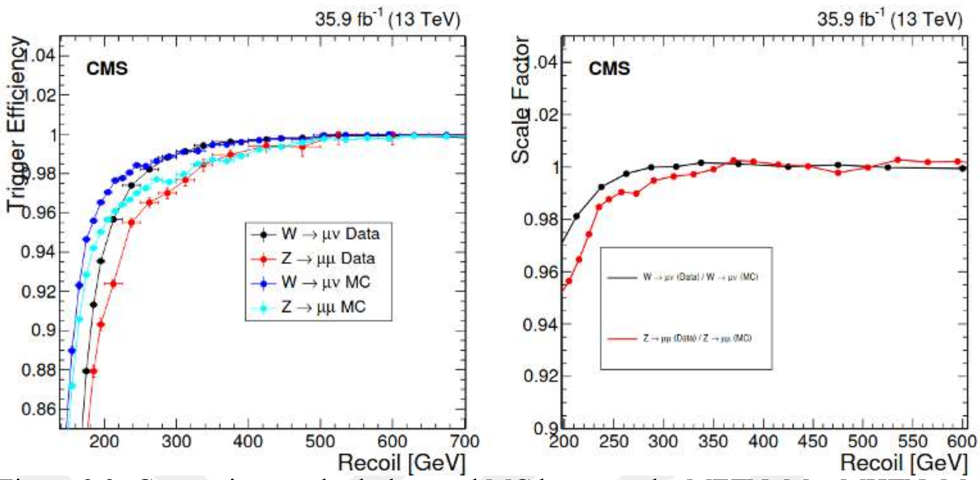



Figure 3.3. Comparison on both data and MC between the METNoMu+MHTNoMu trigger efficiency turn on as measured in $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ events (left). (Right) Comparison between the data to MC scale factors which show a significant disagreement in the low recoil region; taken from (Akchurin et al., 2016).

are corrected using the proper trigger efficiency derived from $Z \to \mu^+\mu^-$ candidates in data. The difference between the $W \to \mu\nu$ and the $Z \to \mu^+\mu^-$ trigger turn on is used as a systematic uncertainty due to the $p_T^{miss}$ triggers and applied to the transfer factors.

Figure 3.4. shows a comparison between the data to MC scale factors obtained computing the MET trigger turn-on in $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ events, which nicely agree over the full recoil spectrum after requiring to have several HLT muons equal to the number of the identified offline muons in the event.

For the single and double electron CRs, events that pass a logical OR of different electron triggers are selected. The electron triggers used depend on the $p_T$ threshold and employ different criteria for the identification and isolation. Low-$p_T$, isolated electron triggers include `Ele27`, `Ele32`, or `Ele35`, while high-$p_T$, non-isolated electron triggers are `Ele105` or `Ele115`, depending on the year.

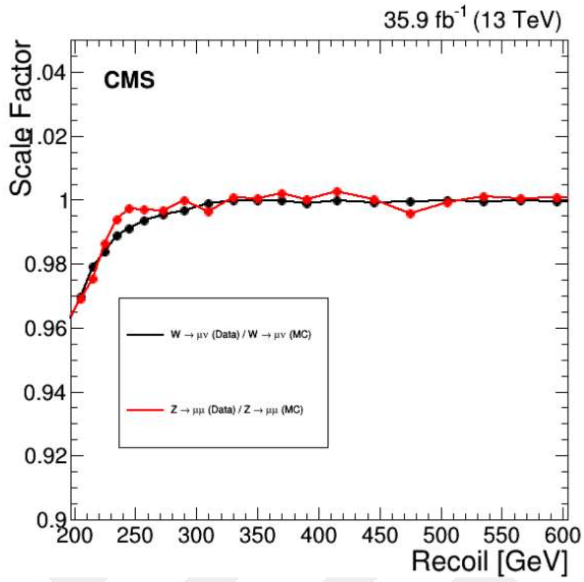



Figure 3.4. Comparison between the data to MC scale factors as obtained computing the MET trigger turn-on in $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ events, which nicely agree over the full recoil spectrum after requiring to have several HLT muons equal to the number of the identified offline muons in the event; taken from (Akchurin et al., 2016).

The low-$p_T$ electron trigger can suffer from inefficiency for $Z \to ee$ events in which the Z boson is boosted, where the two electrons can spoil each other's isolation at the trigger. To recover this inefficiency, the high-$p_T$ trigger, which has no isolation requirement, is added. In 2017, an ECAL reconstruction bug led to a wrong choice on the ADC gain switch. To address this, the CMS electron and photon Physics Object Group (EGamma POG) enforces the usage of a logical OR with the `Photon200` path for analyses that use electrons with $p_T > 300\,\mathrm{GeV}$.

The efficiency of the electron triggers is measured using two approaches in data. The tag-and-probe technique is used, where the trigger efficiency is measured on an electron leg of the Z boson. Events are selected requiring one tag electron, which fires the single electron triggers and passes the tight identification requirements. Then a second probe electron is also required in the event, which is also required to pass the tight selection, and the invariant mass of the dielectron pair is necessary to be consistent with the Z peak. The trigger efficiency is measured in bins of the $p_T$ and η of the electron. However, at high electron $p_T$, there are not enough Z events to use this method, so the electron efficiency is also measured in events from the JetHT primary data set passing the PFHT800 trigger. For electrons with $p_T$ less than 100 GeV, the trigger efficiency is taken from the Z tag-and-probe measurement, while for electrons with $p_T$ greater than 100 GeV, the trigger efficiency is taken from the JetHT data.

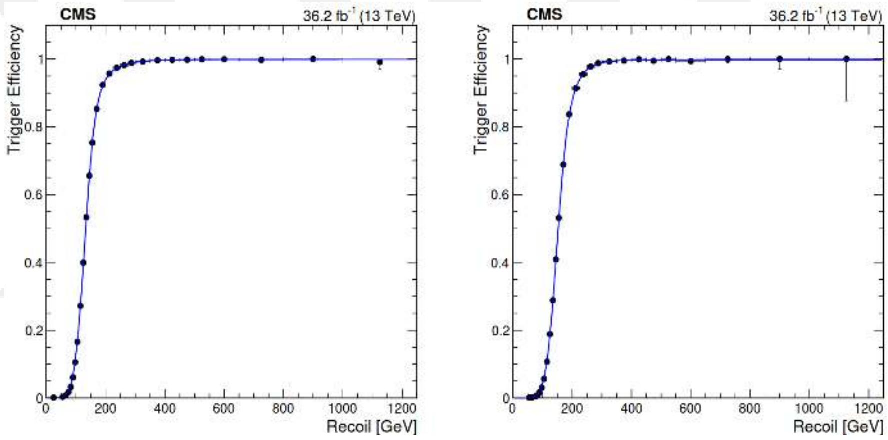



Figure 3.5. MET trigger turn-on curve measured in single muon events (left) and single electron events (right). In the case of the single muon events, the trigger turn-on is computed as a function of the recoil ($p_T^{miss}$ without including the muon), while in the case of the single electron events, it is computed as a function of $p_T^{miss}$. Both measurements agree at the 1-2% level at higher MET, and the efficiencies from the measurement in single muon events are applied to the simulated events in the analysis, taken from (Akchurin et al., 2016).

Figure 3.5. shows the MET trigger turn-on curve measured in single muon events (left) and single electron events (right). In the case of the single muon events,

the trigger turn-on is computed as a function of the recoil ($p_{\mathrm{T}}^{\mathrm{miss}}$ without including the muon), while in the case of the single electron events, it is computed as a function of $p_{\mathrm{T}}^{\mathrm{miss}}$. Both measurements agree at the 1-2% level at higher MET, and the efficiencies from the measurement in single muon events are applied to the simulated events in the analysis.

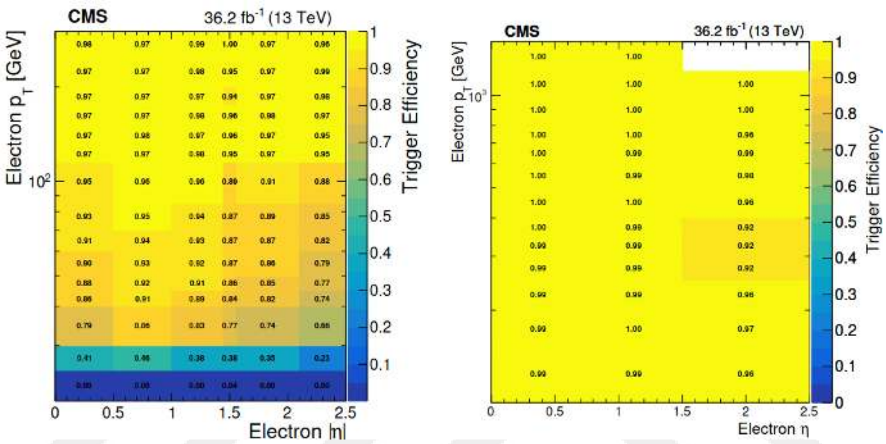



Figure 3.6. Single electron trigger efficiencies as a function of electron $p_{\mathrm{T}}$ and $\eta$. The left plot shows the trigger efficiencies measured using Z tag-n-probe starting with electron $p_{\mathrm{T}}$ of 20 GeV. The right one shows the electron trigger efficiency measured in the JetHT dataset using events passing the PFHT650 or PFHT800 trigger, starting at electron $p_{\mathrm{T}}$ of 100 GeV; taken from (Akchurin et al., 2016).

The electron trigger efficiencies as a function of electron $p_{\mathrm{T}}$ and $\eta$ are shown in Figure 3.6. The left plot shows the trigger efficiencies measured using Z tag-and-probe starting with electron $p_{\mathrm{T}}$ of 20 GeV, while the right plot shows the electron trigger efficiency measured in the JetHT dataset using events passing the PFHT650 or PFHT800 trigger, starting at electron $p_{\mathrm{T}}$ of 100 GeV. The figure is taken from (Akchurin et al., 2016).

For 2017 and 2018, preliminary studies from the monojet group show that an offline 250 GeV recoil threshold can be maintained. The reference triggers are:

- Electron channel in 2017: `HLT_Ele35_WPTight_Gsf OR HLT_Photon200`
- Electron channel in 2018: `HLT_Ele32_WPTight_Gsf OR HLT_Photon200`
- Muon channel in 2017: `HLT_IsoMu27`
- Muon channel in 2018: `HLT_IsoMu24`

### 3.4.2. Background Simulation Samples

MC simulation samples for the SM background are primarily used to cross-check the analysis in CRs and to optimize the selection procedure, which were generated centrally by the CMS collaboration. In general lines, the background to this search are those from SM processes that have real $p_{\mathrm{T}}^{\mathrm{miss}}$ Z in the final state:

- $\mathrm{Z} \to \nu\nu + \mathrm{jets}$
- $\mathrm{W} \to \ell\nu + \mathrm{jets}$
- semileptonic and fully leptonic $t\bar{t}$
- semileptonic and fully leptonic diboson $\mathrm{WW}$, $\mathrm{WZ}$ and ZZ

There are also backgrounds where the $p_{\mathrm{T}}^{\mathrm{miss}}$ comes primarily from mismeasurements of the reconstructed particle momenta in the event. Those include

- multijet production, a.k.a ”QCD”
- $\gamma + \mathrm{jets}$

Finally, DY+jets samples with electrons or muons in the final state are used for CR analysis. The choice of which MC samples will be used to study each process is analysis-dependent. Currently, there are two main options: leading-order (LO) samples, possibly augmented by next-to-leading order (NLO) corrections, a.k.a ”k-factors,” and examples produced directly at NLO. A list of all background MC samples and the corresponding cross-sections are given in Tables 3.4., 3.5. and Tables 3.6., 3.7.

Table 3.4. List of cross sections and their accuracies for various background processes. (2016) Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied. Cross-sections marked with † are (N)NLO

| Dataset name | σ [pb] |
|---|---|
| /QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1735000 |
| /QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 366800 |
| /QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 29370 |
| /QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 6524 |
| /QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1064 |
| /QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 121.5 |
| /QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 25.42 |
| /ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | $136.02^{\dagger}$ |
| /ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | $80.95^{\dagger}$ |
| /ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | $35.85^{\dagger}$ |
| /ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | $35.85^{\dagger}$ |
| /ZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8 | 0.08912 |
| /ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8 | 0.04865 |
| /ggZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8 | 0.014366 |
| /ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8 | 0.007842 |
| /WminusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8 | 0.100 |
| /WplusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8 | 0.159 |
| /ttHTobb_M125_13TeV_powheg_pythia8 | $0.506*0.5824$ |
| /TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 | $831.76^{\dagger}$ |
| /WW_TuneCUETP8M1_13TeV-pythia8 | $118.7^{\dagger}$ |
| /WZ_TuneCUETP8M1_13TeV-pythia8 | $47.2^{\dagger}$ |
| /ZZ_TuneCUETP8M1_13TeV-pythia8 | $16.6^{\dagger}$ |

Table 3.5. List of cross sections and their accuracies for various background processes. (2016) Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied. Cross-sections marked with † are (N)NLO

| Dataset name | σ [pb] |
|---|---|
| /DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 148.0 |
| /DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 40.94 |
| /DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 5.497 |
| /DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 1.367 |
| /DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 0.6304 |
| /DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 0.1514 |
| /DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 0.003565 |
| /WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 1343 |
| /WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 359.6 |
| /WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 48.85 |
| /WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 12.05 |
| /WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 5.501 |
| /WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 1.329 |
| /WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8* | 0.03216 |
| /ZJetsToNuNu_HT-100To200_13TeV-madgraph* | 280.5 |
| /ZJetsToNuNu_HT-200To400_13TeV-madgraph* | 77.7 |
| /ZJetsToNuNu_HT-400To600_13TeV-madgraph* | 10.71 |
| /ZJetsToNuNu_HT-600To800_13TeV-madgraph* | 2.559 |
| /ZJetsToNuNu_HT-800To1200_13TeV-madgraph* | 1.183 |
| /ZJetsToNuNu_HT-1200To2500_13TeV-madgraph* | 0.292 |
| /ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph* | 0.0069 |

Table 3.6. List of cross sections and their accuracies for various background processes, 2017/2018 versions. These samples were produced with the CP5 tune, so the cross-sections are slightly different. Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied. Cross-sections marked with † are (N)NLO.

| Dataset name | σ [pb] |
|---|---|
| ZJetsToNuNu_HT-100To200_13TeV-madgraph* | 302.8 |
| ZJetsToNuNu_HT-200To400_13TeV-madgraph* | 92.59 |
| ZJetsToNuNu_HT-400To600_13TeV-madgraph* | 13.18 |
| ZJetsToNuNu_HT-600To800_13TeV-madgraph* | 3.257 |
| ZJetsToNuNu_HT-800To1200_13TeV-madgraph* | 1.49 |
| ZJetsToNuNu_HT-1200To2500_13TeV-madgraph* | 0.3419 |
| ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph* | 0.005146 |
| WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8* | 1319 |
| WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8* | 1395 |
| WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8* | 407.9 |
| WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8* | 57.48 |
| WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8* | 12.87 |
| WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8* | 5.366 |
| WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8* | 1.074 |
| WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8* | 0.008001 |
| TTJets_TuneCP5_13TeV-madgraphMLM-pythia8 | $831.76^{\dagger}$ |
| ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8 | $80.95^{\dagger}$ |
| ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8 | $136.02^{\dagger}$ |
| ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 | $35.85^{\dagger}$ |
| ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 | $35.85^{\dagger}$ |
| ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8 | $3.36^{\dagger}$ |
| WW_TuneCP5_13TeV-pythia8 | $118.7^{\dagger}$ |
| WZ_TuneCP5_13TeV-pythia8 | $43.5^{\dagger}$ |
| ZZ_TuneCP5_13TeV-pythia8 | $19.1^{\dagger}$ |

Table 3.7. List of cross sections and their accuracies for various background processes, 2017/2018 versions. These samples were produced with the CP5 tune, so the cross-sections are slightly different. Datasets marked with * are LO in QCD and EWK but will have NLO corrections applied. Cross-sections marked with † are (N)NLO.

| Dataset name | σ [pb] |
|---|---|
| QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8 | 23700000 |
| QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8 | 1547000 |
| QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8 | 322600 |
| QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8 | 29980 |
| QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8 | 6334 |
| QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8 | 1088 |
| QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8 | 99.11 |
| QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8 | 20.23 |
| GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8 | 18700 |
| GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8 | 8640 |
| GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8 | 2185 |
| GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8 | 260 |
| GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8 | 85.31 |
| DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8* | 143.0 |
| DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8* | 161.1 |
| DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8* | 48.66 |
| DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8* | 6.968 |
| DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8* | 1.743 |
| DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8* | 0.8052 |
| DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8* | 0.1933 |
| DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8* | 0.003468 |

Table 3.8. NLO $k$ factors for QCD and EW corrections for the W+jets simulation, depending on the $p_T$ of the generated W boson.

| From W $p_T$ | to W $p_T$ | QCD $k$ factor | EW $k$ factor |
|---|---|---|---|
| 150 | 170 | 1.46628 | 0.966481 |
| 170 | 200 | 1.4947 | 0.958351 |
| 200 | 230 | 1.43412 | 0.948513 |
| 230 | 260 | 1.45026 | 0.938304 |
| 260 | 290 | 1.3519 | 0.929394 |
| 290 | 320 | 1.35886 | 0.919688 |
| 320 | 350 | 1.37136 | 0.910425 |
| 350 | 390 | 1.34218 | 0.900288 |
| 390 | 430 | 1.31646 | 0.889122 |
| 430 | 470 | 1.23866 | 0.877255 |
| 470 | 510 | 1.23631 | 0.86653 |
| 510 | 550 | 1.20969 | 0.856032 |
| 550 | 590 | 1.17554 | 0.84536 |
| 590 | 640 | 1.10773 | 0.834675 |
| 640 | 690 | 1.07296 | 0.822288 |
| 690 | 740 | 1.07115 | 0.810779 |
| 740 | 790 | 1.03965 | 0.799689 |
| 790 | 840 | 1.00647 | 0.788566 |
| 840 | 900 | 0.999332 | 0.777413 |
| 900 | 960 | 0.959924 | 0.765405 |
| 960 | 1020 | 0.942234 | 0.753711 |
| 1020 | 1090 | 0.896176 | 0.742262 |
| 1090 | 1160 | 0.858749 | 0.72925 |
| 1160 | 1250 | 0.825848 | 0.716423 |

Table 3.9. NLO *k* factors for QCD and EW corrections for the Z+jets simulation, depending on the $p_T$ of the generated Z boson.

| From Z $p_T$ | to Z $p_T$ | QCD $k$ factor | EW $k$ factor |
|---|---|---|---|
| 150 | 170 | 1.51661 | 0.970592 |
| 170 | 200 | 1.51565 | 0.964424 |
| 200 | 230 | 1.46501 | 0.956695 |
| 230 | 260 | 1.42989 | 0.948747 |
| 260 | 290 | 1.36459 | 0.941761 |
| 290 | 320 | 1.3586 | 0.934246 |
| 320 | 350 | 1.36676 | 0.927089 |
| 350 | 390 | 1.32941 | 0.919181 |
| 390 | 430 | 1.37169 | 0.909926 |
| 430 | 470 | 1.2845 | 0.900911 |
| 470 | 510 | 1.26057 | 0.892561 |
| 510 | 550 | 1.2408 | 0.884353 |
| 550 | 590 | 1.21695 | 0.8761 |
| 590 | 640 | 1.19336 | 0.867687 |
| 640 | 690 | 1.12552 | 0.858047 |
| 690 | 740 | 1.0768 | 0.849014 |
| 740 | 790 | 1.05538 | 0.840317 |
| 790 | 840 | 1.03528 | 0.832017 |
| 840 | 900 | 1.0216 | 0.823545 |
| 900 | 960 | 0.965127 | 0.814596 |
| 960 | 1020 | 0.959297 | 0.806229 |
| 1020 | 1090 | 0.927266 | 0.798038 |
| 1090 | 1160 | 0.89461 | 0.789694 |
| 1160 | 1250 | 0.840332 | 0.781163 |

## 3.5. Physics Objects

All major physics objects are used in this analysis to identify signal-like events, suppress backgrounds, and define control regions in data for background estimation. For leptons, photons, and taus, we adhere to CMS POG-endorsed recommendations. As jets drive our analysis selection, we have made choices that optimize the sensitivity of the analysis. A summary of all physics objects and the selection requirements imposed on them are described below, followed by a section on identifying events with dark Higgs boson candidates.

### 3.5.1. Missing Transverse Momentum and Hadronic Recoil

The $p_{\mathrm{T}}^{\mathrm{miss}}$ used in the analysis is calculated by taking the negative vector sum of the transverse momenta of all Particle Flow (PF) candidates reconstructed in a given event. Our analysis uses reconstructed AK4 jets (Strologas, 2016) and only charged PF candidates associated with the selected primary vertex are used to cluster the jets(Charged Hadron Subtraction, CHS). Corrections to the momenta of AK4 PF CHS jets reconstructed in the event are further propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ (*Pileup Removal Algorithms* 2014). To imitate the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the signal regions, control regions are constructed by selecting events with a lepton. The hadronic recoil against the bosons can be computed by removing the $\vec{p}_T$ of the lepton from the $p_{\mathrm{T}}^{\mathrm{miss}}$ computation. Formally, the recoil is given by:

$$\vec{U} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T}}^{\ell} \tag{3.5}$$

The hadronic recoil vector $\vec{U}$ represents the momentum of the system recoiling against the lepton or boson in the event. The magnitude of the hadronic recoil vector $u_T$ is an important variable used in the analysis, as it provides information on the energy imbalance in the event due to the presence of neutrinos, which is a characteristic signature of the dark Higgs boson decay. The $p_{\mathrm{T}}^{\mathrm{miss}}$ vector $\vec{\mathrm{p}}_{T}^{\mathrm{miss}}$, which is used in the analysis, is related to the hadronic recoil vector $\vec{U}$ by subtracting the $\vec{p}_T$ of the lepton from the $p_{\mathrm{T}}^{\mathrm{miss}}$ computation. The last term is the $\vec{p}_{\mathrm{T}}$ of the muon or electron, depending on the control region. In the case of $\mathrm{Z} \rightarrow \nu\nu$ decays, $\vec{U} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. In this sense, the control regions' recoil can mimic the $p_{\mathrm{T}}^{\mathrm{miss}}$ in $\mathrm{Z} \rightarrow \nu\nu$ events that pass the signal selection.

Events in data can end up with large spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ due to detector noise and beam backgrounds. In order to eliminate these events, several $p_{\mathrm{T}}^{\mathrm{miss}}$ filters listed below were used, which were recommended and provided by the CMS JetMET POG (CMS JetMET POG, 2018). and have been implemented in the analysis. Unless stated otherwise, filters are applied to all data, background MC and signal MC events.

- Good Vertex Filter (At least one good primary vertex in the event)
- Global Beam Halo SuperTight Filter
- HBHE Noise Filter
- HBHE NoiseIso Filter
- ECAL Dead Cells Trigger Primitive Filter
- Bad PF Muon Filter
- EE Bad SC Filter
- ECAL Bad Calibration Filter. We currently apply this filter by vetoing events in a large list centrally compiled. An alternative would be to rerun the filter on MINIAOD using a reduced EE rechits collection.

### 3.5.2. AK4 Jets

Jets used in the analysis are reconstructed by clustering PF candidates in the event using the anti-$k_{\mathrm{T}}$ algorithm (Cacciari, Salam, and Soyez, 2008) with a distance parameter $(\Delta R) = 0.4$ for ordinary AK4 jets. To mitigate the effect of pileup, charged hadrons that do not arise from the primary vertex are removed using the CHS method, which greatly reduces the number of jets originating from pileup.

Jets are corrected for pileup using the event-by-event energy density (L1FastJet corrections). Further corrections are applied as a function of jet pseudorapidity ($\eta$) and transverse momentum (L2, L3 corrections). Further residual jet energy corrections are applied to jets in data to match their response with the simulation (L2L3Residual). These corrections have been derived on AK4 CHS jets (CMS JetMET POG, 2022) and are specifically the versions listed in Table 3.10.. Narrow jets (i.e., AK4 jets) considered in the analysis are required to:

- have $p_T > 30\,\text{GeV}$, $|\eta| < 2.4$;
- pass standard tight jet identification criteria (CMS JetMET POG, 2017);
- have neutral hadron fraction $< 0.8$;
- have neutral EM fraction $< 0.99$;
- have charged hadron fraction $> 0.1$;
- have charged EM fraction $< 0.99$;
- pass the Loose PU ID if they have $p_T < 50\,\text{GeV}$.

The explicit selections on the energy fractions discussed above may be overridden by selections contained in the ”tight jet identification” criteria.

In regions where a lepton is required, we reject a jet if $\Delta R(\text{jet}, \ell) < 0.4$. Jets are also required to be well-separated from fat jets (see below), so we veto a jet if $\Delta R(\text{jet}, \text{fat jet}) < 1.5$. In this way, all AK4 jets in the analysis are ”extra jets” with respect to the fat jet that will make up the Dark Higgs candidate in the event. AK4 jets may also be b-tagged, where the definition is DeepFlavour loose WP (score $>$ 0.0494).Jet control plots are shown in Fig. 3.7..

Table 3.10. shows the jet energy corrections applied to anti-$k_T$ algorithm with a distance parameter of 0.4 (AK4) jets clustered from particle-flow candidates in the

event, which have been corrected for pileup using the event-by-event energy density and further corrected for jet pseudorapidity and transverse momentum, as well as residual corrections applied to match their response with the simulation. Only narrow AK4 jets with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$, and passing standard tight jet identification criteria, are considered in the analysis. They are required to have a neutral hadron fraction less than 0.8, a neutral EM fraction less than 0.99, a charged hadron fraction greater than 0.1, and a charged EM fraction less than 0.99. Additionally, they must pass the Loose PU ID if they have $p_T < 50\,\mathrm{GeV}$. These jet selections may be overridden by selections contained in the ”tight jet identification” criteria. In regions where a lepton is required, a jet is rejected if $\Delta R(\mathrm{jet}, \ell) < 0.4$, and if it is not well-separated from ”fat jets” with a veto $\Delta R(\mathrm{jet}, \mathrm{fat\,jet}) < 1.5$. Finally, AK4 jets may also be b-tagged, where the definition is DeepFlavour loose WP (score > 0.0494).

Table 3.10. Jet Energy Corrections for 2016, 2017, 2018 simulated and collision datasets.

| Sample | Correction |
|---|---|
| Simulation 2016 | `Summer16_07Aug2017_V11_MC` |
| Simulation 2017 | `Fall17_17Nov2017_V32_94X_MC` |
| Simulation 2018 | `Autumn18_V8_MC` |
| Data 2016 | `Summer16_07Aug2017_V11` |
| Data 2017 | `Fall17_17Nov2017_V32` |
| Data 2018 | `Autumn18_RunABCD_V8` |

### 3.5.3. AK15 Fat Jets

To identify the dark Higgs candidate, we cluster jets from PUPPI-weighted PF candidates using the anti-$k_T$ algorithm with a distance parameter of 1.5 (AK15 jets), also known as ”fat jets”. The choice of AK15 jets is motivated by their larger radius, which allows for the capture of more particles within the jet and the potential identification of substructure within the jet.

The fat jets have to be ”dark Higgs tagged” in order to make up a good candidate. The usual handles are the jet mass and the B-hadron content of the jet. For the former, we deploy grooming methods that attempt to remove soft and wide-angle radiation inside a jet produced by pileup interactions, underlying events, or parton shower activity. The soft drop method (A. Larkoski et al., 2014) has been used in this analysis to produce the groomed fat jet. It is performed with parameters $z_{cut} = 0.15$ and $\beta = 1$, which were chosen to optimize the resolution of the mass after grooming.

For the latter, we employ deep-learning-based techniques (DeepFlavour) while trying to retain the independence of the jet mass variable. The full discussion is in Sec. 3.6.. However, we draw attention to the fact that we use the kinematics of the groomed jet throughout the analysis – the $p_{\mathrm{T}}$ requirement, for instance, is applied on the groomed jet.

Fat jets considered in the analysis must have $p_{\mathrm{T}} > 160\,\mathrm{GeV}$ and $|\eta| < 2.4$. Furthermore, to remove jets due to HCAL noise or mismeasurement, we require that a fat jet passes the tight jet ID requirements described in (Akchurin et al., 2016). In regions where a lepton is required, we reject a fat jet if $\Delta R(\mathrm{jet}, \ell) < 1.5$. Jets are corrected using the L1, L2, L3, and L2L3 residual corrections derived for AK8 jets.

Fat jet control plots are shown in Fig. 3.8..

### 3.5.4. Muons

Events containing muons are rejected in the analysis to suppress electroweak backgrounds. To be considered a “loose” muon, an object must pass the following selection:

- $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ (2016 and 2017), $15\,\mathrm{GeV}$ (2018);

- $|\eta| < 2.4$;
- "Loose Muon" identification criteria, the muons are required to pass the Muon POG recommended medium ID (POG, 2023);
- $|d_{xy}| < 0.5\,\mathrm{cm}$ - this selection is explicitly added on top of the loose muon to ensure that the muon originates from the primary vertex in the transverse plane.
- $|d_z| < 1.0\,\mathrm{cm}$ – this selection is explicitly added on top of the loose muon to ensure that the muon originates from the primary vertex along the beam axis.
- Particle-Flow relative isolation in cone of $\Delta R = 0.4$, $I_{\mathrm{rel}} < 0.4$, the "very loose" working point, with an efficiency above 98%.

To account for differences in muon reconstruction and identification between data and simulation, an event scale factor is applied to MC. The derivation of this SF and the corresponding systematic is described in (Akchurin et al., 2016).

To account for differences in muon reconstruction and identification between data and simulation, an event scale factor is applied to MC. The derivation of this SF and the corresponding systematic is described in (ibid.).

#### 3.5.5. Electrons

As with muons, we use two definitions of electrons for veto and selection. To be considered a "loose" electron, an object must pass the following selections:

- $p_{\mathrm{T}} > 10\,\mathrm{GeV}$;
- $|\eta_{\mathrm{SC}}| < 2.5$;
- Loose electron identification requirements (CMS Collaboration, 2018): For this round of analysis, we use the `Electron_mvaFall17V2Iso_WP90` working point, MultiVariate classifier Algorithms (MVAs) based on an identification that includes isolation inputs and has an efficiency of 90%. The MVA-based algorithm is used for electron identification. We used the cuts recommended

by the CMS EGamma POG and applied them to all of the 2016, 2017, and 2018 data. The loose electrons also need to align with the leading vertex of the event, providing a loose cut on the impact parameter significance, which is the error divided by the closest distance from the peak of the track.

- Separated from loose muons ($\Delta R > 0.3$).

”Tight” electrons must pass the following cuts:

- $p_{\mathrm{T}} > 29\,\mathrm{GeV}$ (2016), $40\,\mathrm{GeV}$ (2017 and 2018);
- $|\eta_{\mathrm{SC}}| < 2.5$;
- Tight electron identification requirements (CMS Collaboration, 2018): It incorporates a slightly higher $p_{\mathrm{T}}$ cut and the barrel-endcap transition region. For this round of analysis, we use the `Electron_mvaFall17V2Iso_WP80` working point, an MVA-based identification that includes isolation inputs and has an efficiency of 80%. As with the loose electron selection, we used the cuts recommended by the CMS EGamma POG and applied them to all data of 2016, 2017, and 2018.
- Separated from loose muons ($\Delta R > 0.3$).

Whenever we refer to ”loose” leptons for selection (for the dilepton control regions), loose requirements are the same as the veto definition given above. Once again, an event scale factor is applied to account for differences in electron reconstruction and ID between data and MC, as described in (Akchurin et al., 2016).

### 3.5.6. Photons

Other identification steps are based on the ratio of energy deposits between HCAL and ECAL, shower shape variables, and isolations. Photon veto is applied in the analysis to suppress backgrounds ($Z(\nu\nu)+\gamma$ + jet, $W(l\nu) + \gamma$ + jet) and ensure clean separation from the mono-photon final state. In this analysis, we reject events with any photons passing the loose identification, defined as follows:

- $p_T > 15\,\mathrm{GeV}$;
- $|\eta| < 2.5$;
- Loose photon selection (CMS EGamma POG, 2015). The cut-based loose ID (94X-V2) recommended by EGamma POG was applied at the loose working point, with an average efficiency of 90%. The photon sample was used to estimate the Z(νν)+jets contribution to the signal region.
- A **veto** on electrons – i.e., an identified photon must not also be an electron. Reconstructed electrons that are compatible with the same supercluster are vetoed. We use the `Photon_electronVeto` selection.

### 3.5.7. Taus

Taus can decay leptonically into neutrinos and an electron or muon about 35% of the time, which can mimic the final state of interest. Therefore, events containing tau leptons are removed to suppress electroweak backgrounds.

A loose selection is applied to reject taus as follows:

- $p_T > 18\,\mathrm{GeV}$: Taus with transverse momentum less than 18 GeV are rejected since they are unlikely to pass the analysis requirements and contribute to the signal.
- $|\eta| < 2.3$: Only taus with pseudorapidities less than 2.3 are selected since the CMS detector has better coverage in this region and better tau reconstruction performance.
- **Old** DecayModeFinding: The DecayModeFinding algorithm is used to identify taus that decay hadronically. Since we are neither using DeepTau v2 nor particularly high-$p_T$ taus, there is no particular reason to use the new algorithm, and we use the old one instead.

- MVA-based very loose isolation provided by the POG (CMS Tau POG, 2016): Tau isolation is a measure of the hadronic activity around the tau candidate, and it is used to reduce the contribution from QCD multijet events. The isolation variable is computed using the energy deposited in the calorimeter and the tracks associated with the tau candidate. In this analysis, we use a very loose working point provided by the tau POG and labeled as `"byVLooseIsolationMVArun2017v2DBoldDMwLT2017"`. This working point has an efficiency of about 99%, which ensures high signal acceptance.
- Separated from loose muons ($\Delta R > 0.4$) and from loose electrons ($\Delta R > 0.4$): To ensure that taus do not overlap with other leptons in the event, a separation requirement is applied. Taus that are within a cone of 0.4 in the $\Delta R$ distance measure with respect to loose muons or loose electrons are rejected.

By applying this loose tau selection, we reject events containing taus, which reduces the contribution from electroweak backgrounds and improves the purity of the signal sample.

### 3.5.8. Preselection Summary

The goal of the preselection is to define a set of events that will be used to construct the signal region and various control regions. The preselection criteria are summarized in Table 3.11. and include:

- AK15 jet $p_{\mathrm{T}}$ threshold $> 160$ GeV: Events are required to have at least one AK15 jet with a transverse momentum greater than 160 GeV.
- $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold (signal region): Events are required to have missing transverse momentum greater than a threshold value that is specific to the signal region.
- Recoil $|\tilde{U}|$ threshold (control region): Events are required to have recoil $|\tilde{U}|$

greater than a threshold value that is specific to each control region.

- $p_T$ threshold for AK4 jets $> 30$ GeV: Events are required to have at least one AK4 jet with a transverse momentum greater than 30 GeV.
- Pass Loose PU ID if $p_T < 50$ GeV: AK4 jets with transverse momentum less than 50 GeV are required to pass the loose pileup identification criteria.
- $\Delta\phi$(jet, $p_T^{miss}$ or $\tilde{U}$) $> 0.8$: The azimuthal angle between the leading AK4 jet and the missing transverse momentum or recoil must be greater than 0.8 radians to reduce QCD multijet backgrounds.
- Loose tau leptons vetoed: Events containing tau leptons that pass the loose selection criteria are vetoed to reduce electroweak backgrounds.

Table 3.11. Summary of the preselection criteria used in the analysis.

| | |
|---|---|
| AK15 jet $p_T$ threshold | $> 160$ GeV |
| $p_T^{miss}$ threshold (signal region)<br>Recoil $\lvert\vec{U}\rvert$ threshold (control region) | $> 250$ GeV |
| AK4 jet $p_T$ threshold | $> 30$ GeV |
| Pass Loose PU ID | if $p_T < 50$ GeV |
| $\Delta\phi(\text{jet}, \vec{p}_T^{miss}$ or $\vec{U})$ | $\Delta\phi > 0.8$ |
| Loose tau leptons | Vetoed |

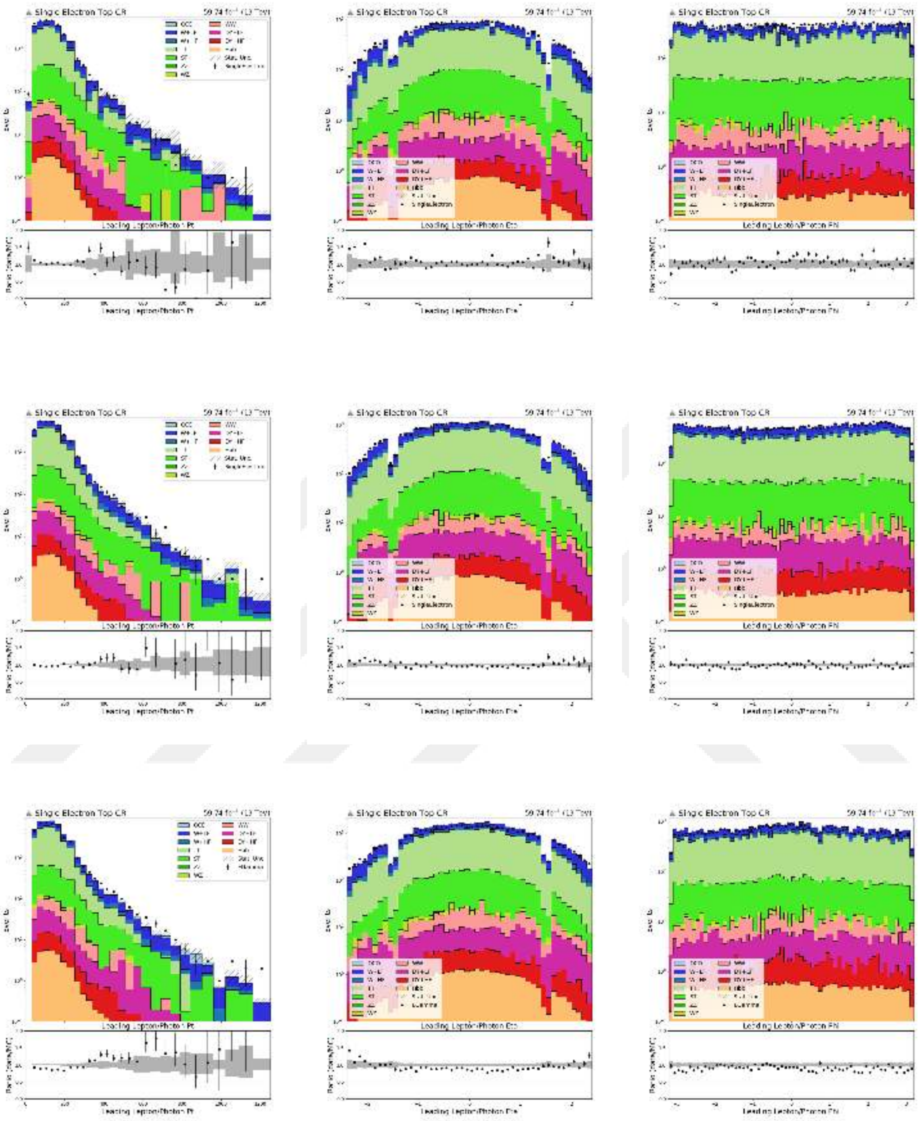


Figure 3.7. Distribution of jet transverse momentum $p_T$, pseudorapidity $\eta$, and azimuthal angle $\phi$ for the years 2016, 2017, and 2018, respectively, after applying the preselection criteria.

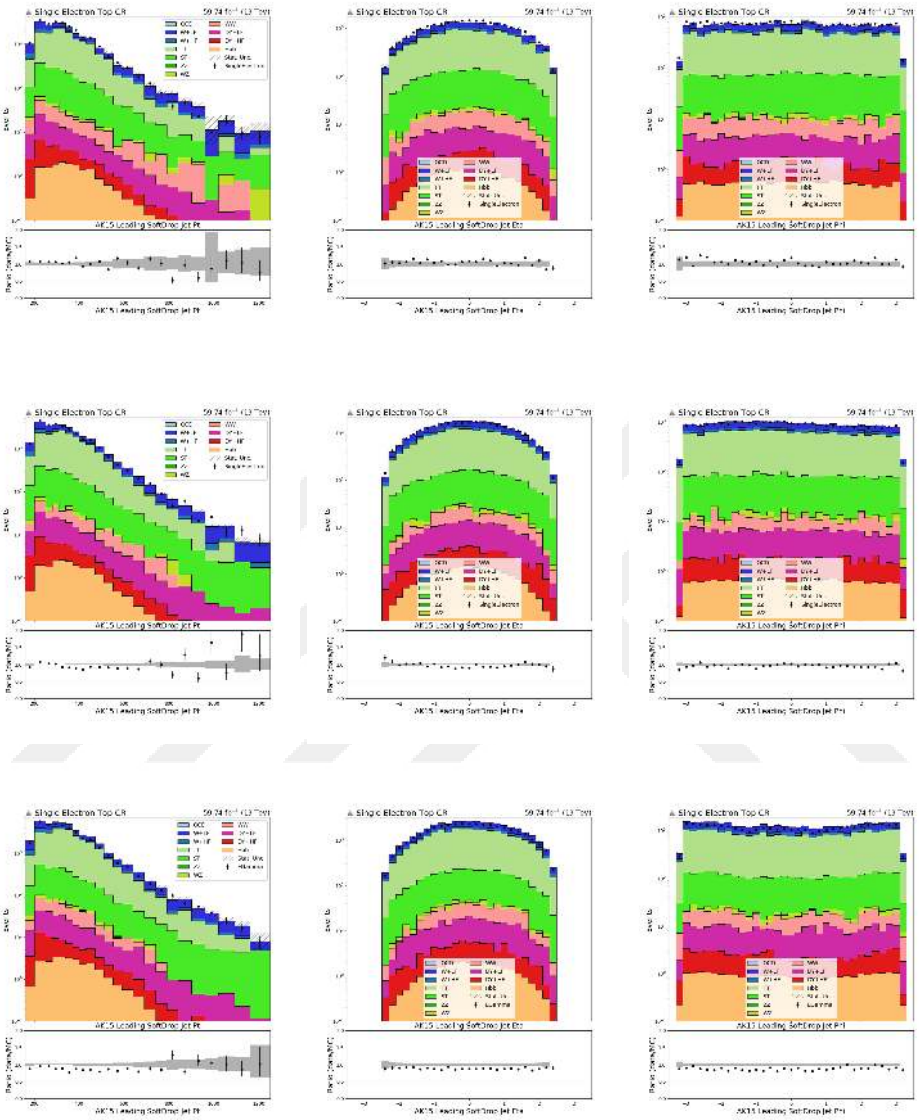


Figure 3.8. Fat jet $p_T$, η and ϕ for 2016 (top), 2017 (middle), and 2018, after the preselection.

### 3.6. DeepAK15

This section describes the use of the DeepAK15 jet tagger for the reconstruction of the dark Higgs candidate in the CMS experiment. The dark Higgs candidate is reconstructed through its mixing with the SM Higgs boson and subsequent decay in the $\mathrm{H} \to \mathrm{b\overline{b}}$ channel. The analysis aims to address two main classes of background: processes containing light-flavour hadronic decays, but that have the right mass scale (e.g. $\mathrm{W} \to \mathrm{q\overline{q}}$, for a dark Higgs mass around 80 GeV); and processes containing real $\mathrm{b\overline{b}}$ hadronic decays, but at the wrong mass scale (e.g. pure $\mathrm{g} \to \mathrm{b\overline{b}}$ splitting in multijet events).

To reduce background, the analysis focuses on boosted dark Higgs candidates that lead to the appearance of merged jets, and the usual handles on the jet substructure are used. To efficiently use all of those handles, the DeepAK15 jet tagger is chosen as a new-generation hadronic tagger, capable of tagging heavy objects (W, Z, H, t) and also tagging the flavour of the hadronic decay (light vs. charm vs. bottom). The DeepAK15 tagger is based on Machine Learning techniques, using as inputs both the list of PF candidates that comprise the jet, and the list of secondary vertices contained in it.

One of the main strengths of the DeepAK15 tagger is that it can be built to be almost completely uncorrelated with the (SoftDrop) jet mass $m_{\mathrm{SD}}$ itself (Dolen et al., 2016). This allows independent selections on both variables, defining multiple signal regions, and improves the efficiency of the tagger. In this analysis, the mass decorrelated version of the tagger is used exclusively to further improve its performance. Overall, the DeepAK15 tagger is an effective tool for reducing background and improving the accuracy of jet reconstruction in the CMS experiment.

### 3.6.1. AK15 Jet Masses

We choose to use the SoftDrop (A. J. Larkoski et al., 2014) jet grooming algorithm to calculate the AK15 jet mass.

To calculate the softdrop jet, first the jet components are reclustered with the Cambridge-Aachen algorithm (Collaboration, 2009). Since it is a sequential recombination algorithm, the whole clustering history is saved. Then, starting from the last recombination:

1. The recombination is undone, and the fraction of shared momentum $z$ between the two subjets
$$z = \frac{\min\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)}{p_{\mathrm{T},1} + p_{\mathrm{T},2}}$$
and the angle $\theta$ between them is calculated.

2. The softdrop condition $z > z_{cut} \cdot \theta^{\beta}$ is calculated. If it is true, the process stops and the jet is considered the final softdrop jet.

3. Otherwise, the softest subjet is processed, while the process starts anew from the first recommendation for the most complicated subjet.

The softdrop jet is further processed by the following algorithm:

- The inverse of the JEC correction is applied to the subjets, and their invariant mass is calculated to form a "raw softdrop mass" $m_{\mathrm{SD}}^{\mathrm{raw}}$.
- A PUPPI softdrop mass correction weight $w_{psmc}$ is calculated from the JEC-corrected groomed jet $p_{\mathrm{T}}$ and $\eta$The formula is reproduced here for

convenience:

$$g = 1.00626 - 1.06161 \cdot (0.07999 \cdot p_{\mathrm{T}})^{-1.20454}$$

$$r_{cen} = 1.09302 - 1.50068 \cdot 10^{-4} p_{\mathrm{T}} + 3.44866 \cdot 10^{-7} p_{\mathrm{T}}^2 - 2.68100 \cdot 10^{-10} p_{\mathrm{T}}^3 + 8.67440 \cdot 10^{-14} p_{\mathrm{T}}^4 - 1.00114 \cdot 10^{-17} p_{\mathrm{T}}^5$$

$$r_{fwd} = 1.27212 - 5.71640 \cdot 10^{-4} p_{\mathrm{T}} + 8.37289 \cdot 10^{-7} p_{\mathrm{T}}^2 - 5.20433 \cdot 10^{-10} p_{\mathrm{T}}^3 + 1.45375 \cdot 10^{-13} p_{\mathrm{T}}^4 - 1.50389 \cdot 10^{-17} p_{\mathrm{T}}^5$$

$$w_{psmc} = g \times r_{cen} \text{ if } |\eta| < 1.3, g \times r_{fwd} \text{ otherwise}$$

and the corrections are plotted in Fig. 3.9. This formula relates several variables used in the particle physics analysis of the CMS detector. Here is each term:

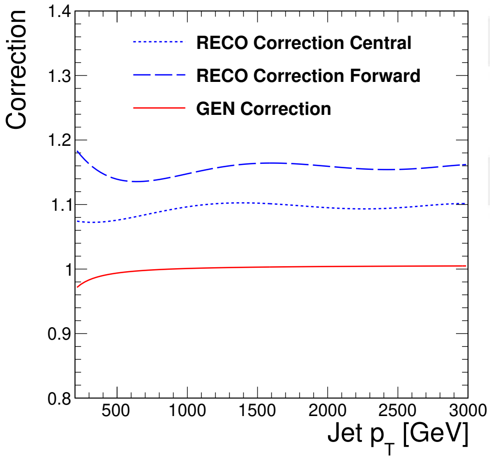



Figure 3.9. PUPPI softdrop mass correction weights, as function of the fat jet $p_{\mathrm{T}}$. The red solid line is the GEN correction, whilst the blue dotted (dashed) lines are the central (forward) RECO corrections.

- $g$ and $r_{cen}$ are parameters used to calculate the energy resolution of the ECAL (electromagnetic calorimeter) in the central region of the detector ($|\eta| < 1.3$).
- $r_{fwd}$ is a similar parameter used to calculate the energy resolution in the forward region of the detector ($|\eta| \geq 1.3$).

- $w_{psmc}$ is the width of the energy distribution (in units of GeV) for a photon or electron measured by the ECAL.

The parameters $g$ and $r_{cen}$ are calculated as a function of the transverse momentum of the photon or electron ($p_T$). The formulas for $g$ and $r_{cen}$ are both functions of $p_T$ and involve several coefficients. The values of these coefficients were determined from simulation studies and calibration measurements.

Finally, $w_{psmc}$ is calculated by multiplying $g$ and $r_{cen}$ if the measured particle is in the central region of the detector, or $g$ and $r_{fwd}$ if it is in the forward region. This formula is used to correct for the energy resolution of the ECAL when measuring the energy of particles like photons and electrons in the detector. Ref provides further details on this formula.

The final quantity that we consider for the AK15 jet mass is the $m_{\mathrm{SD}}$ variable defined as $m_{\mathrm{SD}} \equiv m_{\mathrm{SD}}^{\mathrm{raw}} \times w_{psmc}$. We consider AK15 jets with $m_{\mathrm{SD}}$ in the [0, 300] GeV range. For distribution plots, we consider 10 GeV uniform binning, whilst for background modeling we consider a binning optimized against signal significance: [40, 50, 60, 70, 80, 90, 100, 120, 150, 180, 240, 300] GeV

Figure 3.10. shows, at the preselection level, the jet mass variable $m_{\mathrm{SD}}$ for the main backgrounds, the dark Higgs benchmarks and monojet, mono-W, and mono-Z signals; it also shows the good agreement between observed data and background simulation in the single muon control region for all three years.

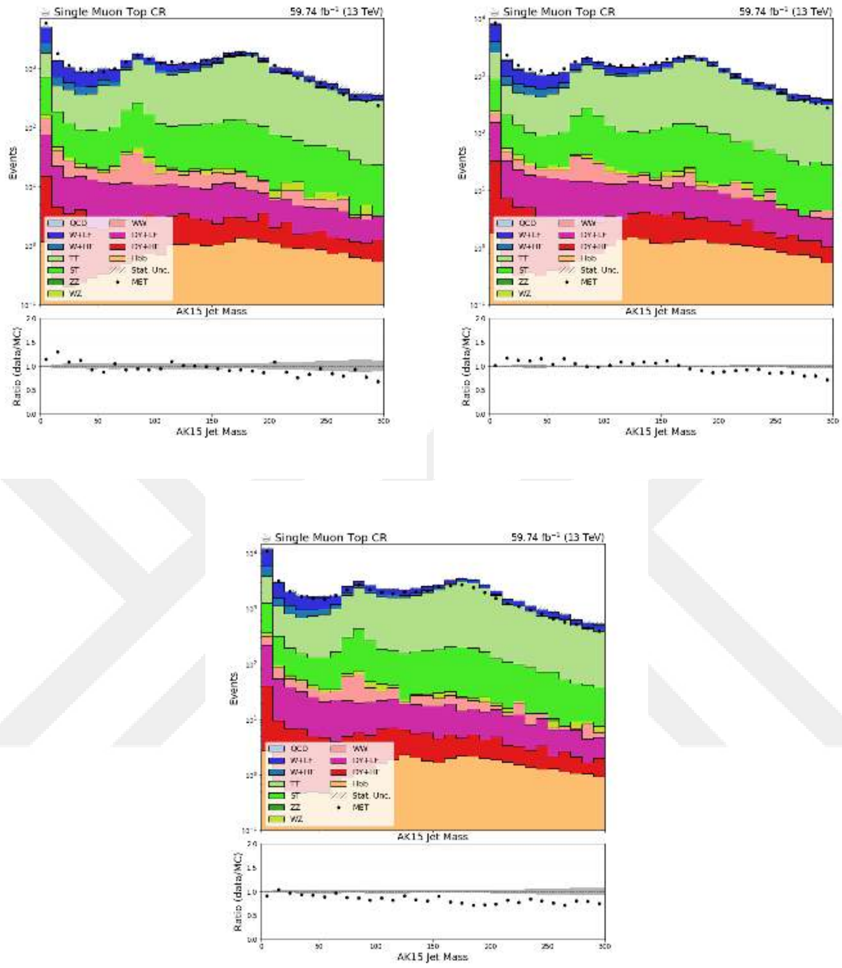



Figure 3.10. The jet mass variable $m_{\mathrm{SD}}$ presents good discrimination power between the dark Higgs signal benchmarks and the standard model backgrounds for 2016 (top left), 2017 (top right), and 2018 (bottom) data-taking periods. It is also well-modeled in the control regions studied (shown for the single muon control region) for all three years.

### 3.6.2. AK15 b-tagging

We use the DeepAK15 tagger output to define signal categories and further improve the significance, $S/\sqrt{B}$. The standard `ZHbbvsQCD` tagger implemented in CMSSW and included in NanoAOD is used. The tagger score is defined as follows:

$$\texttt{ZHbbvsQCD} = \frac{P(Zbb) + P(Hbb)}{P(Zbb) + P(Hbb) + P(QCD)}, \tag{3.6}$$

where $P(Zbb)$, $P(Hbb)$, and $P(QCD)$ are the probabilities for the AK15 jet to be initiated by a Z boson, a Higgs boson, and a quark or gluon, respectively. Although it is typically defined for AK8 jets, the same aggregation of individual probabilities can be applied to AK15 jets.

The main advantage of using a mass-decorrelated tagger is the tagging efficiencies are approximately constant across mass bins. This reduces systematic effects and allows us to use a single threshold on the `ZHbbvsQCD` tagger output as the selection. We define the *mono-hs* category as `ZHbbvsQCD` $> 0.65$. Figure 3.11. shows, at the preselection level, the tagger output for the main backgrounds, the dark Higgs benchmarks, and monojet, mono-W, and mono-Z signals. Figure 3.12. shows the good agreement between observed data and simulation in the single muon control region for all three years.

For the purposes of demonstrating the additional separation power of the `ZHbbvsQCD` tagger, we define three auxiliary mass windows: 30–60 GeV, 60–80 GeV, 80–120 GeV, centered around the three dark Higgs signal benchmarks. Table 3.12. shows the simulated background and signal yields, for each dark Higgs mass hypothesis, in each mass window, before and after the `ZHbbvsQCD` selection. It can be seen that sensitivity for mono-hs signals improves by a factor of 3–4. One may also notice that the $t\bar{t}$ background contribution grows with mass, being responsible for approximately 40% of the expected background in the 80–120 GeV mass bin. This

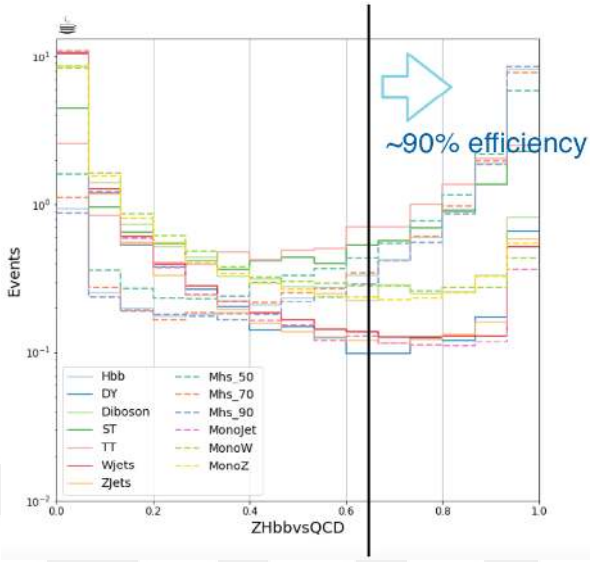



Figure 3.11. The `ZHbbvsQCD` tagger score allows for good separation between resonant $X \to \mathrm{b\overline{b}}$ production (exemplified by the three dark Higgs benchmarks and the $\mathrm{H} \to \mathrm{b\overline{b}}$ itself) and all other processes. Notice that the monojet, mono-W, and mono-Z signal processes are also selected against by the tagger. We define the *mono-hs* category as `ZHbbvsQCD` $> 0.65$; it is more than 90% efficient for the benchmark dark Higgs processes.

effect is due to the presence of something that we did not understand yet. However, we note that reducing the $t\bar{t}$ to a fifth of its current value whilst keeping 90% of signal efficiency would lead to only a 10% increase in sensitivity. Therefore, the usage of a dedicated tagger to exclude $t\bar{t}$ events is unwarranted.

Table 3.12. The ZHbbvsQCD tagger has approximately constant efficiency with mass, increasing signal sensitivity by a factor of 4 for tested dark Higgs benchmarks. However, the $t\bar{t}$ background increases with mass and needs to be addressed. Yields are for 59.97 $\mathrm{fb}^{-1}$.

| Before ZHbbvsQCD selection | | | |
|---|---|---|---|
| Process | Mass window [GeV ] | | |
| | 30--60 | 60--80 | 80--120 |
| W+HF | 1963.8 | 1008.4 | 1894.4 |
| W+LF | 14588.7 | 7389.7 | 12636.1 |
| Z+HF | 2541.6 | 1377.2 | 2505.8 |
| Z+LF | 27974.2 | 14144.4 | 24044.0 |
| TT | 307.6 | 339.0 | 1663.6 |
| ST | 122.0 | 92.8 | 406.1 |
| DY+HF | 12.4 | 6.0 | 14.1 |
| DY+LF | 187.6 | 96.8 | 171.2 |
| WW | 158.9 | 149.5 | 926.6 |
| WZ | 236.7 | 233.4 | 1279.4 |
| ZZ | 112.0 | 85.7 | 755.7 |
| Hbb | 4.5 | 3.1 | 17.8 |
| QCD | 583.0 | 328.1 | 398.5 |
| Total expected background | 48793.0 | 25253.9 | 46713.3 |
| $h_s$ 50 GeV | 2469.3 | 489.7 | 282.0 |
| $h_s$ 50 GeV, $S/\sqrt{B}$ | **11.2** | 3.1 | 1.3 |
| $h_s$ 70 GeV | 409.7 | 1570.1 | 941.6 |
| $h_s$ 70 GeV, $S/\sqrt{B}$ | 1.9 | **9.9** | 4.4 |
| $h_s$ 90 GeV | 173.8 | 317.6 | 2041.2 |
| $h_s$ 90 GeV, $S/\sqrt{B}$ | 0.8 | 2.0 | **9.4** |
| After ZHbbvsQCD selection | | | |
| Total expected background | 2767.0 | 1458.5 | 3268.8 |
| $h_s$ 50 GeV | 2155.7 | 384.0 | 146.7 |
| $h_s$ 50 GeV, $S/\sqrt{B}$ | **41.0** | 10.1 | 2.6 |
| $h_s$ 70 GeV | 311.8 | 1422.9 | 798.6 |
| $h_s$ 70 GeV, $S/\sqrt{B}$ | 5.9 | **37.3** | 14.0 |
| $h_s$ 90 GeV | 114.8 | 269.9 | 1865.0 |
| $h_s$ 90 GeV, $S/\sqrt{B}$ | 2.2 | 7.1 | **32.6** |

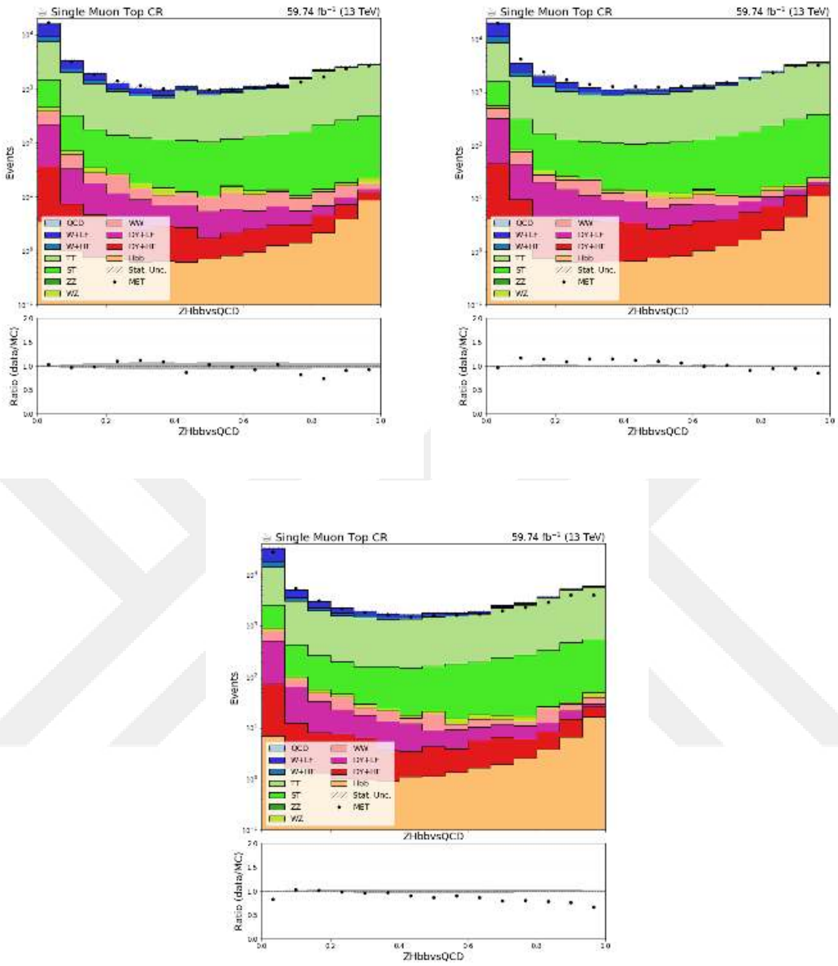



Figure 3.12. Simulation to data comparison for the `ZHbbvsQCD` tagger score, in the single muon control region, for 2016 (top left), 2017 (top right), and 2018 (bottom) data-taking periods.

## 3.7. Event Selection

The event selection process begins with a preselection that is common to both signal and control regions, as described in Table 3.11. This preselection ensures a topology that closely resembles the expected signature of the dark Higgs signal. Additional criteria, such as lepton and photon multiplicities and the number of b-tagged AK4 jets that do not overlap with the fat jet, are then used to differentiate between the signal and control regions, while minimizing uncertainties due to extrapolation.

### 3.7.1. Control and Signal Regions

This search's three main SM backgrounds are W, Z + jets, and $t\bar{t}$ production. To facilitate the signal extraction process defined in Sec. 3.8 3.8., we choose the smallest possible number of control regions to constrain the background shapes and yields. There are two main criteria to define the regions: the leading candidate jet in the event must pass the DeepAK15 selection, and the number of well-reconstructed leptons $\ell$ (e or μ) in the event. Signal events are mostly expected to populate the pass, $0\ell$ category, while background events are expected to mostly populate the other categories (pass, $1\ell$, fail, $0\ell$, fail, $1\ell$). We use the distribution of events in these categories to estimate the remaining SM background in the signal-populated category.

The regions are defined as follows:

Zero lepton: requires jet HEM mitigation, no loose leptons (muons or electrons), no loose photons, selected by $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers. For the 2018 period, we also require the additional $p_{\mathrm{T}}^{\mathrm{miss}}$ HEM mitigation described above. Single muon: requires presence of one tight muon, no loose electrons, no loose photons, no extra b-tagged jets, jet HEM mitigation, selected by $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers. Single electron: requires presence of one tight electron, no loose muons, no loose photons, no extra b-tagged jets, jet

HEM mitigation, selected by single electron triggers. Also requires $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$ and $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}) > 0.7$ to improve $p_{\mathrm{T}}^{\mathrm{miss}}$ modelling. Top muon: requires presence of one tight muon, no loose electrons, no loose photons, at least one extra b-tagged jet, jet HEM mitigation, selected by $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers. Top electron: requires presence of one tight electron, no loose muons, no loose photons, at least one extra b-tagged jet, jet HEM mitigation, selected by single electron triggers. Also requires $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$ and $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}) > 0.7$ to improve $p_{\mathrm{T}}^{\mathrm{miss}}$ modelling.

All the regions are further split into pass/fail regions depending on whether or not they pass the threshold on the fat-jet b-tag of `ZHbbvsQCD` $> 0.65$. The pass subset of the zero lepton region is the *signal region* where we search for evidence of Dark Higgs events. Our challenge is to predict the distributions of the SM backgrounds ($\mathrm{W}, \mathrm{Z}$ + jets, and $t\bar{t}$) in the signal region. The following section shows how to use the control regions to achieve this goal. Figure 3.13. gives an overview of the regions, and Table 3.13. summarizes the different selection criteria.

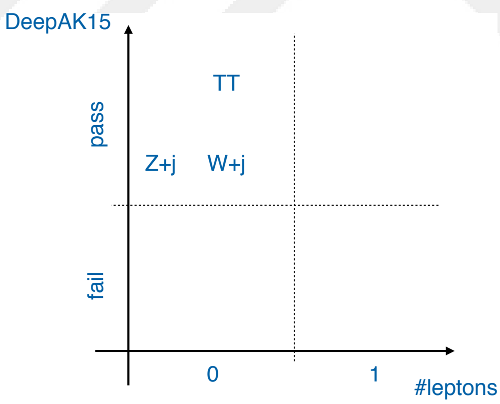



Figure 3.13. Overall picture of the regions.

Table 3.13. Event selection table. Common to all regions is a preselection, including any year-dependent requirements. The zero lepton pass region is the signal region, where we expect most Dark Higgs events to appear. The zero lepton fail region is a control region mostly dominated by W/Z + jets events. The single lepton control regions are used to constrain the W + jets background; a special $t\bar{t}$ control region is defined, with the requirement of additional b-tagged jets, to constrain that background.

| Region | Leptons | Photons | Fat-jet b-tag | Extra b-tags | Additional |
|---|---|---|---|---|---|
| Signal region | 0 | 0 | Pass | 0 | N/A |
| Zero lepton | 0 | 0 | Fail | 0 | N/A |
| Single lepton | 1 | 0 | Pass/fail | 0 | $p_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV for electron. |
| $t\bar{t}$ | 1 | 0 | Pass/fail | $\geq 1$ | $p_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV for electron. |
| Photon | 0 | 1 | Pass/fail | 0 | N/A |

### 3.8. Background Estimation and Signal Extraction

As previously discussed, the three main Standard Model (SM) backgrounds to this search are W, Z + jets, and $t\bar{t}$ production. To predict the W, Z + jets $m_{SD}$ versus recoil shape in the signal region (i.e., the pass subset of the zero-lepton region), we use the data in the fail subset of the zero-lepton region, and calculate pass-fail transfer factors. In other words, we determine the ratio of the W, Z + jets $m_{SD}$ versus recoil shapes in the pass and fail subsets, which allows us to estimate the background shape in the signal region. This approach has no signal contamination and is completely dominated by W, Z + jets.

To further constrain the transfer factors, we also use the single-lepton regions, which are enriched in W + jets events, as well as the top lepton regions, which are enriched in $t\bar{t}$ events. The implicit W-Z link from electroweak symmetry is used to constrain the Z + jets background in the signal region.

#### 3.8.1. Fit Description

To extract the results, a binned fit is performed using the Higgs Combine tool (Higgs PAG, 2019) on the hadronic recoil spectrum in bins of $m_{SD}$. The bin edges for the hadronic recoil spectrum are defined as [250, 310, 370, 470, 590, 3000] GeV, while those for the $m_{SD}$ spectrum are [40, 50, 60, 70, 80, 90, 100, 120, 150, 180, 240, 300] GeV. It should be noted that the fit is made under the assumption of factorization of all relevant event distributions between the hadronic recoil and the mass variable, and that there is a well-behaved function between the Z+jets and W+jets processes, thanks to electroweak symmetry. This is the same assumption made in the standard CMS monojet analysis. Additionally, the data in the zero-lepton fail region is used to derive a pass-to-fail ratio in bins of $m_{SD}$ for Z+jets in the signal region, which is then propagated to W+jets through the aforementioned W-Z link.

### 3.8.2. Z jets Background Estimation

We will employ the widely-used *rhalphabet* method (TRootABCD TWIKI, 2023) for estimating the Z + jets background component in the signal region. This method is an extension of the standard alphabet (a.k.a "ABCD") method for distributions and is based on the assumption that there exists a reasonable relationship, known as the *transfer factor*, between the distributions of a given background component in two related regions. The transfer factor connects the Z + jets background component in the zero-lepton region to that in the signal region, enabling its estimation in the latter from the observed data in the former.

It should be noted that the factorisation between the two variables ($m_{\mathrm{SD}}$ and $U$) is not perfect, and this imperfection is taken into account in our analysis. Specifically, we assume that the pass-fail ratio is sensitive to the differences in the distributions of $m_{\mathrm{SD}}$ and $U$ between the two regions. The pass-fail ratio is defined as the ratio of the expected number of events in the signal region to that in the zero-lepton fail region, both in the same $m_{\mathrm{SD}}$ bin, and is used to estimate the Z + jets background in the signal region. The difference between the signal region and the zero-lepton region lies in the category of the fat jet tagging, which we denote as pass or fail, and this difference is accounted for in the pass-fail ratio.

The pass-fail ratio is a crucial component in estimating the Z + jets background in the signal region, and it must account for two effects. First, it must cover any residual tagger-kinematics correlation, including dependence on the varying heavy-flavor fraction. Second, it must cover the effects of data-simulation discrepancies in the zero-lepton region. To handle these effects, we choose to factorize the pass-fail ratio into two parts: a simulation-based pass-fail ratio that incorporates systematic effects, and a polynomial with free parameters.

The pass-fail ratio is a ratio between two 2D binned distributions, in $m_{\mathrm{SD}}$ and recoil $U$, and is therefore also a 2D binned function divided into the 5 recoil $\times$ 11 $m_{\mathrm{SD}}$ bins discussed above. The polynomial correction is assumed to be defined over the 2D bin barycentres and estimated on the Bernstein basis. Thus, the final form of the pass-fail ratio can be written as:

$$R^{\mathrm{Z+jets}}(m_{\mathrm{SD}},U) = \sum_{l=0}^{n_U} a_{k,l} b_{k,n_{m_{\mathrm{SD}}}}(m_{\mathrm{SD}}) b_{l,n_U}(U) \times \varepsilon^{\mathrm{Z+jets}}(m_{\mathrm{SD}},U) \qquad (3.7)$$

where $k$ and $l$ are Bernstein polynomial indices, and $\varepsilon^{\mathrm{Z+jets}}(m_{\mathrm{SD}},U)$ is the simulation-based pass-fail ratio that incorporates systematic effects. The Z + jets distribution in the signal region is given by:

$$N^{\mathrm{Z+jets}} 0\ell, \mathrm{pass}(m_{\mathrm{SD}},U) = N^{\mathrm{Z+jets}} 0\ell, \mathrm{fail}(m_{\mathrm{SD}},U) \times R^{\mathrm{Z+jets}}(m_{\mathrm{SD}},U) \qquad (3.8)$$

where the subscripts $0\ell$, pass/fail are used to explicitly indicate the relationship between the signal region and zero-lepton control region.

### 3.8.3. W + jets Background Estimation

We use the same approach as for Z + jets, including the same TF (transfer factor) factorization and the usage of a heavy flavour fraction systematic uncertainty. As such, the W + jets distribution in the signal region is expressed as

$$N^{\mathrm{W+jets}}_{0\ell,\mathrm{pass}}(m_{\mathrm{SD}},U) = N^{\mathrm{W+jets}}_{0\ell,\mathrm{fail}}(m_{\mathrm{SD}},U) \times R^{\mathrm{W+jets}}(m_{\mathrm{SD}},U). \qquad (3.9)$$

There are some additional constraints in this background estimation. First, we assume that Z + jets and W + jets are related together through a W–Z link, as done in the monojet analysis:

$$N^{\mathrm{W+jets}}_{0\ell,\mathrm{fail}}(m_{\mathrm{SD}},U) = \frac{N^{\mathrm{Z+jets}}_{0\ell,\mathrm{fail}}(m_{\mathrm{SD}},U)}{E^{\mathrm{W-Z}}(m_{\mathrm{SD}},U)} \qquad (3.10)$$

Second, we use the single-lepton ($1\ell$) regions to constrain the W + jets distributions in both the pass and fail categories:

$$N^{\mathrm{W+jets}}_{1\ell,\mathrm{pass/fail}}(m_{\mathrm{SD}},U) = \frac{N^{\mathrm{W+jets}}_{0\ell,\mathrm{pass/fail}}(m_{\mathrm{SD}},U)}{T^{\mathrm{W+jets}}_{\mathrm{pass/fail}}(m_{\mathrm{SD}},U)} \tag{3.11}$$

So, to summarise the three kinds of transfer factors:

- A transfer factor $R$ that connects the irreducible Z + jets background in between the zero-lepton pass signal region and the zero-lepton fail control region (CR). This is a pass-fail ratio related to the inefficiency of the DeepAK15 tagging. There is an equivalent factor for the W + jets background where the lepton has been lost.
- A transfer factor $E$ (electroweak) that connects the W and Z + jets backgrounds in the fail region. Considering the lepton from W + jets is already lost, the ratio of contributions of each process in this region is dominated by electroweak symmetry considerations. The effect of W/Z NLO correction uncertainties appears mostly through this link.
- A transfer factor $T$ constrains the W + jets background in between the zero-lepton and one-lepton regions, separately in the pass and fail categories. This is directly related to the lepton misreconstruction and misidentification probabilities.

Figure 3.14. shows the approach to the transfer factors amongst the regions.

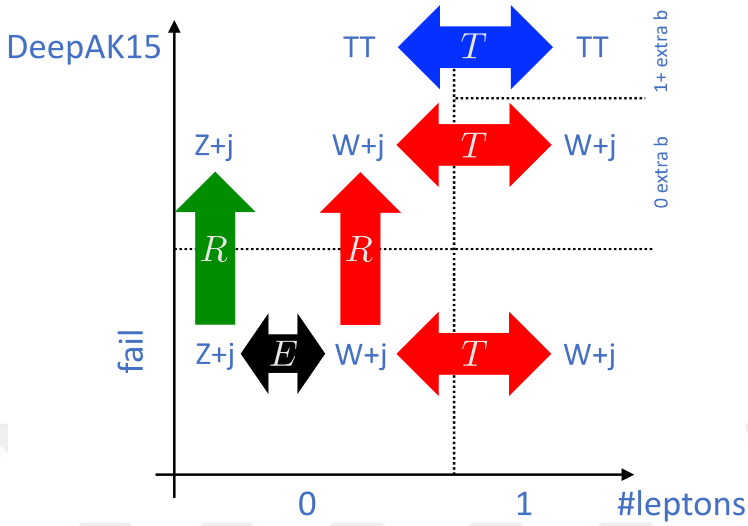



Figure 3.14. Transfer factors amongst the regions. The $R$ transfer factors represent pass-fail ratios related to the inefficiency of the DeepAK15 tagging. The $E$ (electroweak) transfer factor connects the W and Z + jets backgrounds in the fail region. The $T$ transfer factor relates the W + jets background in the zero-lepton and one-lepton regions; a similar transfer factor does the same for $t\bar{t}$, using the one-lepton + one b-tagged jet region.

### 3.9. Scale Factors and Systematic Uncertainties

#### 3.9.1. Heavy-flavor fraction uncertainty

When using our CRs to constrain the signal region, we rely on knowing the heavy-flavor fraction of the backgrounds in order to be able to accurately build TFs between regions with different $b$-tagging requirements. Unfortunately, the heavy-flavor fraction is not well-modeled from theory. Therefore, an estimation on the uncertainty is made using experimental results. CMS has measured the differential cross-section of $W$+jets as a function of $N_{\text{jets}}$ CMS Collaboration, 2012a. From this study, we make the conservative estimate that the uncertainty on the cross-section of $W$+fat jet is less than 20% (Figure 3.15.). Next, using CMS measurements of $W+c$

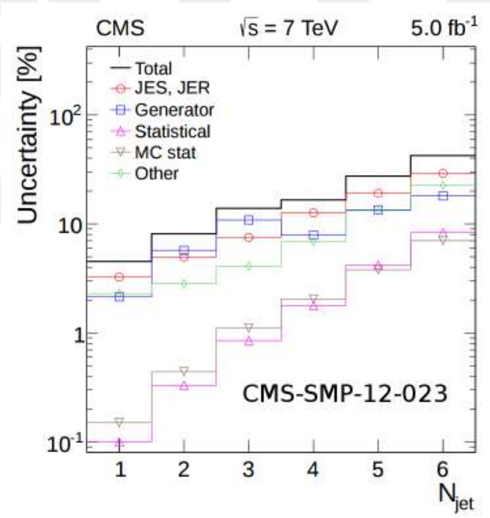



Figure 3.15. Figure showing uncertainty in $W$+jets production as a function of $N_{\text{jets}}$. To estimate an uncertainty on the production of $W$+fat jet, we look at the $N_{\text{jets}} = 2,3$ bins (CMS Collaboration, 2012a).

(CMS Collaboration, 2013) and $W+b\bar{b}$ (CMS Collaboration, 2012b) production, we extract the following uncertainties on heavy-flavor production:

$$\frac{\delta\sigma(W+c, W \to \ell\nu)}{\sigma(W+c, W \to \ell\nu)} = \frac{7.65 \text{ pb}}{107.7 \text{ pb}}, \quad \frac{\delta\sigma(W+b\bar{b}, W \to \ell\nu)}{\sigma(W+b\bar{b}, W \to \ell\nu)} = \frac{0.206 \text{ pb}}{1.06 \text{ pb}} \tag{3.12}$$

The uncertainties are then propagated as follows:

$$\frac{\delta\sigma(W+\text{HF})}{\sigma(W+\text{fat jet})} = \left[\left(\frac{\delta\sigma(W+\text{HF})}{\sigma(W+\text{HF})}\right)^2 + \left(\frac{\delta\sigma(W+\text{fat jet})}{\sigma(W+\text{fat jet})}\right)^2\right]^{1/2}$$

$$= \left[\frac{\delta\sigma(W+c)^2 + \delta\sigma(W+b\bar{b})^2}{\left(\sigma(W+c)+\sigma(W+b\bar{b})\right)^2} + \left(\frac{\delta\sigma(W+\text{fat jet})}{\sigma(W+\text{fat jet})}\right)^2\right]^{1/2}$$

Using the values discussed above, we arrive at:

$$\frac{\delta\sigma(W+\text{HF}, W\to \ell\nu)}{\sigma(W+\text{fat jet}, W\to \ell\nu)} \approx 0.21 \tag{3.13}$$

A similar calculation is done for $Z$+jets. First, we estimate the uncertainty $Z$+fat jet production (CMS Collaboration, 2014b) as 15% using Figure 3.16. Then, from (CMS

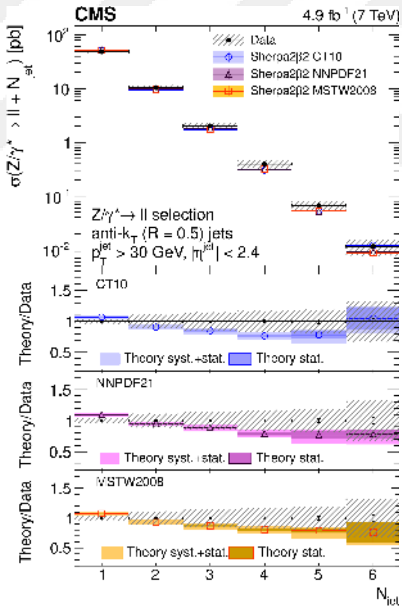



Figure 3.16. Figure showing uncertainty in $Z$+jets production as a function of $N_{\text{jets}}$. To estimate an uncertainty on the production of $Z$+fat jet, we look at the $N_{\text{jets}} = 2, 3$ bins (CMS Collaboration, 2014b).

Collaboration, 2014a), we get the the uncertainty on $Z+b(b)$ production as:

$$\frac{\delta\sigma(Z+b(b))}{\sigma(Z+b(b))} = \frac{0.909 \text{ pb}}{5.84 \text{ pb}} \tag{3.14}$$

The resulting uncertainty on the heavy-flavor fraction is:

$$\frac{\delta\sigma(Z+\text{HF}, Z\to\ell\ell)}{\sigma(Z+\text{fat jet}, Z\to\ell\ell)} = \left[\left(\frac{\delta\sigma(Z+\text{HF})}{\sigma(Z+\text{HF})}\right)^2 + \left(\frac{\delta\sigma(Z+\text{fat jet})}{\sigma(Z+\text{fat jet})}\right)^2\right]^{1/2} = 0.22 \tag{3.15}$$

All HF-fraction uncertainties are then propagated through as rate uncertainties on the respective processes. That is, two nuisances are added to the fit: $Z$+HF and $W$+HF. These nuisances are correlated across the relevant processes in all regions. Depending on the process and region, the rate uncertainties range from 4% to 5%.

### 3.9.2. *b*-tagging scale factors

The BTV POG provides scale factors and associated uncertainties for *b*-tagging of narrow jets. These object scale factors are propagated into event scale factors for this analysis. The associated uncertainties are propagated as two sets of shape variations: narrow jet *b*-tag and narrow jet mis-tag. The recommendations for the scale factors and uncertainties are from Ref (CMS Jet/BTV POG, n.d.).

The scale factors and uncertainties provided by the POG largely cover the $p_T$ range probed in this analysis. AK4 jet SFs are provided for $p_T < 1000$ GeV. For jets beyond this $p_T$ range, the recommendation is to use the scale factor from the $p_T$ boundary and double the uncertainty (Ref ibid.). As shown in Figure 3.17., very few jets in this analysis extend past these $p_T$ bounds.

As recommended by the BTV POG, we derive the MC efficiency for *b*-tagging jets in the phase space used by this analysis. The efficiencies are derived using the $t\bar{t}$ Monte Carlo sample and applying a preselection of $U > 200$ GeV. The jets are split

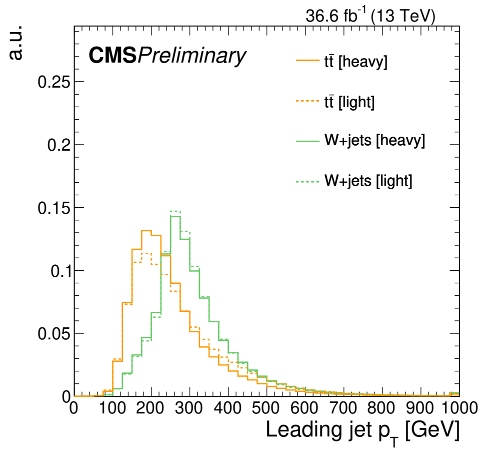



Figure 3.17. $p_{\mathrm{T}}$ spectra of the leading AK4 jet in $t\bar{t}$ and $W$+jets MC events, after a basic selection ($U > 200$ GeV, fat jet $p_{\mathrm{T}} > 200$ GeV) is applied. The shapes are split based on whether the jet is matched to a truth-level $b$ or $c$ (heavy) or not (light). In all cases, the bulk of the spectra are well within the range supported by the POG scale factors and uncertainties.

into three categories: $b$ jets (matched to a truth $b$), $c$ jets (matched to a truth $c$), or light-flavor jets (unmatched). Figure 3.18. shows the efficiencies as a function of the reconstructed jet $p_T$ and η.

### 3.9.3. Lepton Efficiency Scale Factors

We apply data-to-MC scale factors to events in the CRs used in the analysis. These data-to-MC scale factors are derived from the efficiencies that are measured for the electron and muon selections in both data and simulation. The uncertainty on the lepton efficiency scale factor is correlated between the single muon/electron and dimuon/dielectron control regions. There are uncertainties for:

- Electron ID and tracking (1% and 0.5% per lepton leg, respectively)
- Muon ID and tracking (1% and 0.5% per lepton leg, respectively)
- Tau veto (3%)

A more thorough description of the lepton uncertainties we use are described

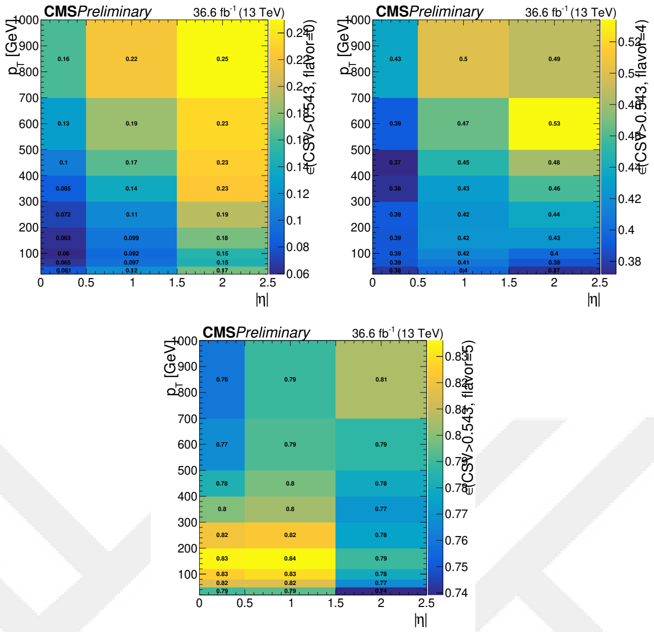



Figure 3.18. $b$-tagging efficiency in the analysis phase space ($U > 200$ GeV). The efficiency is shown in three cases for flavor =0 (top left), flavor =4 (top right), and flavor =5 (bottom) categories depending on the truth flavor of the jet.

in Akchurin et al., 2016.

### 3.9.4. Trigger Efficiencies

Single-electron triggers are used to select events with electrons. The efficiencies of these triggers in data are parametrized as a function of the electron $p_{\mathrm{T}}$ and applied to MC. The derivation of these efficiencies is described in (ibid.).

$p_{\mathrm{T}}^{\mathrm{miss}}$ triggers are used to select events with real $p_{\mathrm{T}}^{\mathrm{miss}}$ or muons. The efficiency is parametrized as a function of the hadronic recoil $U$ in such events. More specifically, it is a function of $p_{\mathrm{T}}^{\mathrm{miss}}$ computed without muons. However, in the monojet analysis (ibid.), it was found that the efficiency has a small dependence on the number of muons in the event. The effect comes from muons faking jets in the HLT, lowering the efficiency of the trigger. At $U = 200$ GeV, the effect is small ($\sim 2.5\%$), and decreases at larger $U$. Since we are operating in the realm of $U > 250$ GeV, this uncertainty is not applicable. The prescription used is the same as in (Akchurin et al., 2016).

### 3.9.5. QCD Normalization Uncertainty

In this analysis, the QCD background is estimated directly from Monte Carlo simulation in all regions. A 100% uncertainty is applied to the normalization of the QCD. This uncertainty is correlated between regions with the same source of the fake. That is, one uncertainty is applied to QCD in the signal region, a separate uncertainty is applied to QCD in muon control regions, and similarly for electrons. The 100% uncertainty is estimated by looking in a QCD-enriched region:

- Veto electrons, muons, taus, photons
- Require $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
- Require $\min\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}},\mathrm{jet}) < 0.1$

The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of this region is shown in Figure 3.19. The disagreement between data and simulation is at most 50%. We double this difference between data and Monte Carlo and use it as the uncertainty on the normalization of QCD.

### 3.9.6. DeepAK15 Scale Factor Uncertainties

For W/Z+jets production, scale factor are measured simultaneously with the signal extraction directly in the final fit. This in-situ calibration is made possible by using events that fail the double-b requirement. The mistag scale factor is taken as a correction to the double-b efficiency for W/Z+jets and included in the fit as an unconstrained nuisance. The double-b efficiency estimated on MC is:

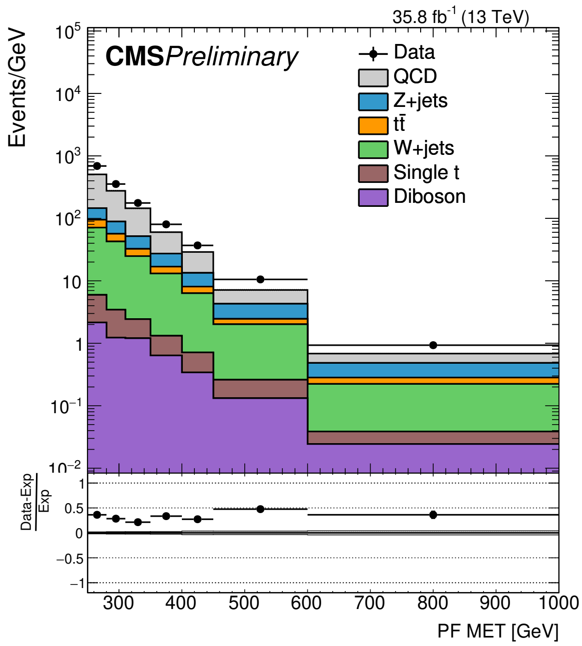



Figure 3.19. Recoil distribution in a QCD-enriched region. Used to validate the 100% normalization uncertainty applied to QCD. The disagreement is at most 50%.

$$\varepsilon_{\mathrm{MC}}^{\mathrm{W/Z+jets\ mis-tag}} \& = \frac{N_{W/Z+jets}^{\mathrm{pass}}}{N_{W/Z+jets}} \tag{3.16}$$

where $N_{W/Z+jets}^{\mathrm{pass}}$ is the yield of W/Z+jets production in the pass signal/control region (double-b tag $> 0.75$). To take into account the variation of such efficiency introduced by the uncertainty on the heavy flavor fraction of W/Z+jets events, two efficiencies are estimated performing the same calculation described above, but with different $N_{W/Z+jets}^{\mathrm{pass}}$ obtained by pushing up and down the heavy flavor fraction. The difference between the up and down efficiency with respect to the central value is taken as a systematic uncertainty. The mistag scale factor for W/Z+jets production is obtained by imposing

$$N_{W/Z+jets}^{\mathrm{total}} \& = N_{W/Z+jets}^{\mathrm{pass}} + N_{W/Z+jets}^{\mathrm{fail}} \tag{3.17}$$

considering

$$\begin{aligned} N^{\text{pass}}_{W/Z+jets} &= \varepsilon^{\text{W/Z+jets mis-tag}}_{\text{data}} \times N^{\text{total}}_{W/Z+jets} \\ &= SF^{\text{W/Z+jets mis-tag}}_{pass} \times \varepsilon^{\text{W/Z+jets mis-tag}}_{\text{MC}} \times N^{\text{total}}_{W/Z+jets} \end{aligned} \tag{3.18}$$

and deriving $N^{\text{fail}}_{W/Z+jets}$ as

$$N^{\text{fail}}_{W/Z+jets} = N^{\text{total}}_{W/Z+jets} \times (1 - SF^{\text{W/Z+jets mis-tag}}_{pass} \times \varepsilon^{\text{W/Z+jets mis-tag}}_{\text{MC}}) \tag{3.19}$$

so that $SF^{\text{W/Z+jets mis-tag}}_{fail}$ can be expressed in terms of $SF^{\text{W/Z+jets mis-tag}}_{pass}$ as

$$SF^{\text{W/Z+jets mis-tag}}_{fail} = \frac{1 - SF^{\text{W/Z+jets mis-tag}}_{pass} \times \varepsilon^{\text{W/Z+jets mis-tag}}_{\text{MC}}}{1 - \varepsilon^{\text{W/Z+jets mis-tag}}_{\text{MC}}} \tag{3.20}$$

making $SF^{\text{W/Z+jets mis-tag}}_{pass}$ the only parameter to be extracted in the fit together with the overall W/Z+jets normalizations. By splitting each control region in pass and fail we have two independent control samples used to constrain two unknowns. The $SF^{\text{W/Z+jets mis-tag}}_{pass}$ are simultaneously applied to the W/Z+jets processes in signal region.

A similar approach is used for the $t\bar{t}$ mistag SF, where the independent measurement of the mistag efficiency is used as prior for the central value of the $SF^{t\bar{t}\text{ mis-tag}}_{pass}$ nuisance and the measured uncertainty as its prior uncertainty in the fit. In this case, the efficiency is fixed since no fluctuation is introduced by any effect such as the varying heavy flavor fraction for W/Z+jets. The exact values for the efficiencies going into the fit and the values of the scale factors extracted from it are as follows:

- **Prefit:** $\varepsilon^{\text{W+jets mis-tag}}_{\text{MC}} = 0.0316 \pm 0.00126$, $\varepsilon^{\text{Z+jets mis-tag}}_{\text{MC}} = 0.0346 \pm 0.00174$, $\varepsilon^{t\bar{t}}_{\text{MC}} = 0.1678 \pm 0.0$
- **Postfit:** $\varepsilon^{\text{W+jets mis-tag}}_{\text{MC}} = 0.0316 \pm 0.00124$, $\varepsilon^{\text{Z+jets mis-tag}}_{\text{MC}} = 0.0346 \pm 0.00171$, $\varepsilon^{t\bar{t}}_{\text{MC}} = 0.1678 \pm 0.0$

- **SF:** $SF_{pass}^{\text{W+jets mis-tag}} = 1.21 \pm 0.24$, $SF_{pass}^{\text{Z+jets mis-tag}} = 1.12 \pm 0.15$ ,$SF_{pass}^{t\bar{t}\text{ mis-tag}} = 1.033 \pm 0.058$

As studies have shown, tying together the W+jets and Z+jets backgrounds to have one floating SF and one $p_{\mathrm{T}}$-dependent component for V+jets does not impact our result significantly, and we prefer to leave them as separate parameters/nuisances, as there is a good physics motivation (different heavy flavor content and production mechanisms for W or Z) to do so.

#### 3.9.7. Additional Uncertainties

Additional rate uncertainties on the luminosity measurement (2.5%) and on the theoretical cross section for MC-driven processes are included and correlated through the different samples. We apply a 5% rate uncertainty to cover MET bin migration, as done in Akchurin et al., 2016. It is derived by propagating the jet energy scale and resolution uncertainties to the MET CMS JetMET POG, 2022; CMS JetMET POG, n.d., and is the signal uncertainty on the MET. An additional 4% rate uncertainty is considered to cover the effect of the CA15 jet energy scale. This is derived by propagating the AK8 jet energy scale uncertainties from the POG, with a factor of two added to cover the extrapolation from AK8 to CA15. These rate uncertainties are only applied to processes that are predicted by non-data-driven methods.

### 3.10. Results

This section presents a search for dark matter in the mono-Higgs final state. The exclusion limits for the Z'-2HDM and Baryonic Z' models are obtained using final distributions of the control regions and the signal region. However, no observed or expected exclusions are given in this section due to statistical limitations. The Z'-2HDM model yields the most stringent limits to date, and the Baryonic Z' model is also excluded for masses below a certain threshold.

The mass sideband of the signal region is used to perform a validation of the scale factors that the main backgrounds receive in the fit. Additional details on this validation may be provided in future results.

Expected and observed exclusions for the Z'-2HDM and Baryonic Z' models can also be found in future results. Other assumptions are made for the Z'-2HDM model, while for the Baryonic Z' model, couplings are assumed.

In the i2HDM model, various mono-jet signals are observed at the LHC, including the particle process $pp \to h_1 h_1 j$, where $h_1$ refers to the Higgs boson and $j$ refers to the jet produced in association with the Higgs boson. In this process, the dark matter particle candidate compensates for the momentum imbalance created by the single jet, and the Higgs mass, $M_H = 125$ GeV, serves as the carrier of dark matter.

Another process contributing to mono-jet is the particle process $q\bar{q} \to h_1 h_2 g$ ($gq \to h_1 h_2 g$), where $h_1$ and $h_2$ are two Higgs bosons with a small mass separation, and $h_2$ decays to $h_1$ plus a low-energy jet or lepton.

The samples used in this study include the $h_1 h_1$ set, consisting of 28 examples in different energy and mass ranges, and the $h_1 h_2$ dataset, consisting of 19 examples. Cross-sections were calculated by analyzing these samples and creating $n$-tuple groups.

The exclusion limits for the $h_1 h_1$ and $h_1 h_2$ processes were obtained using the

HiggsCombine tool within the CMSSW software package, based on the asymptotic limits approach. The cross-section values obtained were compared to the abandoned cross-section values to determine the limit value, as shown in Figure 3.20.. The results were presented and evaluated in group meetings, and future plans for the study were discussed.

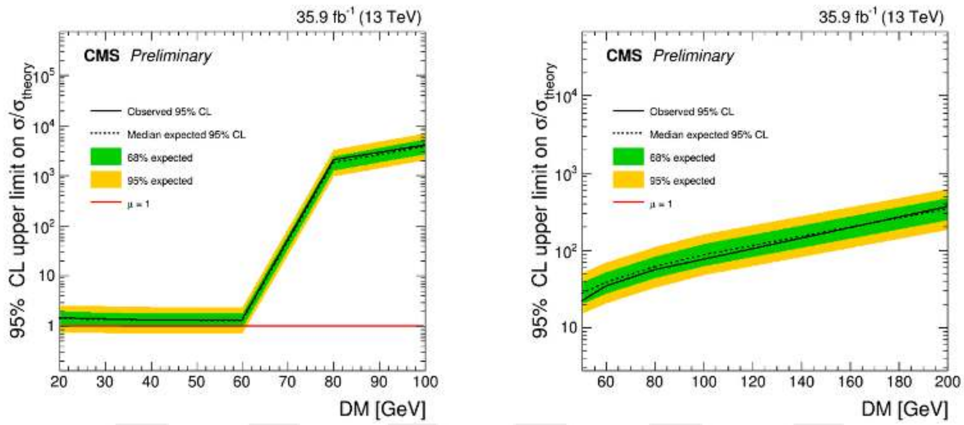



Figure 3.20. Exclusion limits for the $h_1h_1$ and $h_1h_2$ processes obtained by comparing the cross-section values with the abandoned cross-section values. The break in the $h_1h_1$ plot near 60 GeV indicates a region where the sensitivity of the experiment decreases, while the relatively straight line in the $h_1h_2$ plot with no significant drop in sensitivity indicates that the analysis is more effective in detecting signals in this mass range. The distance from the red line in the $h_1h_2$ plot could be due to a lower cross-section for the $h_1h_2$ process compared to the $h_1h_1$ process, or it could be due to differences in the analysis sensitivity between the two processes.

# 4. THE DUNE EXPERIMENT

Neutrinos, one of the most abundant yet least understood particles in the universe, have become increasingly prominent in scientific research due to their potential to unlock fundamental insights into the nature of matter and the universe as a whole. The Deep Underground Neutrino Experiment (DUNE) aims to study the properties of neutrino oscillations, pioneering an international collaboration of over 1,000 scientists and engineers from more than 30 countries. The DUNE experiment, shown in Figure 4.1., is located at the Fermilab research center in Illinois, USA, and consists of two detector modules placed deep underground at the Sanford Underground Research Facility in South Dakota, USA. Utilizing a beam of neutrinos generated by an intense proton beam, located at Fermilab, the detectors will measure the properties of neutrinos and provide new insights into their fundamental nature, including their masses and mixing angles (B. Abi et al., 2020c).

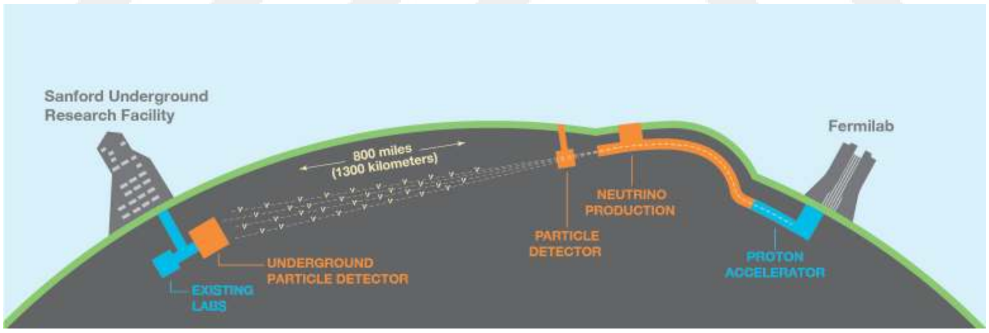



Figure 4.1. This figure depicts the three principal components of the DUNE experiment, namely the Beamline, the Near Detector, and the Far Detector. The illustration emphasizes the trajectory of the neutrino beam generated at Fermilab, which passes through the Near Detector before arriving at the Far Detector, located at the Sanford Underground Research Facility. (*DUNE illustration* 2023).

Neutrinos are of particular interest to scientists because of their potential to reveal insights into fundamental physics and the evolution of the universe. They are also essential to the life and death of stars, emitting neutrinos and anti-neutrinos, of all flavors, with energies of $\sim$10 MeV for a given supernova burst. Despite their abundance and importance, neutrinos remain mysterious and difficult to study. The DUNE experiment aims to change this by measuring the properties of neutrinos, including their masses and mixing angles (*DUNE.org* 2023).

The DUNE experiment is a long-term endeavor, expected to collect data for more than 20 years. During this time, scientists hope to uncover new subatomic phenomena that could transform our understanding of neutrinos and the universe as a whole. In particular, the experiment's detectors may shed light on the asymmetry between matter and antimatter in the universe, a mystery that has puzzled scientists for decades. One of the key strengths of the DUNE experiment is its international collaboration. With scientists and engineers from over 30 countries working together, DUNE represents a shining example of the power of global scientific endeavors. The experiment is also an important investment in the future, as it has the potential to spur technological innovation and inspire a new generation of scientists (*fnal.gov* 2022).

DUNE's main physics objectives are to measure leptonic CP-Violation and determine the neutrino mass hierarchy. The Near and Far Detector aims to measure both low and high energy spectra for neutrinos and antineutrinos simultaneously to obtain all neutrino oscillation parameters simultaneously. Determining $\Delta m^2_{31}$ is vital for understanding a nonzero CP-Violating phase $\delta_{CP}$, which would indicate leptonic CP-Violation. With DUNE's high precision measurements of $\nu_\mu$ and $\nu_e$ oscillations, we can determine the neutrino mass hierarchy with a significance of at least 5$\sigma$ within years of operations. In addition, it will reach 5$\sigma$ sensitivity to CP-Violation for 50% of possible values in approximately 10 years (Babak Abi et al., 2020b). Beyond

DUNE's main physics objectives, it aims to study supernova neutrino bursts, search for nucleon decay, and carry out dark matter and beyond the standard model (BSM) searches (Babak Abi et al., 2020a). DUNE's Far Detector is sensitive to near-Earth ($< 100$ kpc) supernova neutrino burst (SNB) events. Detecting such events can not only help DUNE, but this detection can also become an early supernova signal, signaling observatories to orient their telescopes in the same direction to gain more insight (via photon detection) and better understand the physical processes in the early stages of a star core-collapse and possibly, black hole formation (ibid.). In addition, the Far Detector *shields* against cosmic rays, since it is located 1.5 km underground. Together with its large active detector mass, this shield allows DUNE to search for nucleon decay. Grand unification theories (GUTs) aim to unify the electromagnetic, weak, and strong forces at high energies, predict that baryon number conservation is not exact, and that protons and bound neutrons are unstable. For DUNE to reach its physics goals, it must record a large volume of neutrino oscillation data over a considerable period. DUNE will start recording data while it is only partially completed, being modular by design, to produce the statistics needed. The amount of exposure required for measuring the neutrino mass hierarchy to at least $5\sigma$ depends on the final beam design and other oscillation parameters, namely $\delta_{CP}$ and the mass hierarchy itself.

### 4.1. DUNE Near Detector

The Near Detector's (ND) primary function is to provide a control of systematic uncertainties for the experiment. It provides a high-statistics definition of the neutrino beam close to the source, measuring the unoscillated neutrino interaction rate. This information is particularly useful for adjusting the neutrino interaction model used to move between the beam model and the observed data. The ND is also skilled in taking data at different off-axis beam positions, providing data sets of various beam spectra, and allows for the deconvolution of the beam and cross-section models (Acciarri et al., 2015). The conceptual design of the ND is based on the collaborative experience of many neutrino experiments, such as MINOS, MiniBooNE, T2K, NOvA, MINERvA, and the Short-Baseline Neutrino [SBN] program.

The DUNE ND is a multi-component detector: (1) liquid argon time projection chamber with pixelated readout (ND-LAr), (2) A high-pressure gas argon TPC in a magnetic field (ND-GAr), (3) System for on-Axis Neutrino Detection (SAND). The ND-LAr will use the LArTPC (liquid argon time projection chamber) technology and will operate at 0.6 T magnetic field. This technology will be discussed in more detail in Section 4.3. The ND-GAr will use the gas argon technology and will operate at 0.2 T magnetic field. Both detectors will be located 574 m from the neutrino source at Fermilab and will be able to measure the neutrino flux and spectra before oscillation. The ND-GAr will use the technology of time projection chambers with micro-pattern gas detectors (MPGD) and will allow the study of exclusive channels of neutrino interactions with argon, such as electron-neutrino elastic scattering and neutral current coherent pion production.

The ND-LAr will have a total mass of 17 t and will be approximately 6 m x 6 m x 6 m in size, while the ND-GAr will have a total mass of 0.5 t and will be 1 m x 1 m x 1 m in size. The ND-LAr will provide a neutrino flux measurement with an uncertainty of less than 1%, and the ND-GAr will measure the flux of the un-oscillated beam to an uncertainty of less than 3%. The Near Detector systems are expected to collect more than 100 TB of raw data per year, with a data rate of up to 5 GB/s, which presents a significant challenge for data acquisition and processing (B. Abi et al., 2018).

## 4.2. DUNE Far Detector

The DUNE Far Detector (FD) layout is based on state-of-the-art LArTPC technology, which provides excellent neutrino energy regeneration and imaging capabilities. The DUNE FD consists of four independent cryostats with 70 kt of liquid argon at 87 K (-186◦ C). Each cryostat houses a detector module based on LArTPC technology, with a functional detector volume of approximately 10 kt, for a total of 40 kt sensitive FD volume for DUNE. The internal dimensions of a DUNE FD cryostat are approximately 15.1 m wide, 14.0 m high, and 62 m long.

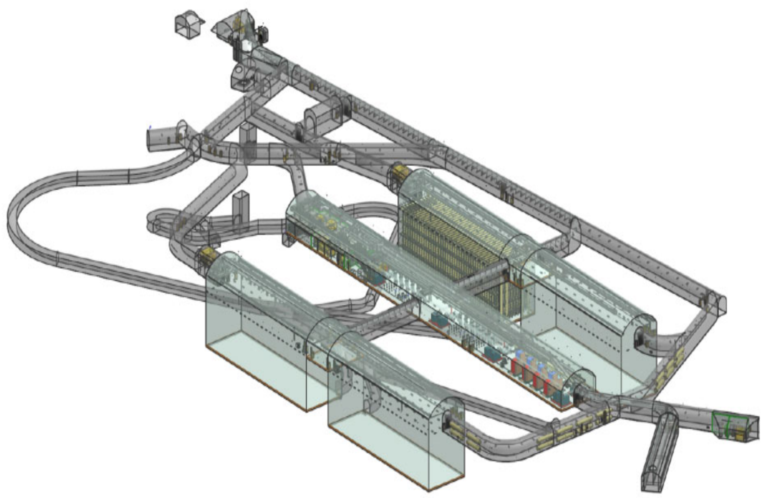


Figure 4.2. The DUNE FD cavern located at the Sanford Underground Research Facility in Lead, South Dakota. The detector will consist of four individual modules, each filled with a 10 kt fiducial (17 kt actual) volume of liquid argon (B. Abi et al., 2020b).

The FD complex is necessary to ensure the safe and stable operation of the detector during several years of data acquisition. Therefore, vital research and development are underway to meet the DUNE requirements. Located at CERN in Geneva, Switzerland, the ProtoDUNE-SP detector is the first completed large-scale prototype of the first DUNE FD module. Figure 4.2. shows the DUNE FD cavern, located at the Sanford Underground Research Facility in Lead, South Dakota. The detector will consist of four individual modules, each filled with a 10 kt fiducial (17 kt actual) volume of liquid argon.

### 4.3. Concept of LArTPC

LArTPC is a detector that performs calorimetric measurements, particle identification, and reconstructs a three-dimensional track from an incoming charged particle traversing LAr. When a charged particle crosses the active LAr volume of the TPC, it leaves an ionization track composed of ionized atoms and electrons (Rubbia et al., 1977). In a LArTPC, an anode and cathode plane generate an electric field that drifts the ionized free electrons from the charged argon atoms and starts to drift parallel to the electric field. The uniformity of the field is provided by a field cage surrounding the active volume. The LArTPC is illustrated in Figure 4.3..

LArTPC technology plays a crucial role in the DUNE experiment. The Far Detector of DUNE consists of four LArTPC modules, each containing over 10 ktons of liquid argon. LArTPC technology provides high-precision 3D imaging of neutrino interactions, allowing for the identification and study of different particle types, including electrons, muons, and pions.

LArTPC technology is ideal for studying neutrinos because it allows for the reconstruction of the direction and energy of each interacting particle. The FD modules of DUNE use this technology to detect neutrinos that have traveled 1,300 kilometers from their source at Fermilab. When a neutrino interacts with the argon nuclei in the detector, it produces charged particles that ionize the argon atoms. These charged particles are then detected and recorded, providing information on the properties of the neutrino. The high-precision imaging capabilities of LArTPC

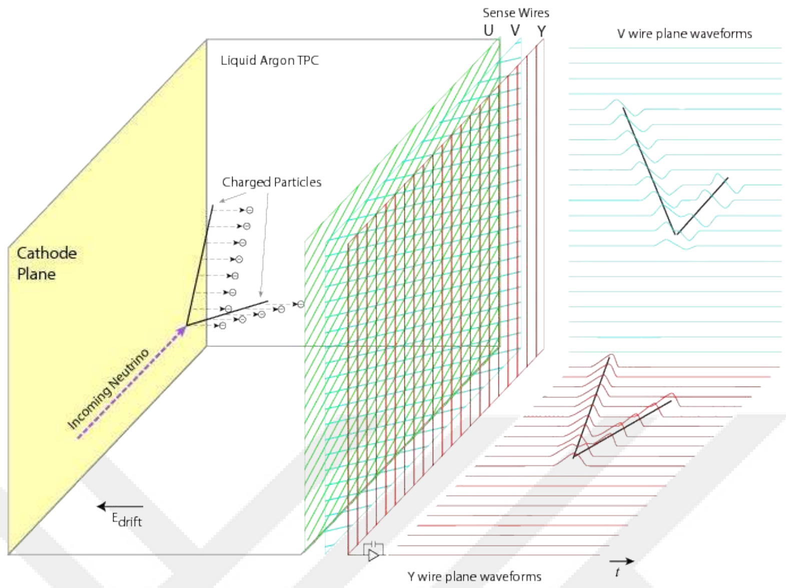



Figure 4.3. This figure shows an example of a LArTPC, which consists of a cathode (left) and an anode (right). The cathode is biased negative, with a high potential, and the anode is set to ground. The result is an electric field that drifts ionization electrons towards the anode, which produces a signal. The ionization signal is detected using a set of induction and collection wire planes wrapped around the Anode Plane Assembly (APA). The far right of the figure shows two 2D projections viewing the same track (Acciarri et al., 2017).

technology, combined with the long baseline of the DUNE experiment, allow for the study of neutrino oscillation at a level of precision not achievable by any other experiment to date. This precision allows for the study of important open questions in neutrino physics, such as the neutrino mass hierarchy and the existence of CP-violation in the leptonic sector.

In addition to its role in DUNE, LArTPC technology has been used in other neutrino experiments, such as the MicroBooNE and ICARUS experiments. The success of LArTPC technology in these experiments and its pivotal role in DUNE highlight the importance of developing and improving this technology for future neutrino experiments (Rubbia et al., 1977).

### 4.4. ProtoDUNE-SP

The ProtoDUNE-SP is a large-scale prototype of one of the four modules of the DUNE FD. It is a single-phase LArTPC located at CERN in Geneva, Switzerland. The ProtoDUNE-SP detector is the first completed prototype of the first DUNE FD module, and it is a vital step toward the construction and operation of the full-scale DUNE FD. The main goal of ProtoDUNE-SP is to demonstrate the functionality and performance of the LArTPC technology on a larger scale and to provide data for the ongoing research and development of the DUNE far detector.

The event reconstruction capability combined with the exposure of the detector's large active volume to the CERN charged-particle beams leads the way to investigating new physics and detector development for particle interaction processes. The use of the CERN charged-particle beams of known particle type and incident energy provides the means to achieve a better understanding of the interaction processes occurring within an argon target and helps to optimize event reconstruction techniques, particle identification algorithms, and calorimetric energy measurements. Beam measurements also provide a calibration data set for tuning Monte Carlo simulations and a reference data set for the DUNE experiment.

ProtoDUNE-SP's large active volume allows for good containment of the hadronic and electromagnetic interaction products in the few GeV range, while the LArTPC features precise 3D tracking and accurate measurement of energy deposited. This combination is a first of its kind in one detector.

Pion and proton beams in an energy range from about half to a few GeV were used primarily to study hadronic interaction mechanisms and secondary particle production. At higher energies, these beams are used to study shower reconstruction and energy calibration. Electrons are used to benchmark and tune electron/photon separation algorithms, to study electromagnetic cascade processes, and to calibrate electromagnetic showers at higher energies. Charged kaons produced in the tertiary beamline are rare but are copiously produced by the pion beam interactions inside the detector. These are extremely useful for characterizing kaon identification efficiency for proton decay sensitivity studies. Samples of stopping muons with Michel electrons from muon decay (or without them, in the case of negative muon capture) are used for energy calibrations in the low-energy range of the SN neutrino events and for the development of charge-sign determination methods.

ProtoDUNE-SP has collected more than $10^6$ events in each of the following: beam data, cosmic data, random trigger data, and calibration data. ProtoDUNE's achievements have included understanding test-beam data for different particle types at momenta 0.3 – 7 GeV/c. Our first performance results have been published B. Abi et al., 2020a and are discussed in detail in the next chapter. Furthermore, ProtoDUNE-SP has been key in validating full-scale DUNE detector technology and engineering components, continuing to demonstrate the long-term operational stability of all detector components. The technical components of the experiment will be discussed in this chapter and are detailed further in our technical paper (Abud et al., 2022). ProtoDUNE-SP is the prototype of the DUNE as shown in Figure 4.4.

ProtoDUNE-SP uses many of the full-scale components the DUNE FD-Horizontal Drift (HD) plans to use. However, it being a prototype, ProtoDUNE-SP uses 6 Anode Plane Assemblies (APAs) instead of 150, and two drift volumes instead of four. As shown in Figure 4.5. the active volume is split in two by a central cathode plane made of three Cathode Plane Assemblies (CPAs).In addition, the reduced detector (relative to an FD module) also reduces the requirements for data acquisition, the amount of liquid argon, diagnostic subdetectors, and many other

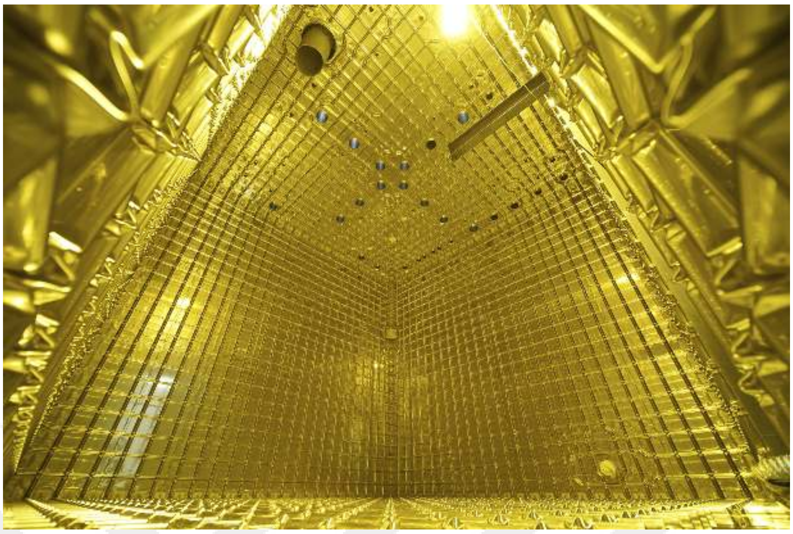


Figure 4.4. ProtoDUNE-SP, inside of the cryostat.

detector-related systems, like the Photon Detection System (PDS).

ProtoDUNE-SP tests for the internal electronics, both those handling beam particle signals (charge + light) and those meant for diagnostics. One core principle of the planned DUNE FD is the electron drift length of 3.6 m, which corresponds to the distance between the cathode and anode planes in the TPC.

APAs are the most crucial building blocks of the DUNE FD-HD; they detect ionization electrons liberated within the detector and measure their properties, such as total charge, position, and time of arrival. Their wire planes are complex, which requires understanding the proper fabrication, as it is difficult to modify.

Another primary goal of ProtoDUNE-SP was to identify the optimal light detection technology for scintillation detection. ProtoDUNE-SP allowed us to use a charged particle beam of known composition and momentum, measuring the light response using three independent photon detector technologies. DUNE's FD-HD will have 150 APAs per module, totaling 600 for the 4× 10kt FD-HD modules. Each

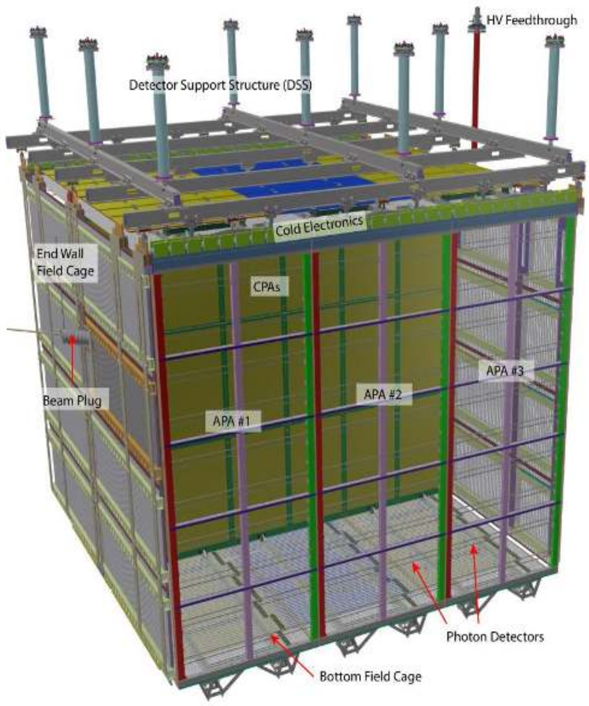



Figure 4.5. 3D model of the ProtoDUNE-SP detector with the labeling of all major components (B. Abi et al., 2020b).

APA can integrate 10 of the PDS modules per APA. Therefore, ProtoDUNE-SP was essential for full-scale testing APAs and determining which PDS detector technology is optimal is vital to understand before production of the DUNE FD.

# 5. STUDY OF PROTODUNE-SP SCINTILLATION

Liquid argon scintillation is a process in which a high-energy particle interacts with liquid argon and excites its atoms, causing them to emit light. This light is then detected by photosensors, such as photomultiplier tubes or silicon photomultipliers, which are used to reconstruct the particle's energy and position. The scintillation light produced by liquid argon is in the ultraviolet range, with a peak emission wavelength of about 128 nm. This process is one of the key features of detectors used in neutrino and dark matter experiments, as it provides an additional signal to the ionization signal used for tracking charged particles.

In January 2020, a stand-alone X-Arapuca detector was installed in the ProtoDUNE-SP cryostat system. The X-Arapuca is the light detection system of the DUNE single-phase far detector. It comprises Silicon Photo Multipliers (SiPMs) sensors, a reflective surface, and a cavity between the photo sensors and the reflective surface. 127 nm photons arrive at the X-Arapuca. They are converted and then reflected in the internal surface back and forth until they reach the photo-sensor array. The X-Arapuca data acquisition was operated between February and July 2020 during the COVID-19 pandemic. The purpose of the operation was to validate the performance of the X-Arapuca detector, the liquid argon TPC performance, and the light detection system as a function of xenon injection.

Xenon injection refers to introducing a small amount of gaseous xenon into a liquid argon detector to study the response of the detector's light detection system. When the xenon atoms interact with the argon, they produce scintillation light, which photosensors can detect. By measuring the amount and characteristics of this light, researchers can gain valuable information about the properties of the detector and improve its performance.

Xenon injection can also help mitigate nitrogen contamination's impact in liquid argon detectors. Nitrogen is a common impurity in liquid argon that can negatively impact the detector's performance by reducing the scintillation light yield and slowing the drift velocity of ionization electrons. By injecting a small amount of gaseous xenon into the detector, the xenon atoms can act as a scavenger to remove the nitrogen impurities from the liquid argon. This process, known as cryogenic distillation, has been used in various liquid argon experiments to improve their performance. During cryogenic distillation, the xenon atoms absorb nitrogen impurities in the liquid argon, forming a xenon-nitrogen mixture. This mixture is then removed from the detector and processed to separate the xenon and nitrogen. The purified xenon is then returned to the detector while the nitrogen is removed. This technique has been used successfully in several liquid argon experiments, including the ICARUS and DarkSide detectors, to remove nitrogen contamination and improve their performance. It has also been tested in ProtoDUNE-SP, and the results have shown that xenon injection can effectively reduce the amount of nitrogen in the detector and improve its performance (Acciarri et al., 2008).

## 5.1. X-Arapuca

Figure 5.1. illustrates the trapping mechanisms of the standard ARAPUCA (right) and the X-Arapuca (left). The standard ARAPUCA is a light-trapping device consisting of two refractive surfaces read by SiPMs (Machado, 2016). When 128 nm photons strike the detector, they are shifted to 350 nm by a thin p-terphenyl (pTp) coating on top of a dichroic filter with a 400 nm transparency cutoff. A second coating layer, Tetraphenyl Butadiene (TPB), converts 350 nm photons to 420 nm. In contrast, the X-Arapuca is an elevated version of the Arapuca that replaces the second coating layer with a wavelength shifter light guide, improving photon collection efficiency

(Machado, 2018). As a result, the produced 420 nm photons are trapped inside the detector and are fully reflective above the 400 nm cutoff, leaping back and forth until they reach the SiPMs. The X-Arapuca is expected to provide improved performance


Figure 5.1. Comparison of the trapping mechanisms of the standard ARAPUCA (right) and the X-Arapuca (left). The X-Arapuca traps scintillation photons through total internal reflection, improving photon collection efficiency and enabling it to capture more scintillation photons than the standard ARAPUCA.

compared to other light detection systems in liquid argon experiments. It is currently being tested and evaluated in the ProtoDUNE-SP detector at CERN, and the results have shown promising performance in terms of light yield and signal-to-noise ratio. The X-Arapuca's trapping through total internal reflection allows it to capture more scintillation photons than the standard ARAPUCA, improving photon collection efficiency and providing a more accurate measurement of particle interactions.

The X-Arapuca consists of a rectangular box with a series of dichroic filters and wavelength shifters that direct scintillation photons to two photon detectors. The box has a window that allows scintillation light to enter and can be made of either quartz or non-quartz materials. The choice of window material can affect the detector's performance, with quartz windows providing better transparency and non-quartz windows providing better mechanical stability.

Overall, the X-Arapuca represents a significant advancement in the development of light detection systems for liquid argon experiments, and its promising performance in the ProtoDUNE-SP detector suggests that it may be a valuable tool in future neutrino and dark matter experiments.

### 5.2. Data Acquisition

The experiment involved taking over 302 runs under different conditions, including varying drift field values and Xe concentrations, each containing 10,000 fixed cosmic-triggered events. These runs were collected for six different X-Arapuca channels, resulting in a total of $18.12x10^6$ signals. In this study, a subset of $7.92x10^6$ signals were used to select a steady concentration and zero electric field conditions for analysis. The Xe concentration was varied over five different eras, with different amounts of Xe concentration in parts per million (ppm): 1.10 ppm, 4.20 ppm, 11.60 ppm, 16.00 ppm, and 18.80 ppm.

Figure 5.2. shows changes in Xenon concentration and light as a function of time. The figure displays a time series data representation of the integrated light ratio of quartz and non-quartz windows at a steady electric field of zero. Each color represents a steady xenon concentration era. More information about the runs and data collection process can be found in the RunTables standalone available at the following link: `https://docs.google.com/spreadsheets/d/1j22lp80CuadYQdBHWP0Ihcr5mThz8V35QL-e9XvqxN0/edit?usp=sharing`.

Two types of X-Arapuca were used in the data collection process: No window and quartz window. The readout channels Ch1, Ch2, and Ch3 used the No window type, which were sensitive to the 127 nm scintillation wavelength of liquid argon, while the readout channels Ch4, Ch5, and Ch7 used the quartz window type, which filtered the UV light on the top of the wavelength shifter. This configuration was important in understanding the effects of Xe concentration on the experiment.

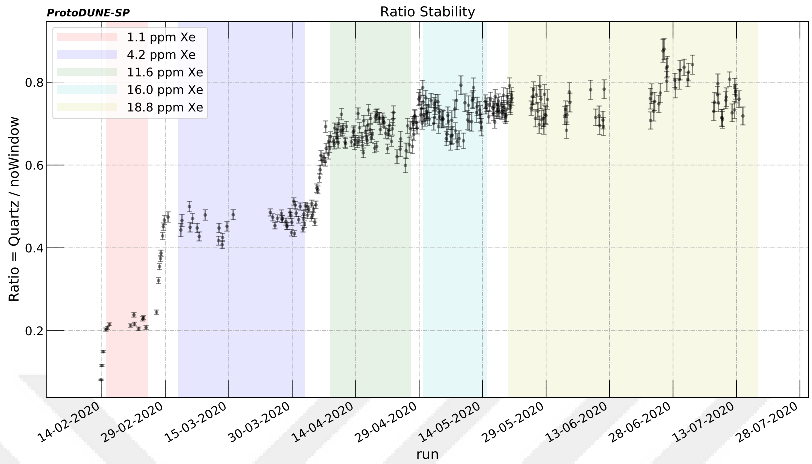



Figure 5.2. A time series data representation, ’E=0’, an integrated light ratio of quartz and non-quartz windows. Each color expresses steady xenon concentration eras.

### 5.3. Noise and Baseline Subtraction

The first challenge in our analysis was to mitigate the unknown noise sources on the waveforms. After performing a visual scan of a large number of waveforms, we found that the noise was mainly induced by trigger electronics. We then used a Supervised Machine Learning (ML) model to distinguish a subset of triggered events with no detectable physical signal. This model was trained using the extensive data collected in the ProtoDUNE-SP experiment. Figure 5.3. shows an example of the averaged waveforms found by the ML model for randomly selected runs across all channels.

We found that approximately 7% of the total recorded triggers were empty events, meaning they had no detectable physical signal. We used a linear regression algorithm to distinguish between the signal and noise waveforms. This involved measuring the slope of the waveform, which is expected to be different between signal and noise, with noise having a slope close to zero and the signal having a non-zero

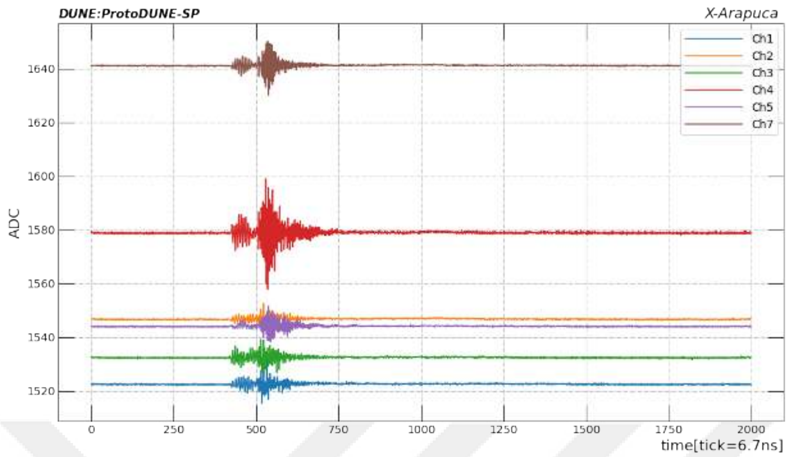


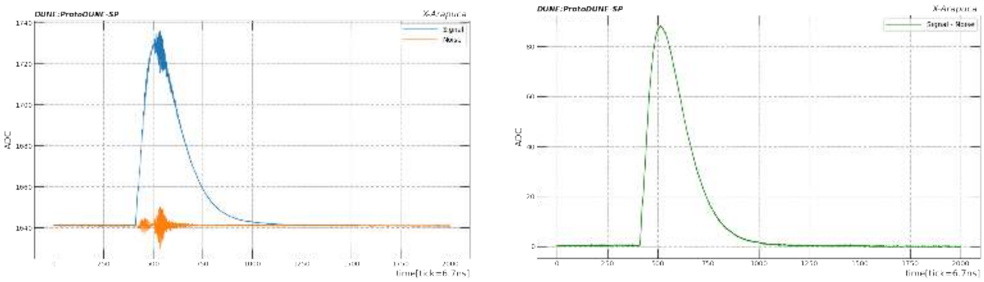

Figure 5.3. Averaged waveforms found by the ML model for randomly selected runs across all channels.

slope due to the presence of a physical signal.

We used the averaged empty triggers to remove the noise from the actual waveform, which also removed the baseline drift automatically. This process not only provided the additional benefit of baseline subtraction without additional calculation but also made it possible to correct the overshoot and undershoot of the signals. Figure 5.4.a shows the signal and noise separation for a randomly selected run, which is divided into signal and noise averaged waveforms for all channels.

In our case, the baseline around 1640 ADC was automatically subtracted by the noise removal algorithm, resulting in a clean signal with no baseline drift. This was particularly important for our subsequent analysis, as the presence of baseline drift could lead to inaccuracies and errors when applying deconvolution techniques to the data. The automatic removal of the baseline drift allowed us to more accurately perform deconvolution on the signal, resulting in more precise measurements of the physical processes of interest.

Furthermore, the noise removal algorithm allowed us to correct the overshoot




(a) The run is divided into signal and noise averaged waveforms for all channels. The x-axis unit is time ticks, each corresponding to 6.7ns.

(b) Clean signal after noise removal. The x-axis unit is time ticks, each corresponding to 6.7ns.

Figure 5.4. Signal and noise separation using machine learning for a randomly selected run. (a) shows the signal and noise separation for the run, while (b) shows the resulting clean signal after noise removal. The noise removal algorithm results in a clean signal with no baseline drift or noise.

and undershoot of the signals, which could lead to incorrect measurements of the signal amplitude. The resulting clean signal allowed us to accurately measure the pulse width and amplitude, which are important parameters for characterizing the detector response.

In summary, the noise and baseline subtraction using machine learning was a crucial step in our analysis pipeline. It allowed us to remove the unknown noise sources and baseline drift from the waveforms, resulting in clean signals that were suitable for subsequent analysis. The resulting clean signals allowed us to accurately measure the pulse width and amplitude, which are important parameters for characterizing the detector response and for making precise measurements of the physical processes of interest.

### 5.4. Single Photo-Electron Response and Deconvolution

Ensuring the stability of the light collection system is critical for subsequent analysis during the Xe doping period. To monitor this, the response of each

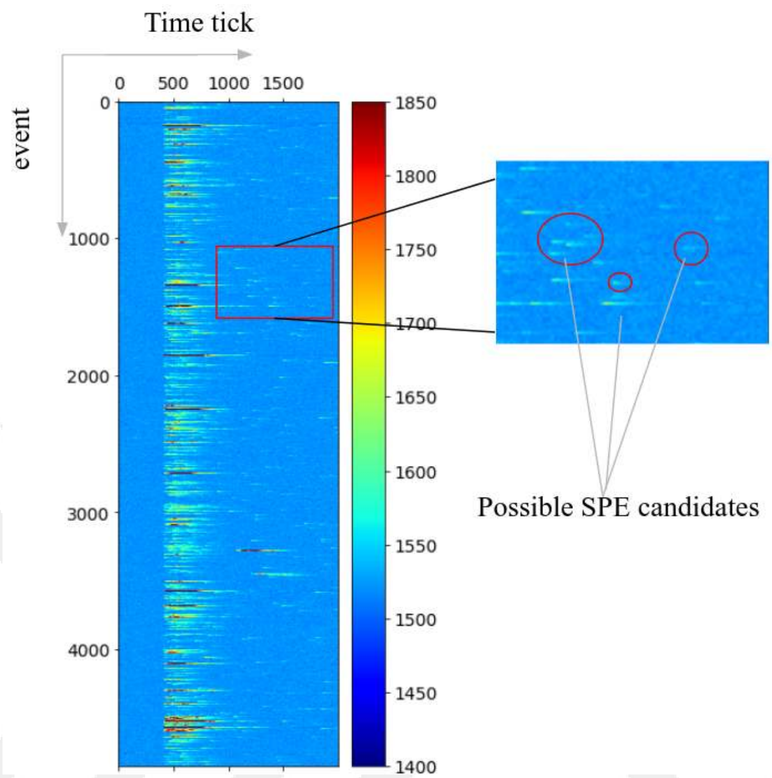



Figure 5.5. 2D plot showing waveforms, with the Y-axis representing the waveform number, the X-axis representing time ticks, and the Z-axis representing the ADC value.

X-Arapuca channel to a single photo-electron (SPE) is regularly extracted. This also allows the calibration of the collected light.

SPE pulses are typically located in the tail of each signal. Visualizing the SPE response is often accomplished by representing waveforms in a 2D plot, where the Y-axis shows the waveform number, the X-axis represents the time ticks, and the Z-axis represents the analog-to-digital converter (ADC) value. An example of such a plot is shown in Figure 5.5.

This plot assists in localizing the SPE response on the waveform, which is necessary for further analysis, such as deconvolution. Collected SPE charge distributions are fitted with a double Gaussian, one for the pedestal and the other for the actual SPE peak. Then, the averaged SPE waveform is extracted by selecting the subset of pulses with a charge within one sigma of the mean SPE value. This average waveform represents the SPE response of the detector and is later used for the deconvolution of multi-photon pulses.

Deconvolution is necessary because the shape of the SiPM signal washes out the timing information due to the convolution of overlapping SPE waveforms. A reliable deconvolution algorithm extracts the timing information. Figure 5.6. displays the extracted SPE distribution for each XA channel, fitted with a double Gaussian function to account for the pedestal and the actual SPE peak.

This process is regularly repeated during the Xe doping period. Figure 5.6. displays the extracted SPE distribution for each XA channel, while Figure 5.7. shows the SPE charge stability across the entire run period. The X-Arapuca system demonstrated stable operating conditions during the entire doping campaign.

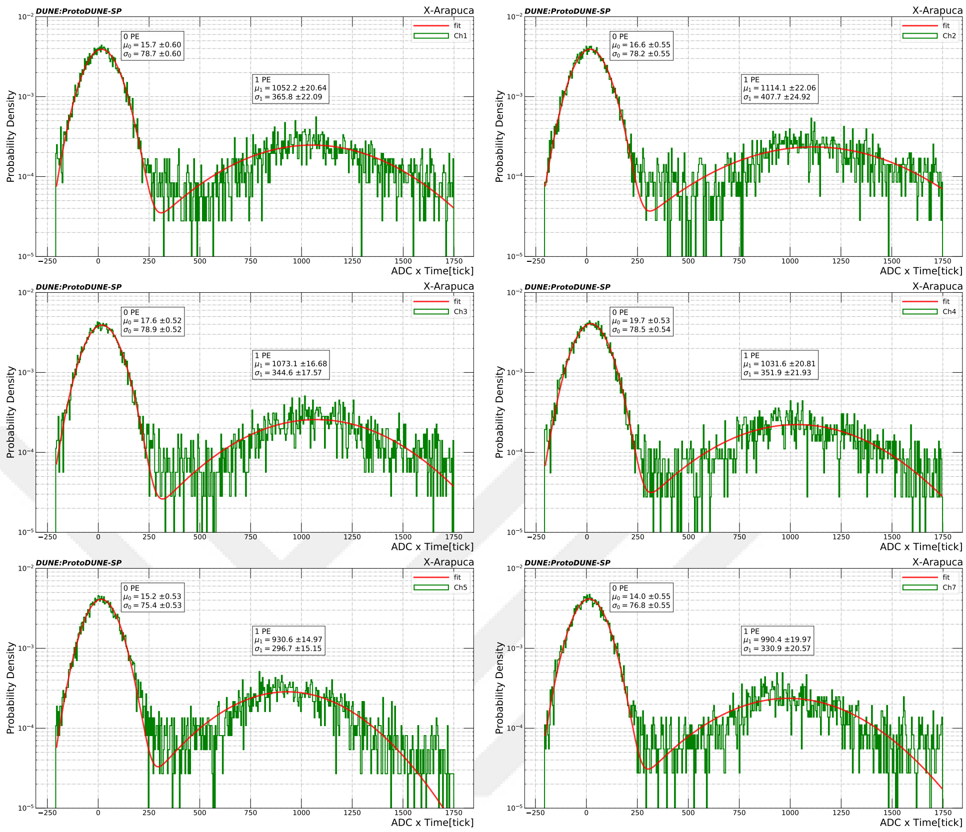



Figure 5.6. SPE response distribution for each X-Arapuca channel, fitted with a double Gaussian function to account for the pedestal and the actual SPE peak. The first Gaussian represents the baseline of the signal, while the second Gaussian represents the signal generated by the detection of a single photon.

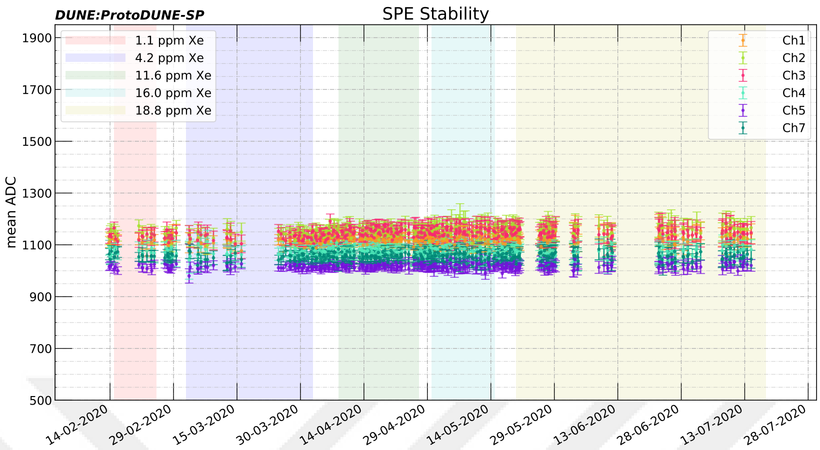



Figure 5.7. Average SPE charge stability for all runs and each channel. Runs cover an overall six-month doping period; coloured areas represent specific dopings.

### 5.5. Ar-Xe excitation transfer model in the presence of nitrogen

The excitation transfer model from $Ar_2^*$ to $Xe_2^*$ in the presence of nitrogen is described by a set of equations, which detail the concurrent processes involved. The first process involves the decay of $Ar_2^*$ dimers to two argon atoms by emitting light at 128 nm. The triplet states of $Ar_2^*$ can also interact non-radiatively with xenon atoms to form an intermediate excited state $ArXe^*$. Depending on the xenon concentration, a fraction of $ArXe^*$ molecules will de-excite, emitting 150 nm radiation, while the remaining molecules will interact with single xenon atoms, producing $Xe_2^*$ dimers. These dimers will decay with decay constants of 4 and 22 ns, emitting 175 nm photons.

In the presence of a quencher such as nitrogen, all the aforementioned processes occur concurrently with non-radiative interactions with nitrogen, leading to dimer destruction without light emission. The $Ar_2^*$ triplet state $^3\Sigma_u^+$ is particularly affected due to its very large lifetime, but all reactions are affected to some degree, depending on the concentration of both the quencher ($N_2$) and the dopant (Xe).

In this section, drawing from previous literature results (McFadden et al., 2021), we detail the model for the excitation transfer from $Ar_2^*$ to $Xe_2^*$ in the presence of nitrogen. The concurrent processes involved are as follows:

- $Ar_2^*$ dimers would normally decay to 2 Ar atoms by emitting light at 128nm, with the usual two very different decay times already mentioned in the introduction;
- in the presence of xenon, non-radiative interactions with the long-lived triplet states of $Ar_2^*$ can lead to the formation of intermediate excited state $ArXe^*$;
- According to the concentration of xenon (up to around 1ppm), a fraction of $ArXe^*$ molecules will de-excite, emitting a characteristic 150nm radiation; the rest (all of them at higher concentrations) will interact again with single xenon

atoms, producing $Xe_2^*$ dimers.

- $Xe_2^*$ dimers will then decay with decay constants of 4ns and 22ns, emitting 175nm photons.
- If a quencher is present in the liquid argon (like it is the case with $N_2$ for ProtoDUNE-SP), then all described processes become concurrent with non-radiative interactions with $N_2$, that cause dimer destruction, without light emission. The most affected species is the $Ar_2^*$ triplet state $^3\Sigma_u^+$, due to its very large lifetime, however all reactions are affected, according to the concentration of both the quencher ($N_2$) and the dopant (Xe).

Specifically, equation 5.1 represents the formation of an intermediate excited state of argon and xenon, equation 5.2 represents the formation of xenon dimers from this intermediate state, and equation 5.3 represents the subsequent decay of these dimers through the emission of photons.

$$Ar_2^* + Xe \rightarrow ArXe^* + Ar \tag{5.1}$$

$$ArXe^* + Xe \rightarrow Xe_2^* + Ar \tag{5.2}$$

$$Xe_2^* \rightarrow Xe + Xe + \gamma(175\,nm) \tag{5.3}$$

$$\frac{\mathrm{d}(Ar_2^*)}{\mathrm{d}t} = -\frac{Ar_2^*}{\tau_{128}} - \frac{Ar_2^*}{\tau_{N2}} - \frac{Ar_2^*}{\tau_{AX}} \equiv -\frac{Ar_2^*}{\tau_{TA}} \tag{5.4}$$

$$\frac{1}{\tau_{TA}} \equiv \frac{1}{\tau_{128}} + \frac{1}{\tau_{N2}} + \frac{1}{\tau_{AX}} \tag{5.5}$$

$$\frac{\mathrm{d}(ArXe^*)}{\mathrm{d}t} = +\frac{Ar_2^*}{\tau_{AX}} - \frac{ArXe^*}{\tau_{150}} - \frac{ArXe^*}{\tau_{N2}} - \frac{ArXe^*}{\tau_{TX}} \tag{5.6}$$

$$\equiv +\frac{Ar_2^*}{\tau_{AX}} - \frac{ArXe^*}{\tau_{TX}} \tag{5.7}$$

$$\frac{1}{\tau_{TX}} \equiv \frac{1}{\tau_{150}} + \frac{1}{\tau_{N2}} + \frac{1}{\tau_{XX}} \tag{5.8}$$

$$\frac{\mathrm{d}(Xe_2^*)}{\mathrm{d}t} = +\frac{Ar_2^*}{\tau_{XX}} - \frac{Xe_2^*}{\tau_{175}} \tag{5.9}$$

Here the various time constants $\tau_i$ are related to decay processes or interaction rates. For example, $\tau_{AX}$ and $\tau_{XX}$ are related to the rates of interactions 5.1 and 5.2, respectively. $\tau_{TA}$ and $\tau_{TX}$ instead summarise the average decay time of excited dimers, due to many concurrent processes (actual decay, destruction due to non-radiative collisions). Finally, $\tau_{N2}$ represents the nitrogen quenching effect due to the presence of a fixed amount of $N_2$ in ProtoDUNE-SP.

In the specific study of light emission, two different configurations of X-ARAPUCA detectors are used. One of them, denoted as $XA_{NQ}$, does not have a quartz window and is sensitive to the sum of all three spectrum components. The other, denoted as $XA_Q$, features a quartz window that cuts all wavelengths except 175 nm.

The equations for the two supercells are given as:

The ”Ar+Xe-XA” supercell is sensitive to all three components:

$$\frac{\mathrm{d}XA_{NQ}}{\mathrm{d}t} = K\left[\frac{1}{\tau_{128}}\frac{e^{-t/\tau_{TA}}}{\tau_{TA}} + \left(\frac{\tau_{XX}+\tau_{150}}{\tau_{XX}\tau_{150}}\right)\left(\frac{\tau_{TA}\tau_{150}}{\tau_{AX}}\right)\frac{(e^{-t/\tau_{TA}} - e^{-t/\tau_{TX}})}{(\tau_{TA}-\tau_{TX})}\right] \tag{5.10}$$

The ”Xe-XA” supercell is only sensitive to 175 nm:

$$\frac{dXA_Q}{dt} = K\frac{1}{\tau_{XX}}\frac{(e^{-t/\tau_{TA}} - e^{-t/\tau_{TX}})}{(\tau_{TA} - \tau_{TX})} \tag{5.11}$$

The theoretical model predicts that the argon fast component should not be present in the waveforms of the ”Xe-XA” detector. However, a residual argon fast component is observed, which may be due to poor light tightness of the supercell. Therefore, in order to extract the time constants of the processes relevant to the ternary Ar-Xe-$N_2$ mixture studied in this experiment, waveforms from both X-ARAPUCAs are fitted with three exponential functions. The coefficients of the fits are dependent on the quantity of dopant (Xe) and are reported in Table 5.1..

To measure the inter-calibration between the Quartz and non-Quartz window detectors in this experiment, the ratio of XQ/XN is considered. This ratio is calculated using the equation:

$$\frac{XQ}{XN} = \frac{\tau_{150}}{\tau_{XX} + \tau_{150}} \tag{5.12}$$

Here, $\tau_{150}$ is the time constant of the 150 nm region, $\tau_{XX}$ is the time constant of the non-Quartz window, and $XQ$ and $XN$ are the average signals in the Quartz and non-Quartz window detectors, respectively. The inter-calibration data is shown in Figure 5.8., which illustrates the strong correlation and validity of the inter-calibration method between the two detector types.

The simultaneous fit of waveforms from both detectors allows for the extraction of $1/\tau_{TA}$ and $1/\tau_{TX}$ at all concentrations, which can be plotted as a function of Xe concentration in Figure 5.10.a. Linear fits are superimposed on the data points, and it should be noted that in Equations (5.5) and (5.8), the coefficients $1/\tau_{128}$, $1/\tau_{150}$, and $1/\tau_{N2}$ are assumed to be independent of Xe concentration and are represented by the

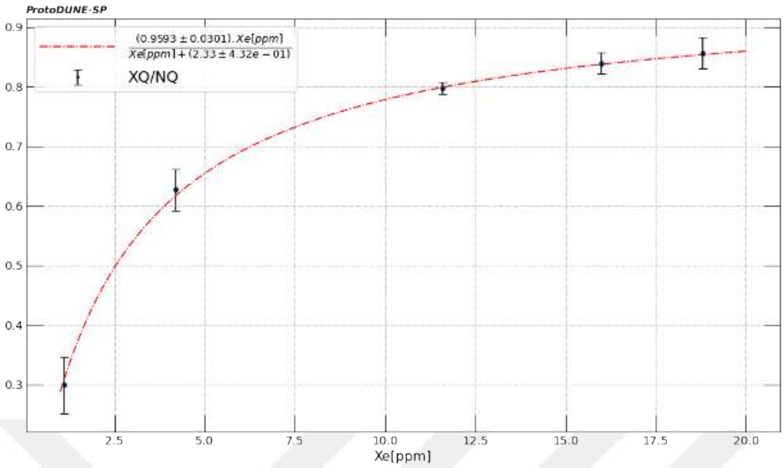



Figure 5.8. Inter-calibration data between Quartz and non-Quartz window detectors as a function of Xe concentration, demonstrating strong correlation and validity of inter-calibration method.

intercept of the fit. Similarly, $1/\tau_{AX}$ and $1/\tau_{XX}$ are dependent on Xe concentration and are represented by the slope of the linear fit.

The pulse charge (Q) dependence on the ratio of charge collected in the narrow and wide readout channels (NQ), which reflects the position of the ionization in the detector, is shown in Figure 5.9.. The figure displays simultaneous fits of the signals collected with and without a quartz window for five Xe injection concentrations. The fits provide information about the light collection efficiency and help to characterize the performance of the detector at different Xe concentrations.

Table 5.1. Values extracted for time constants of the characteristic processes involved in the ternary Ar-Xe-$N_2$ mixture studied in ProtoDUNE-SP. The values depend on xenon concentration and derive from fitting average waveforms collected with both X-ARAPUCA detectors ($\tau_{fast}$ is the variable used to fit the residual fast component present in both detectors waveforms).

| **Xe**[**ppm**] | $\tau_{\mathbf{fast}}$[**ns**] | $\tau_{\mathbf{TA}}$[**ns**] | $\tau_{\mathbf{TX}}$[**ns**] | $\tau_{\mathbf{AX}}$[**us**] | $\tau_{\mathbf{XX}}$[**us**] |
|---|---|---|---|---|---|
| 1.1 | 26.64 $\pm_{0.34}$ | 589.60 $\pm_{76.33}$ | 1220.08 $\pm_{35.89}$ | 4.79 $\pm_{0.13}$ | 12.77 $\pm_{0.18}$ |
| 4.2 | 24.91 $\pm_{0.30}$ | 495.57 $\pm_{12.77}$ | 1013.93 $\pm_{12.17}$ | 1.25 $\pm_{0.03}$ | 3.34 $\pm_{0.05}$ |
| 11.6 | 26.09 $\pm_{0.32}$ | 270.59 $\pm_{2.21}$ | 711.85 $\pm_{2.58}$ | 0.45 $\pm_{0.01}$ | 1.21 $\pm_{0.02}$ |
| 16.0 | 27.52 $\pm_{0.54}$ | 234.10 $\pm_{2.64}$ | 545.28 $\pm_{3.00}$ | 0.33 $\pm_{0.01}$ | 0.88 $\pm_{0.01}$ |
| 18.8 | 30.99 $\pm_{0.99}$ | 207.13 $\pm_{3.43}$ | 488.53 $\pm_{3.80}$ | 0.28 $\pm_{0.01}$ | 0.75 $\pm_{0.01}$ |

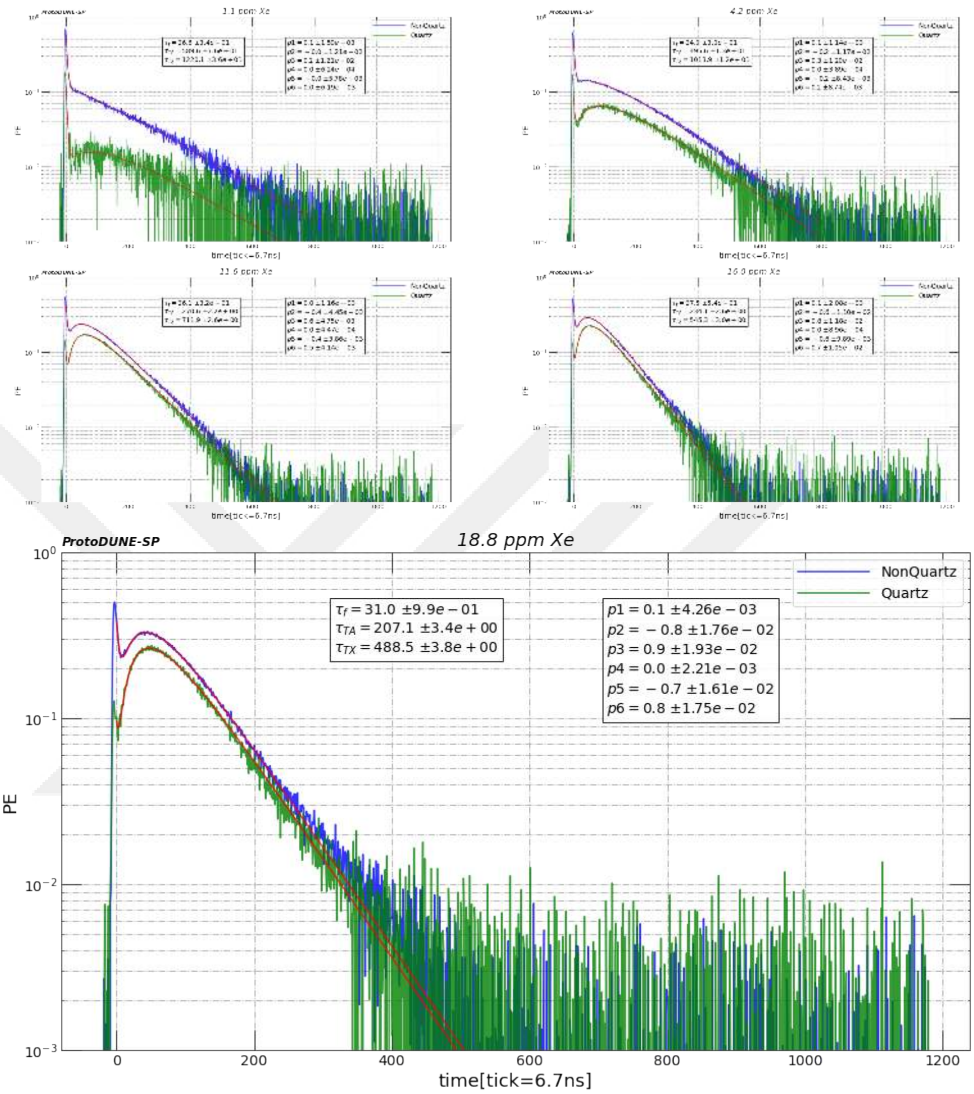
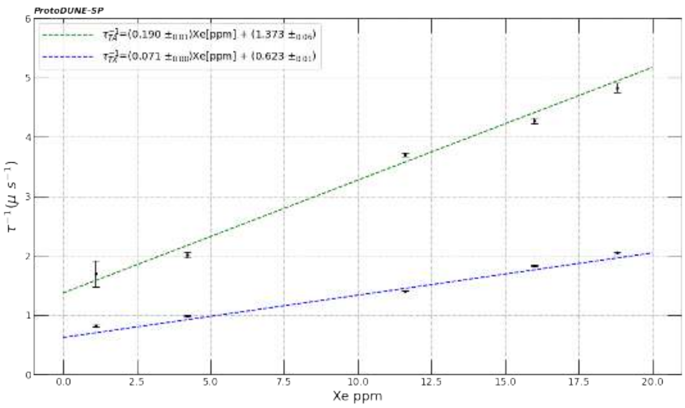
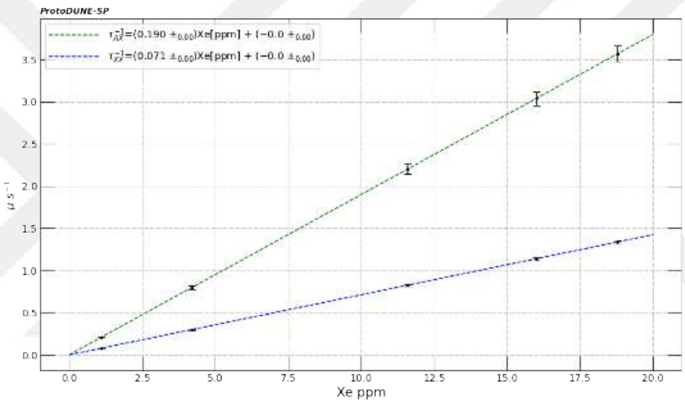



Figure 5.9. These plots depict the relationship between the pulse charge (Q), which corresponds to the charge collected in the quartz readout channel, and the ratio of charge collected in the narrow and wide readout channels (NQ), which corresponds to the position of the ionization in the detector, for five different Xe injection concentrations. Simultaneous fits of the signals collected with and without a quartz window provide valuable insights into the light collection efficiency and aid in characterizing the detector's performance at varying Xe concentrations. Additionally, the time constants (taus) for each injection concentration are provided in Table 5.1. for further analysis.




(a) Overall scintillation time constant for argon and xenon.




(b) Interaction time constant for argon-xenon and xenon-xenon interactions.

Figure 5.10. Time constants for scintillation and interaction processes in argon and xenon. a shows the overall scintillation time constant for argon and xenon, which is the decay time of the scintillation light after an event. b shows the interaction time constant for argon-xenon and xenon-xenon interactions, which is the characteristic time for the transfer of energy between the two particles involved in the interaction. These time constants are important for understanding the properties of liquid argon and xenon used in dark matter detectors.

# 6. CONCLUSION

In this study, we have explored various dark matter and dark Higgs related models and obtained exclusion limits for the $Z'$-2HDM and Baryonic $Z'$ models as a function of $m_{Z'}$ and $m_{A^0}$ or $m_\chi$, respectively. Our analysis included assumptions such as $tan\beta = 1.0$, $g_{Z'} = 0.8$, and $m_\chi = 100$ GeV for the $Z'$-2HDM, and $g_q = 0.25, g_\chi = 1$ for the Baryonic $Z'$ model.

Our results show that for the $Z'$-2HDM model, we were able to exclude a $Z'$ mass of 2.3 TeV with an expected exclusion of 2.55 TeV for a mass $m_{A^0} = 300\,\text{GeV}$. The lower exclusion boundary was given by $m_{Z'} = 600\,\text{GeV}$. The largest $A^0$ mass we can exclude is 700 GeV for a corresponding Z' mass of 1.75 TeV, with an expected exclusion of $M_{A^0} = 800\,\text{GeV}$. These are the most stringent limits put on this model to this date. For the $Z'$-Baryonic model, we can exclude masses below $m_{Z'} = 1.45\,\text{TeV}$, with an expected exclusion of 1.75 TeV. This is the first time that limits on this model have been set in the mono-H(bb) channel.

Furthermore, we also studied the i2HDM model and obtained exclusion limits for the $h_1h_1$ and $h_1h_2$ processes. We observed various mono-jet signals at the LHC, including the particle process $pp \rightarrow h_1h_1j$ and $q\bar{q} \rightarrow h_1h_2g$ ($gq \rightarrow h_1h_2g$), which were used to create samples of $h_1h_1$ and $h_1h_2$ events. The exclusion limits for the $h_1h_1$ and $h_1h_2$ processes were obtained using the HiggsCombine tool within the CMSSW software package, based on the asymptotic limits approach. Our study found that the $h_1h_2$ process is more effective in detecting signals in the mass range studied, with no significant drop in sensitivity.

Moreover, we investigated the use of xenon doping in liquid argon to enhance light collection efficiency for TPCs, a technique that will be used for the DUNE Far Detector. We tested this technique in the ProtoDUNE-SP experiment, when the cryostat was contaminated with a few ppm of nitrogen levels, which is known reduce the amount of scintillation light. Our study investigated the combination of nitrogen and xenon in the liquid argon using an X-ARAPUCA detector, and a three-exponential decay model was used to demonstrate the light transfer process between argon, xenon, and nitrogen. Based on our model fit, we were able to show successful results.

Finally, our study provides valuable insights into the exclusion limits of various dark matter and dark Higgs related models, as well as a detailed evaluation of the timing mechanism of the scintillation process of argon in the presence of nitrogen and xenon. By using an X-ARAPUCA detector and a three-exponential decay model, we demonstrated the successful enhancement of light collection efficiency for TPCs through the use of xenon doping in liquid argon. These findings can inform future research in the field of particle physics and contribute to the search for a better understanding of the nature of dark matter.

## REFERENCES


Abercrombie, Daniel et al. (2015). “Dark matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum”.

Abi, B. et al. (2018). “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: Single-Phase Module”. In: *arXiv:1807.10334 [physics.ins-det]*.

— (2020a). “First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform”. In: *JINST* 15, P12004. DOI: 10.1088/1748-0221/15/12/P12004. arXiv: 2007.06722 [physics.ins-det].

— (2020b). “The DUNE Far Detector Interim Design Report Volume 2: Single-Phase Module”. In: *arXiv:2010.02939 [physics.ins-det]*.

— (2020c). “The DUNE Far Detector Interim Design Report, Volume 1: Physics, Technology and Strategies”. In: *arXiv preprint arXiv:2002.03005*.

Abi, Babak et al. (2020a). “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE”. In: *JINST* 15.08, T08008. DOI: 10.1088/1748-0221/15/08/T08008. arXiv: 2002.02967 [physics.ins-det].

— (Feb. 2020b). “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics”. In: arXiv: 2002.03005 [hep-ex].

Abud, A. Abed et al. (2022). “Design, construction and operation of the ProtoDUNE-SP Liquid Argon TPC”. In: *JINST* 17.01, P01005. DOI: 10.1088/1748-0221/17/01/P01005. arXiv: 2108.01902 [physics.ins-det].

Acciarri, R. et al. (2015). "A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam". In: *arXiv e-prints*. arXiv: 1503.01520 [hep-ex].

— (2017). "Design and Construction of the MicroBooNE Detector". In: *JINST* 12.02, P02017. DOI: 10.1088/1748-0221/12/02/P02017. arXiv: 1612.05824 [physics.ins-det].

Acciarri, R. et al. (2008). "Cryogenic Distillation of Nitrogen from Liquid Argon". In: *arXiv:0804.1217*. arXiv: 0804.1217.

Akchurin, N. et al. (2016). *Search for dark matter produced with an energetic jet, or a hadronically decaying W or Z boson, at* $\sqrt{s} = 13$ *TeV with the full 2016 dataset*. CMS Note 2016/473. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2016/473.

Alwall, J., R. Frederix, et al. (2014). "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07, p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

Alwall, J., S. Höche, et al. (Feb. 2008). "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *The European Physical Journal C* 53.3, pp. 473–500. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-007-0490-5. URL: https://doi.org/10.1140/epjc/s10052-007-0490-5.

Belyaev, Alexander et al. (Feb. 2018). "Anatomy of the inert two-Higgs-doublet model in the light of the LHC and non-LHC dark matter searches". In: *Physical Review D* 97.3. DOI: 10.1103/physrevd.97.035011. URL: https://arxiv.org/pdf/1606.07609.pdf.

Cacciari, Matteo, Gavin P. Salam, and Gregory Soyez (2008). “The anti-$k_t$ jet clustering algorithm”. In: *JHEP* 04, p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

Catani, S. et al. (1993). “Longitudinally-invariant kt-clustering algorithms for hadron-hadron collisions”. In: *Nuclear Physics B* 406.1, pp. 187–224. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(93)90166-M. URL: http://www.sciencedirect.com/science/article/pii/055032139390166M.

CMS Collaboration (2012a). *Differential cross section measurements for the production of a W boson in association with jets in proton-proton collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CMS-PAS-SMP-12-023. URL: https://cds.cern.ch/record/1712497.

— (2012b). *Measurement of the production cross section for a W boson and two b jets in pp collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CMS-PAS-SMP-12-026. URL: https://cds.cern.ch/record/1639945.

— (2013). *Measurement of associated W + charm production in pp collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CMS-PAS-SMP-13-149. URL: https://cds.cern.ch/record/1712497.

— (2014a). *Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CERN-PH-EP-2014-005. URL: https://cds.cern.ch/record/1646977.

— (2014b). *Measurements of jet multiplicity and differential production cross sections of Z+jets events in proton-proton collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CERN-PH-EP-2014-205. URL: https://cds.cern.ch/record/1749688.

CMS Collaboration (2015). "Measurements of the b-tagging performance of the CMS detector". In: *CMS Physics Analysis Summary* BTV-15-001. URL: https://cds.cern.ch/record/1984285/files/BTV-15-001-pas.pdf.

— (2018). "Multivariate Electron Identification for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2. [Online; accessed March 2023].

CMS EGamma POG (2015). https://twiki.cern.ch/twiki/bin/view/CMS/MultivariatePhotonIdentificationRun2. Accessed: March 3, 2023.

CMS Jet/BTV POG (n.d.). https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X.

CMS JetMET POG (2017). "CMS Jet Identification". https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID. Accessed: March 3, 2023.

— (2018). "MissingET Optional Filters for Run 2". https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2. Accessed on 2023-03-05.

— (2022). https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataMC. Accessed on March 3, 2023.

— (n.d.). https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution.

CMS Tau POG (2016). https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendation13TeV. Accessed: March 3, 2023.

Collaboration, CMS (2009). *Missing transverse energy performance in minimum-bias events using proton-proton collisions at* $\sqrt{s} = 900$ *GeV*. Tech. rep. CMS-PAS-JME-09-001. CERN. URL: https://cds.cern.ch/record/1194489/files/JME-09-001-pas.pdf.

Dolen, James et al. (2016). “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”. In: *JHEP* 05, p. 156. DOI: 10.1007/JHEP05(2016)156. arXiv: 1603.00027 [hep-ph].

Duerr, Michael, Alexander Grohsjean, et al. (2017). “Hunting the dark Higgs”. In: *Journal of High Energy Physics* 2017.4, p. 143. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)143. URL: https://doi.org/10.1007/JHEP04(2017)143.

Duerr, Michael, Felix Kahlhoefer, et al. (2016). “How to save the WIMP: global analysis of a dark matter model with two s-channel mediators”. In: *JHEP* 09, p. 042. DOI: 10.1007/JHEP09(2016)042. arXiv: 1606.07609 [hep-ph].

“DUNE illustration” (2023). https://www.dunescience.org/. [Accessed February 21, 2023].

“DUNE.org” (2023). https://www.dunescience.org/about-dune/. [Accessed February 21, 2023].

Ekstedt, Andreas et al. (2016). “Constraining minimal anomaly free U(1) extensions of the Standard Model”. In: *JHEP* 11, p. 071. DOI: 10.1007/JHEP11(2016)071. arXiv: 1605.04855 [hep-ph].

“fnal.gov” (2022). https://www.fnal.gov/science/experiments/dune.html. [Accessed February 21, 2023].

Frandsen, Mads T. et al. (2012). “LHC and Tevatron Bounds on the Dark Matter Direct Detection Cross-Section for Vector Mediators”. In: *JHEP* 07, p. 123. DOI: 10.1007/JHEP07(2012)123. arXiv: 1204.3839 [hep-ph].

Higgs PAG (2019). “SWGuideHiggsAnalysisCombinedLimit”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit. Accessed on 2023-03-05.

Hoeche, Stefan et al. (2005). “Matching parton showers and matrix elements”. In: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, pp. 288–289. DOI: 10.5170/CERN-2005-014.288. arXiv: hep-ph/0602031 [hep-ph].

Kahlhoefer, Felix et al. (2016a). “Implications of unitarity and gauge invariance for simplified dark matter models”. In: *JHEP* 02, p. 016. DOI: 10.1007/JHEP02(2016)016. arXiv: 1510.02110 [hep-ph].

— (Feb. 2016b). “Implications of unitarity and gauge invariance for simplified dark matter models”. In: *Journal of High Energy Physics* 2016.2, p. 16. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2016)016. URL: https://doi.org/10.1007/JHEP02(2016)016.

Karimäki, V et al. (1997). *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN. URL: http://cds.cern.ch/record/368412.

Larkoski, A. et al. (2014). “Soft drop”. In: *JHEP* 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

Larkoski, Andrew J. et al. (2014). “Soft drop”. In: *JHEP* 04, p. 017. DOI: 10.1007/JHEP04(2014)017. arXiv: 1402.2657 [hep-ph].

Machado, P. A. N. (2016). “ARAPUCA: a new concept for detection of scintillation light”. In: *Journal of Instrumentation* 11.08, P08011. DOI: 10.1088/1748-0221/11/08/P08011.

— (2018). “X-ARAPUCA: A highly efficient light trap for the next generation of liquid argon neutrino detectors”. In: *Journal of Instrumentation* 13.06, P06015. DOI: 10.1088/1748-0221/13/06/P06015.

McFadden, N. et al. (Feb. 2021). “Large-scale, precision xenon doping of liquid argon”. In: *Nuclear Instruments and Methods in Physics Research Section*

*A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1011, p. 165575. DOI: 10.1016/j.nima.2021.165575. URL: https://arxiv.org/abs/2006.09780.

*Pileup Removal Algorithms* (2014). Tech. rep. Geneva: CERN. URL: https://cds.cern.ch/record/1751454.

POG, Muon (2023). *Muon POG recommended ID*. Accessed on March 1, 2023. URL: https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Medium_Muon.

Rubbia, C. et al. (1977). "Liquid Argon as a Tracking Material: A New Concept in Neutrino Detectors". In: *CERN-EP-77-08*.

Sjöstrand, Torbjörn et al. (2015). "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191, pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: http://www.sciencedirect.com/science/article/pii/S0010465515000442.

Smith et al. (2005). "*Particle dark matter: evidence, candidates and constraints*". In: *Phys. Rept.* 405, p. 279. DOI: 10.1016/j.physrep.2004.08.031. arXiv: hep-ph/0404175 [hep-ph].

Smith, Nicholas et al. (2020). "Coffea Columnar Object Framework For Effective Analysis". In: *EPJ Web of Conferences* 245. Ed. by C. Doglioni et al., p. 06012. DOI: 10.1051/epjconf/202024506012. URL: https://arxiv.org/abs/2008.12712.

Strologas, John (2016). "Performance of Jet reconstruction in CMS at 13 TeV". In: *PoS* ICHEP2016, p. 736. DOI: 10.22323/1.282.0736. URL: http://cds.cern.ch/record/2287326.

The NNPDF collaboration et al. (2015). "Parton distributions for the LHC run II". In: *Journal of High Energy Physics* 2015.4, p. 40. ISSN: 1029-8479. DOI:

10.1007/JHEP04(2015)040. URL: https://doi.org/10.1007/JHEP04(2015)040.

TRootABCD TWIKI (2023). “ABCD Method”. https://twiki.cern.ch/twiki/bin/view/Main/ABCDMethod. Accessed on 2023-03-04.

**CURRICULUM VITAE**

He studied Physics at Cukurova University and graduated in 2012. For his MSc, he contributed Gamma-Ray Burst analysis observed by SWIFT to evaluate the burst's temporal properties. In addition, he contributed to the optical telescope facility and performed observations at the UZAYMER observatory. After completing his MSc in 2015, he began his PhD with the beyond standard model physics group in 2016 at CMS. He contributed to several studies such as Mono-Jet, i2HDM and Dark Higgs. In 2019 he became a member of DUNE and ProtoDUNE experiments and contributed to photon detection and time projection chambers.