BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

A SEARCH FOR GEV SCALE DARK MEDIATORS DECAYING TO A PAIR OF MUONS USING DATA FROM THE CMS DETECTOR

by

CHRISTOPHER COSBY

B.S., University of Washington, 2018
M.S., Boston University, 2020

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2024

Approved by

First Reader

James W. Rohlf, PhD.
Professor of Physics

Second Reader

David M. Sperka, PhD.
Assistant Professor of Physics

To my loving parents.

Acknowledgments

Many people have played a role in helping to create this thesis, both from my academic circles and my personal life, in ways both large and small. I will not be able to adiquately thank all of them here, but will try to acknowledge some of the people who had the most direct impact on this work and my life in the past few years.

First, I am grateful to the members of my thesis committee; Anushya Chandran, Ed Kearns, Pankaj Mehta, and Martin Schmaltz, for their insights and good questions during my defense, and their feedback on the draft of this thesis.

I wish to thank all of my colleagues, friends, and mentors in the BU CMS group. It has been a privilege to work with people who are so warm, welcoming, and deeply interested in helping one another succeed. My fellow graduate students are always a joy to see around the office. In no particular order, Alp Akpinar, Kaitlin Salyer, Siqi Yuan, Dan Arcaro, and Anna Tsatsos, have all played a big part in shaping my experience as a graduate student. Andreas Albert, Sihyun Jeon, Daniel Spitzbart, Celia Fernandez Madrazo, Carlos Erice Cid, and Elisa Fontanessi are all role models of mine, and I will strive to emulate their expertise and passion in my own post-doc position. Though we did not work together directly, I have benefited greatly from the mentorship and insights of all BU CMS faculty; my thanks to Indara Suarez, Frank Golf, and Zeynep Demiragli.

For their contributions to the analysis that is the centerpiece of this thesis, I also wish to thank CMS collaborators Zhangqier Wang, and Christoph Paus from MIT.

My thesis advisors, David Sperka and Jim Rohlf, have had by far the greatest impact on the physicist I am today. David has been an unyielding source of knowledge and support throughout our time working together, and he has set an incredible standard for how faculty should nurture and care for their graduate students. Likewise, Jim possesses a wealth of wisdom and insight about our field, which he shares freely.

I always enjoy our conversations and I always learn something new from them. I am truly grateful to both of them for the opportunities they have given me, and I hope to make them proud through my future work.

Larry Sulak has also played a large role in helping me to suceed as a physicist. By working as a teaching fellow in his Advanced Lab and Electroncis Lab courses, I gained invaluable experience in basic instrumentation and how to practically engage with hardware problems. He has sponsored me on multiple trips to CERN to aid in the development of the NIM+ project, and helped me to build connections that continue to bear fruit years later.

I have learned much from the staff engineers in the BU EDF. Eric Hazen and Dan Gastler were both fantastic resources for me as I grew into my role as the graduate student AMC13 contact. Dan in particular has imparted many coding best practices onto me that I now compulisvely adhere to, to the betterment of my code and my life generally. Also, Chris Lawlor has always been friendly and dependable, whenever I need a tool, component, or advice on how best to approach a problem.

My thanks to all the front office administrative staff, in particular the graduate program coordinators; Mirtha Cabello, Vikram Chodary, and Christina Crabtree. Mirtha was the first person to greet me when I arrived in Boston, and I relied greatly on her deep knowledge of the physics program in my first couple of years.

I have been blessed with many good times and close friendships during the course of my Ph.D. Amoung those not yet named are David Long, Brian Dawes, Ronnie Tang, Maria Yampolskaya, Nick Russo, Alex Kapadia, Victoria Kovalchuk, Zijie Wan, Matteo Bellitti, and many more. Together we have hosted a thriving book club, multiple tabletop groups, and lively beer socials.

I must give a special acknowledgement to two friends with whom I have shared many treasured memories over the past six years; Thomas Wester and Paige Kunkle.

With them, grueling problem sets became opportunities for joyous collaboration, and many otherwise boring afternoons were filled with warm coffee and delightful discussion. They are in a Zoom call with me as I write these acknowledgements, supporting me even in the final minutes before I submit this dissertation.

Disruptions and setbacks are an expected part of any Ph.D. program. Even so, a cancer diagnosis is a serious derailment. This work has been made possible by the fact that I am still alive, thanks to the efforts of my cancer care team at Massachusetts General Hospital; Dr. Jeremy Abramson, Dr. Julie Haydu, Rosalba Martignetti, and my infusion nurse, Greg Smith. Dr. Rich Galgano and Sheila Hunkeler from BU Student Health were critical in getting my initial diagnosis. Sheila in particular worked with me for many months to understand why I felt sick, and always took my reported symptoms seriously, even when they seemed strange and uncorrelated.

Finally, I must thank my friends and family in Washington state for their love and support over the past six years. This section will continue for 50 more pages if I attempt to recognize all my beloved friends from Woodnville and my time at UW, but I must give special credit to Andrew Leever, Sidd Gorti, and Grant Sportelli, who are indispensable companions that I hope to keep in my life forever. Joe Hutchings, who has played a major role in my personal development and who I always look forward to seeing. My siblings, Claire, Nick, and Amelia, who I love and am very proud of. My parents, who have always supported me and elevated me to achieve my dreams.

A SEARCH FOR GEV SCALE DARK MEDIATORS DECAYING TO A PAIR OF MUONS USING DATA FROM THE CMS DETECTOR

CHRISTOPHER COSBY

Boston University, Graduate School of Arts and Sciences, 2024

Major Professors: James W. Rohlf, PhD.
Professor of Physics

David M. Sperka, PhD.
Assistant Professor of Physics

ABSTRACT

Data from the LHC has enabled groundbreaking searches for physics beyond the standard model to be performed over the past decade. However, given that no concrete signatures of new physics have yet been observed at the TeV scale, many physicists have begun to explore the use of LHC data to probe lower energy processes with exceptional precision. This thesis will describe a search for prompt low-mass dimuon resonances performed using the 13 TeV proton-proton collision data collected by the Compact Muon Solenoid (CMS) detector during LHC Run-II. The analysis exploits a dedicated high-rate trigger stream which allows the recording of events below conventional momentum thresholds. The search is performed by looking for narrow resonance peaks in the dimuon mass continuum in the ranges from 1.1–2.6 and from 4.2–7.9 GeV. Model-independent limits on the production rate of a dimuon resonance within the experimental fiducial acceptance are set. Limits on parameters for a dark photon model and a two Higgs doublet model with an extra scalar, are also discussed.

Contents

List of Tables

List of Figures

List of Abbreviations

2HDM .... Two Higgs doublet model
2HDM+S .... Two Higgs doublet plus scalar model
APD .... Avalanche photodiode
BDT .... Boosted decision tree
BPIX .... Barrel pixel detector
BSM .... Beyond the Standard Model
CL .... Confidence-level
CM .... Center of mass
CMS .... The Compact Muon Solenoid
CR .... Control region
CSC .... Cathode strip chamber
dCB .... Double-sided Crystal Ball
DT .... Drift tube
DY .... Drell-Yan
ECAL .... Electromagnetic Calorimeter
EM .... Electromagnetic
FCNC .... Flavor-changing neutral current
FPIX .... Forward pixel detector
GEM .... Gas electron multiplier
ggF .... Gluon-gluon fusion
HAHM .... Hidden Abelian Higgs model
HB .... HCAL Barrel
HCAL .... Hadronic Calorimeter
HE .... HCAL Endcap
HF .... HCAL Forward
HLT .... High-level trigger

HO .... HCAL Outer
HV .... High-voltage
ISR .... Initial state radiation
ISR .... Intersecting Storage Rings
IT .... Inner Tracker
L1 .... Level-1
LHC .... Large Hadron Collider
LO .... Leading order
MC .... Monte-Carlo
ML .... Maximum-likelihood
NLO .... Next-to-leading order
NNLO .... Next-to-next-to-leading order
OS .... Opposite sign
pdf .... Parton-distribution function
PV .... Primary vertex
QCD .... Quantum chromodynamics
QED .... Quantum electroodynamics
RPC .... Resistive plate chamber
Sp$\bar{\text{p}}$S .... Super Proton-Antiproton Synchrotron
SiPM .... Silicon-photomultiplier
SM .... Standard Model
SPS .... Super Proton Synchrotron
SR .... Signal region
SS .... Same sign
TEC .... Tracker endcaps
TIB .... Tracker inner barrel
TID .... Tracker inner disks
TnP .... Tag-and-probe
TOB .... Tracker outer barrel
TTC .... Trigger, timing, and control
VPT .... Vacuum phototriode

Chapter 1

Introduction

1.1 Particle Physics

Almost all known phenomenon in the universe can be modelled and understood using the language of particle physics. Everything, from the tumultuous dynamics of the early universe, to the nuclear processes that ignite our sun, to the simple interactions your body has with matter here on Earth can be decomposed and understood in terms of interactions between fundamental particles and fields that constitute them.

1.1.1 The Standard Model

The Standard Model (SM) is the state-of-the-art in particle physics for describing the behavior and dynamics of matter in our universe. It contains the properties of all known particles and makes predictions for how they should interact with one another using the computational methods of quantum field theory. The SM contains 12 species of fermions, which can coarsely be described as "matter," 5 unique vector bosons which correspond to "forces" or interactions between fields, and 1 scalar boson.

Fermions

The observable matter of our universe is composed of spin-1/2 fermions, differentiated by their different combinations of charges, masses, and other quantum numbers. The usual visualizations of the SM particles will contain the fermions grouped into quarks and leptons, arranged into three generations that correspond to the masses of the particles and approximately the order in which they were discovered.

Symmetries and Gauge Invariance

Our derivation of the SM will begin with the study of free Dirac fermions, as we observe them in our universe today. The force carriers of the SM are gauge bosons which appear in the theory upon requiring invariance in the action of these fermions under local transformations of the groups SU(3) and SU(2) $\otimes$ U(1).

Generically, local transformations ($U(x)$) derived from the generators of these groups will not conserve the dynamics of the action of free fermions they are applied to. A Lagrangian description of a physical system will contain partial derivative operators (∂_μ), which produce additional terms when they act on the spatially dependant part of the locally transformed field. This will alter the action of the system, and our theory is not invariant under the said operation.

If one believes that the universe does in fact respect the symmetries listed above (something which can be inferred by studying the behavior of particles charged under those groups), then invariance under local transformations can be maintained by reformulating the Lagrangian in terms of a gauge covariant derivative (D_μ). For transformations of the form $\psi \to e^{-g\alpha}\psi$, the gauge covariant derivative has the form:

$$D_\mu = \partial_\mu - igA_\mu, \tag{1.1}$$

which introduces a new gauge field A. In the case of the SM symmetries, these gauge fields correspond to the force carriers of the associated interaction. By Noether's theorem, each of these symmetries has an associated charge, which is conserved in all processes mediated by these fields.

1.1.2 The Electroweak Force

The force carriers of the electroweak sector are the massless photon and the three massive weak bosons Z^0, W^+, and W^-, which are physical states of the gauge fields induced by the $\mathrm{SU}(2)\otimes\mathrm{U}(1)$ group generators. The photon mediates electromagnetic interactions, and is ubiquitous at any energy scale, facilitating a wide range of familiar physical processes, such as binding electrons and nuclei into atoms. By contrast, the weak mediators are more esoteric. At the energy scales we experience on Earth they are unlikely to be produced outside of nuclear decays, neutrino interactions, and collider or reactor settings.

Gauge Structure

For most people, field theory is introduced in the context of scalar electromagnetism. It is shown to emerge from the symmetry $\mathrm{U}(1)_\gamma$, which conserves electric charge, and the associated gauge field possesses all the properties of the Standard Model photon. Because of this, it is seductive to interpret the $\mathrm{U}(1)$ component of the electroweak group in a similar way, such that $\mathrm{U}(1)$ symmetry singlehandedly generates the photon field, and the three $\mathrm{SU}(2)$ generators produce the three weak bosons. In reality, the physical bosons we observe are superpositions of the underlying fields produced by the $\mathrm{SU}(2)\otimes\mathrm{U}(1)$ generators.

The construction begins by understanding the handedness and flavor structure of the fields that the $\mathrm{SU}(2)$ generators act on [1]. Fermion fields can be decomposed into left and right handed chiral components. These components are defined as the remaining parts of the fields when they are subjected to the chiral projection operator:

$$P_{R,L} = \frac{1 \pm \gamma^5}{2}, \tag{1.2}$$

where $P_{R,L}\psi = \psi_{R,L}$ and $\bar{\psi}P_{R,L} = \bar{\psi}_{L,R}$. For massless fermions, the chiral states are equivalent to helicity. The exact form of γ^5 depends on the choice of basis, but it always acts on a 4-component dual spinor description of the field. After this separation is performed, a difference in how the fields transform under $SU(2)$ operations emerges.

The left handed components of the fermion fields transform as doublets of flavor under SU(2) ⊗ U(1), meaning this component of the fields is effected by both facets of the transformations:

$$\psi_L = \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L \qquad \text{OR} \qquad \psi_L = \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L \tag{1.3}$$

In regards to how they transform under these groups, there is parity between the quark and lepton doublets. A global transformation of such a field under SU(2)⊗U(1) would look like:

$$\psi_L \to \psi_L' = e^{i\left(Y\beta + \sum_{i=1}^{3} \frac{\alpha_i \sigma_i}{2}\right)} \psi_L \tag{1.4}$$

where $\sigma_{1,2,3}$ are the 3 Pauli matrices, and β is some rotation value, and Y is the hypercharge of the field the transformation is applied to.

For the right handed components, the fields transform as singlets and are unaffected by the SU(2) operations. They have the form:

$$\psi_{R,1} = q_{u,R}, \quad \psi_{R,2} = q_{d,R} \qquad \text{OR} \qquad \psi_{R,1} = \nu_{\ell,R}, \quad \psi_{R,2} = \ell_R^- \tag{1.5}$$

A global transformation of these fields would resemble:

$$\psi(x)_R \to \psi(x)_R' = e^{iY\beta}\psi(x)_R \tag{1.6}$$

The free Lagrangian for a Dirac spinor is:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.7}$$

When decomposed into the left and right handed components of ψ, it becomes:

$$\mathcal{L} = ((P_L + P_R)\psi)^\dagger\, \gamma^0 (i\gamma^\mu \partial_\mu - m)(P_L + P_R)\psi$$

(using: $\bar{\psi} = \psi^\dagger \gamma^0$)

$$= \psi^\dagger (P_L + P_R)^\dagger \gamma^0 (i\gamma^\mu \partial_\mu - m)(P_L + P_R)\psi$$

$$= \bar{\psi}\gamma^0 (P_L + P_R)^\dagger \gamma^0 (i\gamma^\mu \partial_\mu - m)(P_L + P_R)\psi$$

(exploiting the identity: $\gamma^0(\gamma^5)^\dagger\gamma^0 = -\gamma^5$)

$$= \bar{\psi}(P_R + P_L)(i\gamma^\mu \partial_\mu - m)(P_L + P_R)\psi$$

$$= \bar{\psi}(P_R + P_L) i\gamma^\mu \partial_\mu (P_L + P_R)\psi - m\bar{\psi}(P_R + P_L)(P_L + P_R)\psi$$

(separating mass and kinetic terms)

$$= \bar{\psi} i\gamma^\mu \partial_\mu (P_L + P_R)(P_L + P_R)\psi - m\bar{\psi}(P_R + P_L)(P_L + P_R)\psi$$

(exploiting the identity: $\gamma^5\gamma^\mu = -\gamma^\mu\gamma^5$)

$$= \bar{\psi} i\gamma^\mu \partial_\mu (P_L P_L + P_R P_R)\psi - m\bar{\psi}(P_L P_L + P_R P_R)\psi$$

(removing terms with $P_R P_L = P_L P_R = 0$)

$$= \bar{\psi} P_R i\gamma^\mu \partial_\mu P_L \psi + \bar{\psi} P_L i\gamma^\mu \partial_\mu P_R)\psi - m\bar{\psi}(P_L P_L + P_R P_R)\psi$$

(returning leftmost operators to left side of γ^μ)

$$= \bar{\psi}_L i\gamma^\mu \partial_\mu \psi_L + \bar{\psi}_R i\gamma^\mu \partial_\mu \psi_R - m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)$$

Consider the effect of applying the global transformations shown in Equations 1.4 and 1.6 to the left and right handed fields, respectively, in this Lagrangian. The kinetic terms will remain invariant, as both transformations are unitary and will cancel with their own Hermitian conjugates. However, the transformations will not cancel in the mass terms, violating the symmetry we seek to impose. To restore invariance under SU(2) $\otimes$ U(1), we declare that the SM fermions are all massless ($m = 0$, at least under the Dirac description of mass), removing the transformed terms. The fermion Lagrangian for our electroweak theory is now:

$$\mathcal{L} = \bar{\psi}_L i\gamma^\mu \partial_\mu \psi_L + \sum_{i=1}^{2} \bar{\psi}_{R,i} i\gamma^\mu \partial_\mu \psi_{R,i}, \tag{1.8}$$

where each term corresponds to either the quark or lepton doublet/singlets shown in 1.3 and 1.5. The true "full-glory" form of the Lagrangian will have terms for both.

Now we expand our symmetry constraints to also require invariance under local transformations. To maintain gauge invariance, we introduce the following covariant derivatives to the Lagrangian:

$$\partial_\mu \to D_\mu = \partial_\mu - ig \sum_{a=1}^{3} \frac{\sigma_a}{2} W_\mu^a - ig' Y B_\mu, \tag{1.9}$$

for doublet terms, and:

$$\partial_\mu \to D_\mu = \partial_\mu - ig' Y B_\mu, \tag{1.10}$$

for singlet terms. These invoke four new gauge fields, labeled as $W_\mu^{a=1,2,3}$ for the SU(2) generators and B_μ for U(1), where g and g' are dimensionless gauge coupling parameters. Including kinetic terms for each of these fields requires defining a set of strength tensors with the form $F^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c$, where a, b, and c run over the indices of the gauge group. f^{abc} is the structure constant of the group in question, which encodes the commutation relationship of the different generators. For $B_{\mu\nu}$, the generators of U(1) are just numbers, which trivially commute with themselves [2]. The field strength is therefore:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.11}$$

which is predictably similar to the electromagnetic field strength tensor. The structure constant for the SU(2) group is the three-dimensional Levi-Civita tensor:

$$W^a_{\mu\nu} = \left(\partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu\right). \tag{1.12}$$

The final term in this expression will lead to self-interactions between the W^a_μ-fields upon expanding the kinetic term. The kinetic portion of the Lagrangian for these fields is:

$$\mathcal{L}_{\text{Kinetic}} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a, \tag{1.13}$$

The full Lagrangian with the covariant derivatives substituted in is:

$$\mathcal{L} = \mathcal{L}_{\text{Kinetic}} + \bar{\psi}_L i\gamma^\mu \left(\partial_\mu - ig\frac{\sigma_a}{2} W^a_\mu - ig'YB_\mu\right)\psi_L + \sum_{i=1}^{2} \bar{\psi}_{R,i} i\gamma^\mu \left(\partial_\mu - ig'YB_\mu\right)\psi_{R,i} \tag{1.14}$$

Focusing only on the term which contains the SU(2) fields, we begin to see the sorts of interactions that are enabled by the addition of these fields:

$$\begin{aligned} \mathcal{L} &\subset \bar{\psi}_L \gamma^\mu g \frac{\sigma_a}{2} W^a_\mu \psi_L \\ &= \begin{pmatrix} \bar{q_u} & \bar{q_d} \end{pmatrix} \gamma^\mu g \frac{\sigma_a}{2} W^a_\mu P_L \begin{pmatrix} q_u \\ q_d \end{pmatrix} + \begin{pmatrix} \bar{\nu_\ell} & \bar{\ell^-} \end{pmatrix} \gamma^\mu g \frac{\sigma_a}{2} W^a_\mu P_L \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix} \end{aligned} \tag{1.15}$$

The factor $\frac{\sigma_a}{2} W^a_\mu$ represents the sum of all the Pauli matrices, multiplied by their respective gauge fields:

$$\frac{\sigma_a}{2} W^a_\mu = \frac{1}{2} \begin{pmatrix} W^3_\mu & W^1_\mu + iW^2_\mu \\ W^1_\mu - iW^2_\mu & -W^3_\mu \end{pmatrix} \tag{1.16}$$

The off-diagonal elements of this matrix contract the opposite flavor members of the SU(2) doublets together, creating the "charged current" component of the Lagrangian. Let $W^+_\mu = \frac{W^1_\mu + iW^2_\mu}{2}$ and $W^-_\mu = (W^+_\mu)^\dagger$:

$$\mathcal{L}_{CC} = g\left(\bar{q_u} W^-_\mu \gamma^\mu P_L q_d + \bar{q_d} W^+_\mu \gamma^\mu P_L q_u + \bar{\nu_\ell} W^-_\mu \gamma^\mu P_L \ell^- + \bar{\ell^-} W^+_\mu \gamma^\mu P_L \nu_\ell\right). \tag{1.17}$$

Each of these terms represents a real vertex in the SM, where a quark or lepton can be converted into its isospin doublet partner, via contact with these mixtures of the SU(2) gauge fields. We can now identify these as the physical SM $W^{\pm}$ bosons.

We now address the on-diagonal elements of $\frac{\sigma_a}{2}W_\mu^a$ and the B_μ terms. Our goal is to identify the SM neutral bosons and hopefully recover the theory of quantum electroodynamics (QED), with a photon field the couples to fermions proportional to their electric charge (Q). If we suppose that W_μ^3/B_μ are purely the Z^0 and photon respectively, then the B_μ gauge coupling must be the QED fundamental charge ($g' = e$), and the hypercharge parameters (Y) must be proportional to the electric charge of each fermion. However, because B_μ is a singlet under SU(2), it will assign the same hypercharge to both components of the left handed doublets ($Q_u = Q_d$, and $Q_\ell = Q_\nu$). Further, the QED photon is blind to chirality, acting equally on left and right handed fields, so the right handed components of these fermions would have the same hypercharge as their left handed counterparts. This would mean that all quarks would have the same electric charge as one another, as would all leptons. This is incompatible with our observed reality, so we must seek a different definition of the photon and electric charge.

We will instead try to find the physical neutral bosons by formulating a mixture of W_μ^3 and B_μ, similar to how the physical charged current mediators are mixtures of W_μ^1 and W_μ^2. Consider the rotation:

$$\begin{pmatrix} A_\mu \\ Z_\mu^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \tag{1.18}$$

where θ_W is the "weak mixing angle," which will recover QED by satisfying the relationship $g\sin\theta_W = g'\cos\theta_W = e$, with an appropriate definition of electric charge. The neutral sector Lagrangian now has two components, one of which exclusively contains the A_μ field:

$$\begin{aligned}
\mathcal{L}_{\text{QED}} &= \sum_f \bar{\psi}_f \gamma^\mu A_\mu \left(g\frac{\sigma_3}{2}\sin\theta_W + g' Y_f \cos\theta_W \right)\psi_f \\
&= \sum_f \bar{\psi}_f \gamma^\mu e A_\mu \left(\frac{\sigma_3}{2} + Y_f \right)\psi_f, \qquad (1.19)
\end{aligned}$$

where f runs over all fermions. If the component in parenthesis is electric charge, then this term represents interactions with the SM photon. We define the operator $Q = T_3 + Y_f$, where $T_3 = \frac{\sigma_3}{2}$ is the weak-isospin projection of the fermions, which is $\pm 1/2$ for the doublets and 0 for the singlets. We can now determine hypercharge values for each fermion which will correctly assign their electric charges:

$$Y_{L,u} = Y_{L,d} = +1/6 \qquad Y_{L,\nu} = Y_{L,\ell} = -1/2$$
$$Y_{R,u} = +2/3, \quad Y_{R,d} = -1/3 \quad Y_{R,\nu} = 0 \quad Y_{R,\ell} = -1$$

The right-handed neutrino has no charge under any of the generators, and therefore has no couplings to any of the electroweak bosons.

The other term in the neutral current Lagrangian is:

$$\begin{aligned}
\mathcal{L}_Z &= \sum_f \bar{\psi}_f \gamma^\mu Z_\mu \left(g\frac{\sigma_3}{2}\cos\theta_W + g' Y_f \sin\theta_W \right)\psi_f \\
&= \sum_f \bar{\psi}_f \gamma^\mu Z_\mu \left(eT_3 \frac{\cos\theta_W}{\sin\theta_W} + Y_f \frac{\sin\theta_W}{\cos\theta_W} \right)\psi_f \\
&= \sum_f \bar{\psi}_f \gamma^\mu Z_\mu \frac{e}{\sin\theta_W\cos\theta_W}\left(T_3\cos^2\theta_W + Y_f\sin^2\theta_W \right)\psi_f \\
&= \sum_f \bar{\psi}_f \gamma^\mu Z_\mu \frac{e}{\sin\theta_W\cos\theta_W}\left(T_3 + Q\sin^2\theta_W \right)\psi_f \qquad (1.20)
\end{aligned}$$

This term facilitates interactions between the the Z^0 boson and fermion/anti-fermion pairs of the same species.

The full Lagrangian for the electroweak sector that has been described thus far is:

$$\mathcal{L}_{\mathrm{EWK}} = \mathcal{L}_{\mathrm{Kinetic}} + \mathcal{L}_{CC} + \mathcal{L}_{\mathrm{QED}} + \mathcal{L}_{Z} \tag{1.21}$$

Note that at this stage, all of the fermions are massless, due to our symmetry requirements. This also applies to the gauge bosons, as their mass terms would have the form $m^2 A^\mu A_\mu$, which is not invariant under SU(2)$\otimes$U(1) transformation of the fields. We must invoke an additional field to provide mass terms to all of these particles, which will be described in a later section.

1.1.3 Quantum Chromodynamics

The third symmetry group of the Standard Model is SU(3), which corresponds to the strong force and produces the theory of quantum chromodynamics (QCD). The construction of this sector is in many ways similar to the SU(2) component of the electroweak derivation [3], complicated by the addition of more generators, but made simpler by the lack of mixing with the U(1) gauge field. Where the SU(2) generators acted only on left-handed fermions in doublets of isospin, the SU(3) generators act exclusively on quarks in triplets of "color" (red, blue, green). We will begin again with the Dirac Lagrangian:

$$\mathcal{L} = i \sum_{f}^{\substack{u,c,t\\d,s,b}} \bar{q}_f \gamma^\mu \partial_\mu q_f, \tag{1.22}$$

where f indexes all the flavors of quark. The quarks remain massless due to the SU(2) $\otimes$ U(1) constraints. As stated, q_f is a three component vector which is acted on by the SU(3) generators. These generators are the 8 Gell-Man matrices (λ_a), a set of 3x3 analogues to the Pauli matrices. A global transformation of the quark triplet would look like:

$$q_f \rightarrow q'_f = e^{i \sum\limits_{a=1}^{8} \frac{\theta_a \lambda_a}{2}} q_f, \tag{1.23}$$

which our Lagrangian is invariant under. Promoting to a local transformation ($\theta_a \rightarrow \theta_a(x)$) allows us to construct the covariant derivative:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ig \sum_{a=1}^{8} \frac{\lambda_a}{2} G_\mu^a \tag{1.24}$$

G_μ^{1-8} are the 8 gauge fields, called gluons, which interact with the color triplets like:

$$\mathcal{L}_{\text{QCD}} = \sum_{f}^{\substack{u,c,t\\d,s,b}} \bar{q}_f \gamma^\mu G_\mu^a \frac{\lambda_a}{2} q_f. \tag{1.25}$$

Because all quarks have the same color representation, these fermion interaction terms have significantly less texture than their counterparts in the electroweak Lagrangian.

We also wish to create a kinetic term for the gluon fields, following the same procedure as the previous section. The field strength tensor will be:

$$G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g f^{abc} G_b^\mu G_c^\nu, \tag{1.26}$$

where a, b, and c count over all 8 gluons. There is no single well-known symbol for the QCD structure constant, but it encodes the commutation properties of the generators as:

$$i f^{abc} \lambda^c = \left[\lambda^a, \lambda^b\right] \tag{1.27}$$

The full QCD Lagrangian is:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a + \sum_{f}^{\substack{u,c,t\\d,s,b}} \bar{q}_f \gamma^\mu G_\mu^a \frac{\lambda_a}{2} q_f. \tag{1.28}$$

As with the kinetic term for the SU(2) fields, the gluons will gain self-interaction terms for combinations of generators wherein the structure constant in the strength tensor is not zero. However, because the SU(3) gauge coupling is much larger than the SU(2) coupling, these terms play a much greater role in the physics of QCD:

$$-\frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a = -\frac{1}{4}\left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu + gf^{abc}G_b^\mu G_c^\nu\right)\left(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a + gf^{abc}G_\mu^b G_\nu^c\right) \quad (1.29)$$

The terms that are linear and quadratic in the coupling correspond to three and four gluon vertices respectively:

$$\mathcal{L}_{ggg/gggg} = -2\frac{g}{4}f^{abc}\left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu\right)G_\mu^b G_\nu^c + \frac{g}{4}f^{abc}f_{ade}G_b^\mu G_c^\nu G_\mu^d G_\nu^e \quad (1.30)$$

Gluons are massless due to symmetry constraints, like all other bosons thus far.

1.1.4 The Higgs Mechanism

We have constructed interaction terms for all the expected behaviors of the SM fermions and gauge groups, with one missing component; all the fermions are massless, as are the $W^\pm$ and Z^0 bosons. We have shown that the normal Dirac and Proca mass terms that would be included do not satisfy the symmetry properties we are imposing on our theory. Therefore, we must implement a new method to construct mass terms for these particles. One way to do so is via the Higgs mechanism [4], where the SM includes a complex scalar field (ϕ), with terms in the Lagrangian that are invariant under the generators of the gauge field we want to assign a mass to. As a toy example, we will consider the Abelian case where ϕ transforms under U(1), as though we were trying to give a mass to the photon:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^* D^\mu\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2, \quad (1.31)$$

where μ and λ are unknown parameters. If $\mu^2 > 0$, the potential will have a single minimum at $\phi_0 = 0$, maintaining symmetry under U(1) transformations and creating two massive bosons for the two degrees of freedom in the scalar (the mass term is already in the Lagrangian). Alternatively, if $\mu^2 < 0$, the minimum of the potential moves to $\langle\phi_0\rangle = \sqrt{\frac{-\mu^2}{2\lambda}}$ and the ground state becomes some non-zero value with an infinite set of possible orientations in the complex plane. A general representation of ϕ in polar coordinates looks like:

$$\phi = \frac{v+h}{\sqrt{2}} e^{i\frac{\chi}{v}}, \tag{1.32}$$

where h represents transverse excitations of the field, χ represents rotational excitations, and $v = \frac{-\mu^2}{\lambda}$. Note that the field before had two degrees of freedom for the real and imaginary components of the field, and it still does, though expressed through combinations of h and χ. The above Lagrangian becomes very long when substituting in this field expression, due to many expanding cross-terms. To maintain simplicity, we will focus mainly on the kinetic elements, in particular the effect of the U(1) covariant derivative, when acting on ϕ:

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (\partial_\mu + igA_\mu)\,\phi^*\,(\partial^\mu - igA^\mu)\,\phi + \mu^2\phi^*\phi + ... \\
&= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(e^{-i\frac{\chi}{v}}\partial_\mu h - ie^{-i\frac{\chi}{v}}(h+v)\partial_\mu\frac{\chi}{v} + ie^{-i\frac{\chi}{v}}g(h+v)A_\mu\right) \\
&\qquad\qquad \times\left(e^{i\frac{\chi}{v}}\partial^\mu h + ie^{i\frac{\chi}{v}}(h+v)\partial^\mu\frac{\chi}{v} - ie^{i\frac{\chi}{v}}(h+v)gA^\mu\right) + \mu^2\frac{(v+h)^2}{2} + ... \\
&= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(\partial_\mu h - i(h+v)\left[\partial_\mu\frac{\chi}{v} - gA_\mu\right]\right) \\
&\qquad\qquad \times\left(\partial^\mu h + i(h+v)\left[\partial^\mu\frac{\chi}{v} - gA^\mu\right]\right) + \mu^2\frac{(v+h)^2}{2} + ... \\
&\supset -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(\partial_\mu h\partial^\mu h + (h+v)^2\left(\partial_\mu\frac{\chi}{v} - gA_\mu\right)^2 + \mu^2h^2\right)
\end{aligned} \tag{1.33}$$

The Lagrangian now includes kinetic and mass terms for the transverse mode of the scalar field (a Higgs boson). The parenthesis containing the rotational mode χ and the gauge field A_μ presents an opportunity to greatly simplify this expression and make the physical implications more clear. Because any evolution in the value of χ moves the scalar field along equipotentials, the physics is unchanged by an arbitrary gauge choice of the polar coordinate. We are even free to choose a gauge wherein the term containing χ will perfectly cancel itself out, via:

$$A^\mu \to A^\mu - \frac{1}{gv}\partial_\mu \chi \tag{1.34}$$

The non-interacting sector of the Lagrangian now looks like:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(\partial_\mu h \partial^\mu h + v^2 g^2 A_\mu A^\mu + \mu^2 h^2\right), \tag{1.35}$$

where $v^2 g^2 A_\mu A^\mu$ is a proper mass term for the gauge boson in this theory. It can be said that the gauge boson absorbs or "eats" the degree of freedom associated with χ, in order to access its third polarization state as a, now massive, vector boson. This is a promising start, but practically it is the exact opposite of what we want to do; the SM photon does not need a mass term, and this mechanism has no means by which to give the weak bosons or fermions their own masses. To resolve this, we must extend to the non-Abelian case.

1.1.5 Electroweak Symmetry Breaking

Suppose that the complex scalar field was not a singlet, but a doublet under SU(2):

$$\phi = \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix} \tag{1.36}$$

We wish for it to transform under SU(2)$\otimes$U(1), so we also assign a weak hypercharge value Y. If we choose $Y = +1/2$, then the top component will have an electric charge of +1, and the bottom component will be neutral:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.37}$$

where both entries have a real and imaginary component ($\phi^+ = \frac{1}{\sqrt{2}}[\phi^a + i\phi^b]$, $\phi^0 = \frac{1}{\sqrt{2}}[\phi^c + i\phi^d]$). The Lagrangian is similar to the Abelian case, and subject to the same spontaneous symmetry breaking behavior when $\mu < 0$:

$$\mathcal{L} = (D_\mu\phi)^\dagger D^\mu\phi - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \tag{1.38}$$

where the minimum of the potential is given by $\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}$. This can be satisfied by many possible combinations of the values of the fields $\phi^{a,b,c,d}$. We can freely choose to let the ground state be described entirely by the real component of the neutral entry of the doublet, ϕ^c.

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.39}$$

With this, the ground state is no longer generally invariant under application of the SU(2) $\otimes$ U(1) generators, except for the specific combination that we identified as the electric charge operator, $Q = T_3 + Y_f$. In this way, the U(1) symmetry corresponding to QED remains unbroken and the photon will not gain a mass, as though it were in the ($\mu^2 > 0$) Abelian scenario. We can choose to evolve the field exclusively in the real part of the lower component of the doublet, gauging away the modes of ϕ^a, ϕ^b and ϕ^d, as we did with χ previously:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+h \end{pmatrix} \tag{1.40}$$

We can now plug this into the above Lagrangian with the SU(2) $\otimes$ U(1) covariant derivative:

$$\begin{aligned} D^\mu\phi &= \frac{1}{\sqrt{2}}\left(\partial^\mu - ig\frac{\sigma_a}{2}W_a^\mu - ig'YB^\mu\right)\begin{pmatrix} 0 \\ v+h \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\left(\begin{pmatrix} 0 \\ \partial^\mu h \end{pmatrix} + \frac{(v+h)}{2}\begin{pmatrix} ig(W_1^\mu - iW_2^\mu) \\ ig'B^\mu - igW_3^\mu \end{pmatrix}\right) \end{aligned} \tag{1.41}$$

$$(D_\mu\phi)^\dagger = \frac{1}{\sqrt{2}}\left(\begin{pmatrix} 0, & \partial_\mu h \end{pmatrix} + \frac{(v+h)}{2}\begin{pmatrix} -ig(W_\mu^1 + iW_\mu^2), & -ig'B_\mu + igW_\mu^3 \end{pmatrix}\right) \tag{1.42}$$

The kinetic term evaluates to:

$$\begin{aligned}(D_\mu\phi)^\dagger D^\mu\phi &= \frac{1}{2}\partial^\mu h\partial_\mu h + \frac{(v+h)^2}{8}\left(g^2(W_1^\mu + iW_2^\mu)(W_1^\mu - iW_2^\mu) + (g'B^\mu - igW_3^\mu)^2\right)\\ &= \frac{1}{2}\partial^\mu h\partial_\mu h + \frac{(v+h)^2}{8}g^2W_\mu^- W^{+\mu} + \frac{(v+h)^2}{8}(g'B_\mu B^\mu - gW_\mu^3 W_3^\mu)\\ &= \frac{1}{2}\partial^\mu h\partial_\mu h + \frac{(v+h)^2}{4}g^2W_\mu^- W^{+\mu} + \frac{(v+h)^2}{8}(g'^2 + g^2)Z_\mu^0 Z^{0\mu}, \end{aligned} \tag{1.43}$$

which finally produces mass terms for the weak bosons, proportional to the vacuum expectation value and the SU(2) and U(1) gauge couplings. These terms indicate that $m_{Z^0} = \frac{v\sqrt{g'^2+g^2}}{2}$, and $m_{W^\pm} = \frac{vg}{2}$. These values can be measured via experiment, and are found to be roughly $m_{Z^0} \approx 91.2$ GeV [5] and $m_{W^\pm} \approx 80.4$ GeV [6]. The combination of B^μ and W_3^μ which creates the photon is not represented, and it does not gain a mass term. The kinetic term also contains interactions between the weak bosons and h. The potential part of the Lagrangian contains a mass term for scalar Higgs boson:

$$\mathcal{L} \supset -\mu^2\phi^*\phi = -\frac{1}{2}2\mu^2(h+v)^2 \tag{1.44}$$

yielding a mass of $m_h = \sqrt{2\mu^2}$. This value has also been experimentally determined, with a value of $m_h \approx 125$ GeV.

The final mechanism we need to complete the SM is some way of giving the fermion fields masses, due to our inability to give them Dirac mass terms. We can use the Higgs doublet to create new terms which may satisfy the SU(2) $\otimes$ U(1) symmetry we require while giving the particles mass. We now expand the full SM Lagrangian to include interactions between the left and right handed fermion fields and the Higgs:

$$\begin{aligned}
\mathcal{L}_{Yuk} &= \frac{g_Y}{\sqrt{2}}\bar{\psi_L}\phi\psi_R + \frac{g_Y}{\sqrt{2}}\phi^\dagger\bar{\psi_R}\psi_L \\
&= \frac{g_Y}{\sqrt{2}}\bar{\psi_L}\begin{pmatrix} 0 \\ v+h \end{pmatrix}\psi_R + \frac{g_Y}{\sqrt{2}}\begin{pmatrix} v+h, & 0 \end{pmatrix}\bar{\psi_R}\psi_L \\
&= \frac{g_Y}{\sqrt{2}}\left(\bar{f}_{\downarrow L}(v+h)f_R + \bar{f}_R(v+h)f_{\uparrow L}\right) \\
&= \frac{g_Y}{\sqrt{2}}v\left(\bar{f}_{\downarrow L}f_R + \bar{f}_R f_{\uparrow L}\right) + \frac{g_Y}{\sqrt{2}}h\left(\bar{f}_{\downarrow L}f_R + \bar{f}_R f_{\uparrow L}\right),
\end{aligned}$$

where $f_{\uparrow L}$ and $f_{\downarrow L}$ are the fermions in the upper and lower components of the left-handed doublets, and f_R are the corresponding right handed fields. The term on the left of the above expression is a valid mass term, which can be implemented for am SM fermions, except for neutrinos because they do not have a right-handed field. The mass is proportional to the scalar vacuum expectation value, and the coupling constant g_Y: $m_f = \frac{g_Y}{\sqrt{2}}v$. The term on the right is a coupling between the fermions and the Higgs field, which can be parameterized in terms of the mass:

$$\begin{aligned}
\mathcal{L}_{Yuk} &= m_f\left(\bar{f}_{\downarrow L}f_R + \bar{f}_R f_{\uparrow L}\right) + \frac{m_f}{v}h\left(\bar{f}_{\downarrow L}f_R + \bar{f}_R f_{\uparrow L}\right) \\
&= m_f\bar{f}f + \frac{m_f}{v}h\bar{f}f,
\end{aligned}$$

where f runs over all up-type quarks, down-type quarks, and charged leptons.

1.1.6 Dark Matter and BSM

The SM fields and interactions described up to this point are extremely descriptive for nearly all sub-atomic processes one could hope to witness in a contemporary particle physics experiment. However, some observable phenomena in our universe is not well explained by the fields included in the SM. A glaring blindspot for the theory is the evidence of "dark matter" in astronomical and cosmological measurements. Dark matter manifests as outsized gravitational effects, suggesting some massive presence, with no detectable coupling to any SM particles. These effects manifest in a wide variety of different observables, though it was originally discovered and investigated for its impact on the velocities of stars moving through galaxies.

Speculation that some invisible matter was needed to explain aberrant stellar velocity distributions dates back over 100 years [7], but the theory gained mainstream acceptance in the late 1970s due to the findings of Vera Rubin and Kent Ford [8] on spectroscopic observations of galaxy rotation curves. In the following decades, further evidence for the presence of dark matter emerged via other experimental modes. These include, but are not limited to, anomalous temperature anisotropies at small angular sizes in the cosmic microwave background power spectrum [9], and the effects of gravitational lensing in optical measurements that out scale the predicted effect from pure baryonic matter [10]. These observations have contributed to the current cosmological paradigm (ΛCDM), where it is estimated that universe contains roughly five times as much dark matter as typical baryonic matter. None of the SM particles are good candidates to explain these measurements.

It is natural to consider that dark matter may be composed of some new type of particle that couples feebly, if at all, to the SM fields, but nonetheless has some mechanism to gain mass and induce gravitational effects. If we assume that there is some small coupling between the SM and dark matter sectors, it may be possible to observe it directly or indirectly through experiments here on Earth. This thesis describes a search for hypothetical particles that might mediate this small interaction, making use of collider data from the CERN LHC. Some candidate models to produce this interaction are described in Section 4.2.

Chapter 2

Collider Physics and the LHC

2.1 History of Collider Physics

The history of modern particle physics is entwined with the technological advancement of particle accelerators. The SM as we know it today was not delivered fully formed, with full predictions for all the fundamental particles and their behaviors, but was instead built up gradually, as newer accelerators gave access to higher energies processes. While some predictions and inferences can be made by probing lower energies, a definitive, direct observation of a new particle necessitates collisions with center of mass (CM) energy of at least the rest mass of the particle in question. Because of this, the reach of discoveries in particle physics has been historically gated by the energy of the most power collider at any given time.

Many different methods of raising particles to high energies have been created and applied throughout the past century to study different elements of the SM, however, circular colliders, specifically those colliding hadrons, have been the dominant technology for advancing the energy frontier since they were first introduced. The first of these was the Intersecting Storage Rings (ISR) at CERN, which began construction in 1966 and had its first beam test in late 1970 [11]. The peak beam energy was 31.4 GeV resulting in a CM energy of 62.8 GeV. A major challenge for the ISR, as with for any $p\bar{p}$ collider, is the production and cooling of antiprotons via stochastic cooling [12]. One of its most significant physics triumphs was the first observation of charmed baryons such as the Δ_c^+ [13]. It also reported observations of the quarkonia mesons, J/Ψ and Υ [14, 15], though it famously "missed" being the first to discover them, due to the fiducial volume of the ISR detectors being preferentially forward relative to the beam. The ISR ended operations in 1984.

The next large circular hadron collider project was the Super Proton-Antiproton Synchrotron (Sp$\bar{\text{p}}$S) at CERN, built by upgrading the already operating Super Proton Synchrotron (SPS) to run with two parallel beams in opposite directions, one of protons and another of antiprotons. The SPS was commissioned by the CERN council in 1971, providing its first beam in June of 1976. Though the initial specification was for a 300 GeV accelerator, it was realized during the design process that a 400 GeV beam was possible in the same size ring given the available magnet technology, and the CERN council approved the beam energy to be increased to 400 GeV. Around the time of the first beam, a proposal was raised, led by Carlo Rubbia [16], making a physics case for converting the SPS into a $p\bar{p}$ collider, as a means to observe the, then predicted, weak bosons. The proposal was accepted in 1978, and work began on upgrading the SPS to accommodate a second beam. Precise beam control was the main element of these upgrades, namely tight steering and timing to ensure good luminosity at the collision points, and introducing electrostatic deflectors to avoid unwanted interactions between the beams. Sp$\bar{\text{p}}$S was completed and had its first collisions in 1981 at a CM energy of 540 GeV, which would be increased to 630 GeV in 1984. Data from these collisions collected by the UA1 and UA2 experiments enabled the discovery of the W in 1983 [17, 18], followed by the Z boson in the same year [19, 20]. The Sp$\bar{\text{p}}$S collider program ended in 1991, and was converted back into the SPS for use as a proton and electron source for other CERN experiments.

In 1983, construction was completed on the first ever $p\bar{p}$ collider in the United States, at Fermilab in Batavia, Illinois. The apparatus was called Tevatron, designed raise proton beams to 800 GeV, with collisions at CM energies up to $\sim$1,600 GeV. This was the first time that physics at the TeV scale was accessible in a collider experiment. By the next year, the beam energy had been increased to 900 GeV, then again to 980 GeV in 2001, after the “Main Injector” upgrade campaign [21]. The CDF and DØ detectors were built around the two collision points, with the primary physics goal of discovering the top quark. In 1995, both experiments simultaneously published the first evidence of $t\bar{t}$ pair production via QCD processes [22, 23], completing the fermion sector of the SM. Both experiments also engaged in searches for the Higgs boson, which at the time was the last remaining unobserved component of the current SM. Post-hoc it is known that the Higgs rest mass ($\approx$ 125 GeV) was accessible in Tevatron collisions, and their data likely contained many Higgs events. However, neither experiment was able to resolve a definitive signature with the $\sim$10 fb^{-1} worth of data provided by Tevatron [24], before the program ended in 2011. In 2012, both experiments published a combination result, indicating an excess in their data at invariant masses between 120-135 GeV, with a maximum local significance of 3.3 σ [25], preempting the discovery announcement of the 125 GeV Higgs, which would come just days later.

2.2 The Large Hadron Collider

The current frontier of beam energies in proton-proton collisions is provided by the CERN Large Hadron Collider (LHC). The LHC ring stretches north from the main CERN site into the French countryside, with the far collision point near the city of Cessy. It is constructed in the 27 km tunnel previously occupied by the Large Electron-Positron Collider, which was decommissioned in 2000. The first beam was circulated in 2008, and physics data taking began in early 2010 at a beam energy of 3.5 TeV. The LHC has operated in three “run” periods to date, with multi-year shutdowns in between each run for maintenance and upgrades. Run-I continued until the end of 2012, maintaining a beam energy of 3.5 TeV. Run-II began in mid 2016 and continued until the end of 2018, seeing the beam energy increased by nearly a factor of 2, up to 6.5 TeV. Run-III began in 2022 with a beam energy of 6.8 TeV and is ongoing at the time of writing, planned to stop in 2025.

The LHC is not only advantaged over previous colliders in terms of its available CM energies, but also in its intensity. In Run-I alone, the LHC delivered about $\sim$30 fb^{-1} of collisions; approximately three times as much as the Tevatron produced over its entire lifetime. This is made possible by the fact that the LHC is a pp rather than $p\bar{p}$ collider; ordinary protons are easier to source and cool than antiprotons.

Figure 2·1: Photograph from above of the area of France and Switzerland where the LHC is built. Overlaid is a cartoon of the LHC tunnel, with labels for the locations of the 4 main physics detectors: ATLAS, ALICE, CMS, and LHCb.

There are four collision points around the LHC ring, each of which hosts one of the flagship detector projects; ATLAS, ALICE, LHCb, and CMS. The primary physics objective of the LHC was to discover the Higgs boson, which was promptly achieved just months after the end of Run-I. CMS and ATLAS released results in July of 2012, indicating the presence of a new particle compatible with the SM Higgs at a mass of $\sim$ 125 GeV, observed in a multitude of final states [26, 27]. Both experiments independently reported an observed significance upwards of 5 σ, which is sufficient to claim discovery. The LHC continues to collect data, with the goal of further probing the properties and couplings of the Higgs, as well as searching for evidence of beyond the Standard Model (BSM) phenomena, such as dark sectors or super-symmetry.

The LHC ring and positions of each detector is indicated on Figure 2·1. A detailed technical overview of the CMS detector is provided in the next section.

Chapter 3

The CMS Experiment

CMS is one of the two "general purpose" detectors on the LHC, so called because of its large angular acceptance and diverse set of subdetectors to allow for reconstruction and identification of all known SM particles leaving a pp collision (except for neutrinos). Data collected by CMS can be used for a range of scientific purposes, including precision measurements of known SM quantities, direct searches for undiscovered particles, and broad scope anomaly hunts to find any hints of deviation from the SM. CMS is located at the far collision point in France, directly across the LHC ring from ATLAS and the main CERN campus.

3.1 Specifications

CMS is shaped like a large cylinder resting on its side, skewered by the LHC beam line, with the interaction point directly in the center. The coordinate conventions of the detector are defined relative to this point, taking the interaction point to be the origin, the upwards vertical direction to be the $+y$-axis, the line in the horizontal plane that points towards the center of the LHC to be the $+x$-axis, and the $+z$-axis to be along the beam counter-clockwise when viewed from above. Given the radial symmetry of the detector, it is prudent to define an azimuthal coordinate (ϕ), which is in the $x-y$ plane going up from the $+x$-axis, and a radial coordinate (r), which is the distance from the beam in the same plane. The polar coordinate of the detector (θ) is in the $y-z$ plane, starting from the $+z$-axis. For the purposes of particle physics it is useful to construct a coordinate for this plane that flattens the expected distribution of outgoing particles from the LHC interactions, else the data will seem to be concentrated at very small angles, not much exploiting the functional range of the parameter. The quantity ($\eta = -\ln\tan\left(\frac{\theta}{2}\right)$) called pseudorapidity approximately accomplishes this, as a stand in for true "rapidity" which requires knowledge of the particles mass to calculate. η is the conventional polar coordinate for all CMS analyses. By construction, the angular separation of two particles in the detector is invariant under Lorentz boosts in direction of the beam, when expressed in terms of $\Delta\eta$ [28]. When studying events containing two or more particles presumed to originate from the same process, it is common to describe the angular separation between them in terms of the metric $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

3.2 Detector Structure

The detector is a multi-layered design, composed of various trackers and calorimeters, specialized for recording specific aspects of particles leaving the interaction point. By compositing the information from each of these subdetectors, one can identify the species and energy of the outgoing particles in a given event. The layers of CMS are, ordered from closest to the beam to furthest, the Inner Tracker (IT), Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL), the superconducting magnet, and the muon system. A cut-away illustration of the detector is shown in Fig 3·1.

3.2.1 Magnet

A 3.8 T superconducting solenoid is embedded in the detector to enable charge identification and momentum reconstruction for charged particles in the tracking layers [30]. When charged particles travel through the detector, the magnetic field subjects them to a Lorentz force, deflecting the particle inversely proportional to its momentum in the transverse plane (orthogonal to the beam) [31]. This is especially vital for deter-

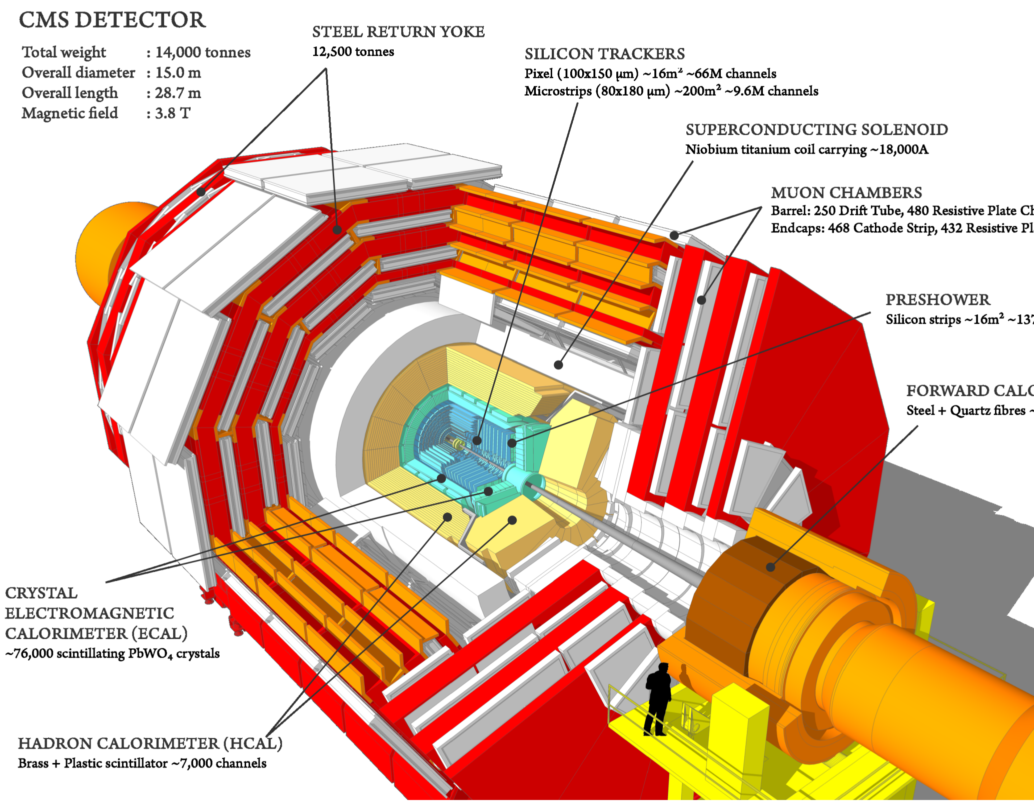

Figure 3·1: A computer generated model of CMS with a portion removed to reveal inner layers. Taken from Ref. [29].

mining the energy of final state muons, as they are not stopped in the calorimeters and their energy must be completely inferred through the shape of their path through the trackers.

The magnet is a hollow cylinder, with an inner radius of $r = 295$ cm, encapsulating the inner tracker and calorimeters. It is composed of wound coils of niobium-titanium alloy, which is superconducting below $T_c = 10$ K [32]. The coils are cooled to 4 K in a cryostat, enabling them to carry the $\sim$18.2 kA of current needed to maintain the 3.8 T magnetic field [33]. A 10,000 ton iron yoke surrounds the magnet, to provide structure and confine the magnetic field.

3.2.2 Inner Tracker

The nearest subdetector to the beam is the IT, an array of multiple layers of silicon pixel and strip detectors that records “hits” from charged particles traveling away from the interaction point. From the pattern of these hits across many layers, tracks of outgoing particles can be identified. Tracks that point back to the same point in space may be associated with the same primary vertex (PV). The IT can be thought of as the combination of two smaller sub-detectors; the inner pixel detector and outer strip tracker.

The pixel detector is composed of four “barrel” layers (BPIX), which are a set of concentric cylindrical pixel arrays parallel with the beam, and three “endcap” layers (FPIX), which sit on both sides of the barrel. The barrel layers are placed at radii of 2.9, 6.8, 10.9, and 16.0 cm and extend out 27 cm in both directions along the z-axis. The endcap layers are placed at distances ± 29.1, ± 39.6, and ± 51.6 cm away from the interaction point along the z-axis, allowing for reconstruction of tracks with η values up to 2.5. The full area of the detector represents about 2 m^2 of silicon when laid flat. The standard size of a tracker pixel is 100 μm x 150 μm, which requires roughly 124 million individual pixels to tile the whole detector detector area. The pixel detector requires this immense level of granularity because it sits closest to the beam and is recording hits from outgoing particles before they have had a chance to spatially separate very much. Without very precise spatial resolution, it would be impossible for the tracker to distinguish hits from tracks that are closely aligned. The spatial resolution is found to be $\sim 10\mu$m in the plane transverse to the beam, and $\sim 20\mu$ along the beam [34].

Encompassing the pixel detector is the strip tracker. The layout of the strip tracker is more complicated that the pixel detector, with two sets of barrel and endcap layers of different geometries. The tracker inner barrel (TIB), has 4 layers between

$26 < r < 49$ cm and extends to $z = \pm 70$ cm. Surrounding this are 6 additional barrel layers, the TOB, spaced between $61 < r < 109$ cm and $z = \pm 110$ cm. A set of 9 endcap disks, the tracker endcaps (TEC), span the range $120 < z < 280$ cm on either side of the TOB. The final component is the tracker inner disks (TID), which are 3 endcap disks that occupy the area on both sides of the TIB that is shadowed by the outer barrel [35]. Figure3·2 illustrates the layout of these subsystems with a cross-section of a single detector quadrant. Between all subdetectors, the average width of the strips (pitch) ranges between 80-180 μm, with finer pitches generally being found at smaller radii. Most of the strip layers are "single-sided" detectors, with the long side of the strips oriented along the plane of the polar angle. For some cells, the long dimension of as strip can be up to 25 cm, giving severely limited spatial resolution along the z-axis in the barrel and the r-axis in the endcaps. However, the first two layers of each subsystem and the fifth layer of the TEC are double-sided, with a second set of strips in the same module, rotated by $\sim 6°$from the first. This enhances the spatial resolution, which is measured to be ~ 30 $(45)\mu$m in $r - \phi$ space and ~ 230 $(530)\mu$m in $r - z$ for the TIB (TOB). [36]. The total active silicon area of the strip tracker is close to 200 m^2, and is contains roughly 9.6 million detector channels [37].

3.2.3 The Electromagnetic Calorimeter

The ECAL is the first of the two calorimetry subsystems, specialized for reconstructing the energy of electromagnetic (EM) showers. The bulk of the detector is made of scintillating lead-tungstate crystal ($PbWO_4$), chosen for its high transparency and radiation hardness [39], and its relatively short electromagnetic radiation length ($X_0 \approx 0.89$ cm) [35]. Like the other subdetectors is constructed with a barrel section, which begins at $r = 129$ cm from the beam-axis and extends to a pseudorapidity of $|\eta| = 1.48$, and an endcap, which begins at $z = 315$ cm and covers the range $1.49 < |\eta| < 3.0$. The barrel volume consists of 61,200 individual crystals, tapered

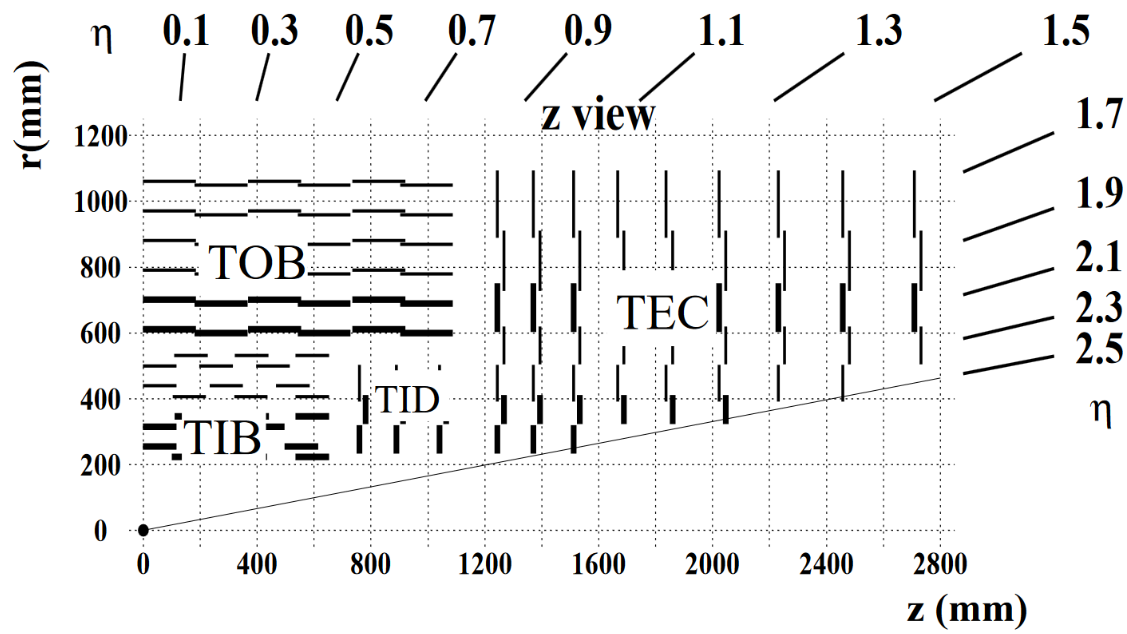

Figure 3·2: Cartoon of strip tracker layout, viewed in the $r-z$ plane. Taken from Ref. [38].

to be narrower closer to the beam, such that they can better fill the cylindrical volume. Crystals in the barrel are typically 22 mm $\times$ 22 mm on the small face and 26 mm $\times$ 26 mm on the larger face, with a depth of $\sim$ 23 cm. Each crystal is aligned such that the length points back towards the beam-spot, but slightly askew, to avoid the case of outgoing particles traveling perfectly along the cracks between crystals. The depth of the barrel crystals gives them full radiation stoppage of $25.8X_0$, giving excellent containment of all EM showers. The endcaps are each composed of 7,324 crystals, 28 mm $\times$ 28 mm on the inside face and 30 mm $\times$ 30 mm on the outside. The endcap crystals are $\sim$ 22 cm deep, equating to 24.7 radiation lengths.

Light yield in the barrel is read out by silicon avalanche photodiode (APD)s mounted to the back of the crystals, and in the endcap is read out by vacuum phototriode (VPT)s. APDs are ideal for the barrel, because implementing any vacuum photo detectors in the low $|\eta|$ regions of the experiment will result in significant loss of performance, due to the magnetic field interacting with the avalanche photoelectrons. Likewise, VPTs are used in the endcap due to the intense forward radiation dose during collisions, which would quickly degrade the performance of silicon devices. The endcap VPTs are comparatively less impacted by the magnet because they are more parallel with the field lines and also use only a single gain stage, limiting the opportunity for charges to be deflected [40]. Both endcaps also include preshower detectors, which particles will pass through before interacting with the endcap crystals. Unlike the rest of the ECAL, these are sampling calorimeters, each consisting of two sets of alternating layers of lead absorber and silicon strips. EM objects will undergo a small shower in the lead, and create a cluster of hits in the silicon. The strip pitch is ~ 2 mm, giving it vastly superior spatial resolution to the rest of the calorimeter. Figure 3·3 shows a cross section of a single quadrant of the ECAL.

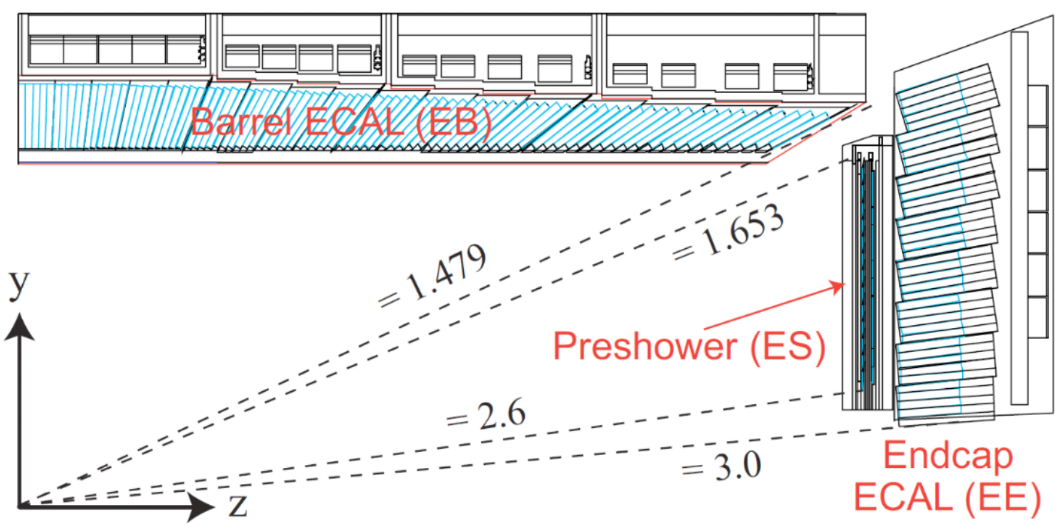

Figure 3·3: Cartoon of ECAL design, viewed in the $r-z$ plane. Taken from Ref. [41].

The main physics role of the ECAL is to aid in the reconstruction of high energy photons and electrons. Electrons traveling through matter will deposit energy via ionization and bremsstrahlung processes. At particle energies of interest to CMS, bremsstrahlung is the dominating effect in electron showering. High energy photons traveling through matter will also create showers, typically beginning with an e^+e^- pair production as the photon passes through the electric field of a nucleus in the material. Regardless of the original particle, both of these processes with repeat in the shower, until the only particles remaining are minimum ionizing, at a number proportional to the energy of the original particle. An EM shower originating from a photon and an electron look identical in the calorimeter, but can be distinguished by using information from the IT. An electron produced at the beam spot will leave hits in the silicon layers as it flies, while a photon will usually penetrate straight through the detector until it encounters the calorimeter bulk. If an ECAL energy deposit is pointed to by a track in the pixel and strip detectors, we can infer that it may be from an electron, while if there is no track it is likely a photon. Photons and electrons can frequently be distinguished in this way, but neutral pions decaying to a pair of photons can be extremely hard to resolve from a single hard photon. This problem is pronounced in the forward region of the detector, which contains many boosted π^0s from QCD processes. For this effect, information from the preshower detector is used. The fine granularity allows for the photons to be resolved separately, even if the showers merge in the endcap crystals.

3.2.4 The Hadronic Calorimeter

The HCAL is designed to record the energy of hadronic showers (jets). Jets are more difficult to study than EM showers, because the mean free path of hadrons in a given material is much longer than electrons or photons. It takes much more calorimeter material to fully stop and record a jet because of the long interaction length, and the light yield is lower per GeV of energy compared to an EM shower [42]. The main body of the HCAL is made up a barrel (HB) and endcap (HE) section. HB sits between the ECAL and the inner edge of the magnet, and extends to $|\eta| = 1.3$.

HE extends the coverage to $|\eta| = 3.0$ [35]. Both sections are sampling calorimeters, made of alternating layers of brass absorber and scintillating plastic, penetrated by wavelength shifting fibers. The fibers collect the light and transport it along a clear transmission fiber to the front end electronics at the edge of the barrel, where it is read out by an array of silicon-photomultipliers (SiPMs) [43]. The calorimeter is divided into towers with constant angular size $\Delta\eta = 0.087$ and $\Delta\phi = 0.087$. This size is constant until $|\eta| = 1.6$ in the endcap, past which the granularity is reduced by a roughly a factor of 2 in both coordinates. Each calorimeter tower is read out separately and is further divided into longitudinal depth segments, each with its own readout channel. Figure 3·4 shows the layout of HB and HE, the grouping of the longitudinal segments, and the location of the front end electronics modules.

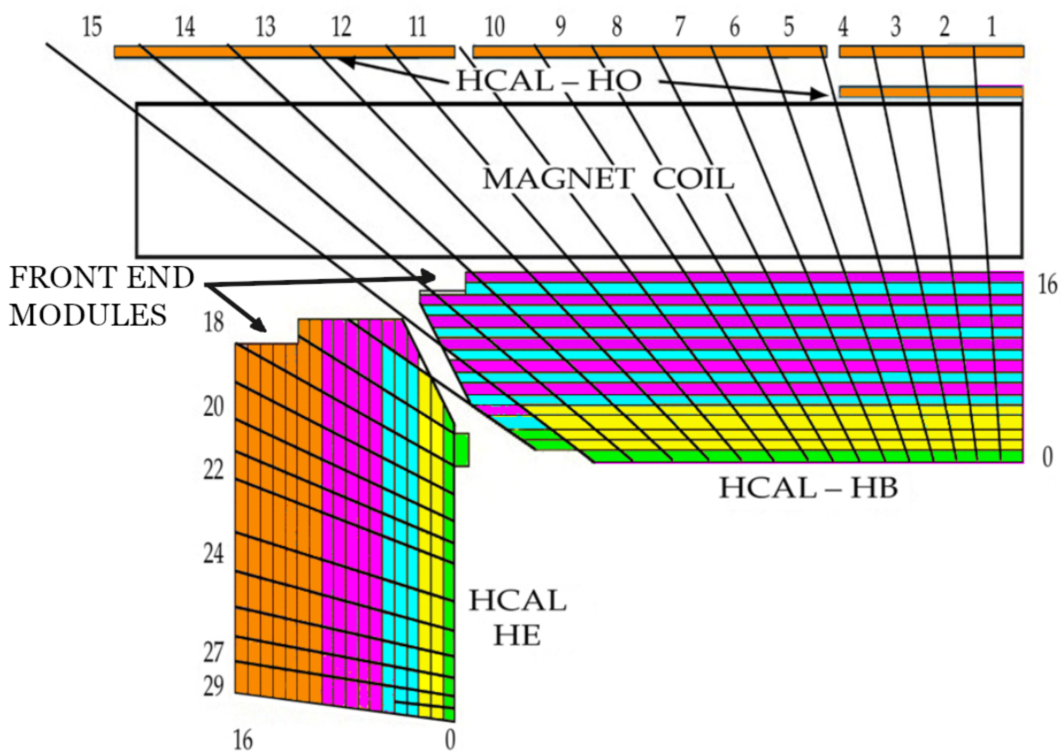

Figure 3·4: Diagram of a single quadrant of HB, HE, and HO. Colors represent different readout depth segments. Taken from Ref. [43].

The stopping power of the endcap is mostly flat as a function of η, with roughly 10 nuclear interaction lengths ($\sim 10\lambda_i$) in each towers. For HB the radius of the barrel is constant, which means that the length of the volume varies with η. At $\eta = 0$, the combined stoppage of both calorimeters is only $\sim 5.8\lambda_i$, which is short enough that some jets may exit the HCAL without producing enough light to accurately sample their energy. To recover some information about these jets, the HCAL Outer (HO) detector is installed outside the magnet volume. For HO, the magnet itself acts as an absorption volume, contributing $\sim 1-2.5$ interaction lengths depending on the angle. A scintillator layer extends out to $|\eta| = 1.3$ to measure these showers at $r = 4.07$ m, just inside the first ring of barrel muon detectors. At critically central angles with $|\eta| < 0.35$, a layer of 19.5 cm steel is placed in front of the scintillator, to further enhance the absorption. An additional plastic sampling layer is also added between the steel and the solenoid. With these additions, the effective energy absorbing depth of the ECAL+HCAL is greater than $11.8\lambda_i$ in nearly all regions of the detector. The position of the HO sampling layers are indicated on Figure 3·4.

The final component of the calorimeter is the HCAL Forward (HF) detector. HF sits far from the other HCAL subsystems, 11.2 m from the interaction point and increases the forward coverage of the detector out to $|\eta| < 5.2$. This ultra-forward region is extremely radiation intense, and requires the sampling material to be swapped for quartz fibers, which is more radiation tolerate than plastic. The absorbing bulk is a steel cylinder with $12.5 < r < 130.0$ cm, and a depth of 165 cm. Unlike the other HCAL subsystems, where the active material is interleaved in layers between sections of absorber, the quartz fibers run through the calorimeter longitudinally, producing Cherenkov light, which flows to the back of the detector to be read out.

Discussion of the AMC13, a backend hardware component that is instrumental to the readout of HCAL data, is provided in Appendix A.

3.2.5 Muon Detectors

The final subdetector is the muon system, which sits entirely outside of the magnet. Between the calorimeters and the magnet volume, it is expected that nearly all charged particles from the collision will have been absorbed before reaching the muon system, save for muons. Fully stopping a muon to reconstruct its energy would require a calorimeter with potentially miles of material depth and the momentum resolution would be quite poor besides. The muon system instead measures the curvature of their paths in the return field of the magnet. At this radius, silicon detectors are infeasible due to cost and engineering considerations, given the massive area it would need to cover. Instead four different varieties of detectors are used to accurately reconstruct the muon path; drift tubes (DTs) and cathode strip chambers (CSCs) are used in the barrel and endcap regions respectively for precise position resolution on the muon hits, and resistive plate chambers (RPCs) are layered between the DTs and CSCs in both regions to enhance timing resolution. gas electron multiplier (GEM) detectors are implemented at very high pseudorapidity, to capture forward going muons.

The DT chambers span the barrel region, up to $|\eta| < 1.2$, woven between the layers of the iron yoke. Each "tube" is roughly 13 mm high and 42 mm across, extending up to 2.4 m [35]. The tubes are filled with a gaseous mixture of argon and carbon-dioxide, which will ionize when a charged particle passes through the tube. A high-voltage (HV) anode wire is suspended in the center of each tube maintained at +3,600 V, as well as +1,800 V electrodes along the top and bottom, and -1,200 V cathode strips on the sides. When a muon ionizes the gas in the chamber, the negative charges will flow towards the wire, creating a current that can be read out to register the hit. The geometry of the tube and a simulation of the electric field is shown in Figure 3·5 (left). The DTs are assembled into stations, each of which has

12 layers of tubes, grouped into 3 sets of 4. The first and last of these sets have the wire oriented along the beam axis to measure the position in ϕ, and the group in the middle is oriented perpendicular to measure the position in η. The arrangement of DTs within a station is shown in Figure 3·5 (right). There are four stations of DTs in the barrel at different radii, with the innermost starting at $r = 402$ cm and the outermost starting at $r = 710$ cm. The outermost station only has 8 layers of DTs, missing the set that runs perpendicular to the beam. The spatial resolution of the DTs is roughly 100 μms in the azimuthal plane, throughout the barrel. [44]

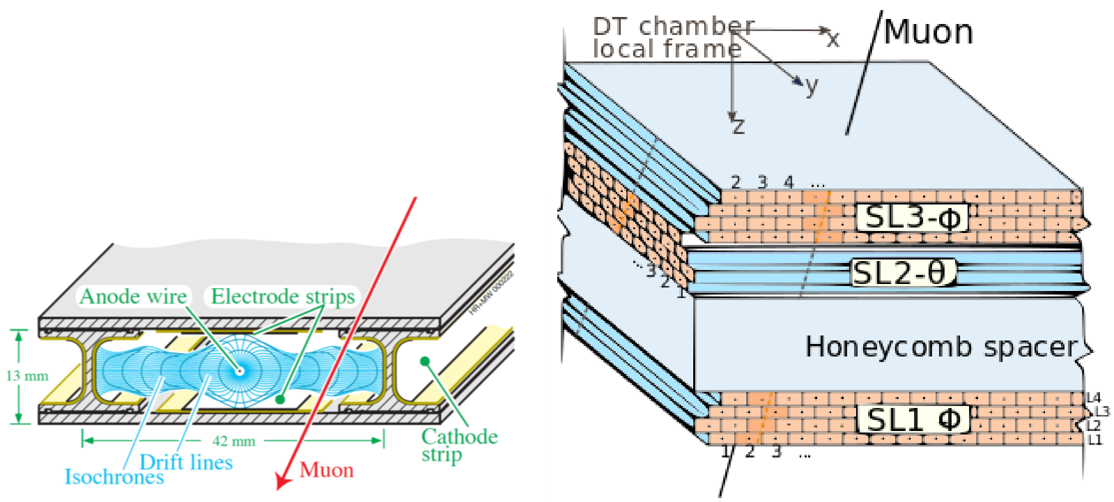

Figure 3·5: Cross section of a single DT chamber, with a visualization of the electric field (left). Arrangement of DT chambers within a station (right). Taken from Ref. [45].

The endcaps cover the range $1.2 < |\eta| < 2.4$, and are assembled from CSCs. The basic working principle of a CSC is similar to that of the DTs; infer hits via the current generated when an active gas is ionized. Rather than crossed layers of tubes, each containing a wire, the CSCs are a larger volume with many HV (+2,900-3,600 V) anode wires running parallel in the ϕ-direction, and planes of cathode strips pointing along the radial direction. Each CSC contains seven cathode planes, spaced by roughly 8 mm in the z direction. The planes are shaped like elongated trapezoids, allowing them to tile the circular endcap geometry. The six regions between the cathode planes are gas-gaps, which are filled with a mixture of argon, carbon-dioxide, and carbon-tetrafluoride. Each gas-gap has a plane of anode wires suspended in the middle, with the wires spaced approximately 3 mm apart. Figure 3·6 shows a cartoon of typical CSC segment. When a muon passes through the CSC and ionizes the gas, negative charges flow to the wires, giving a measurement of the r coordinate, and positive ions will drift towards the cathodes, indicating the ϕ coordinate. The CSC system consists of multiple stations throughout the endcap, extending out to approximately $z = 10.5$ m along the beam axis, and $r = 7.0$ m in the radial direction. The spatial resolution of the CSCs ranges from 50-140 μms in the azimuthal plane, degrading with increased η.

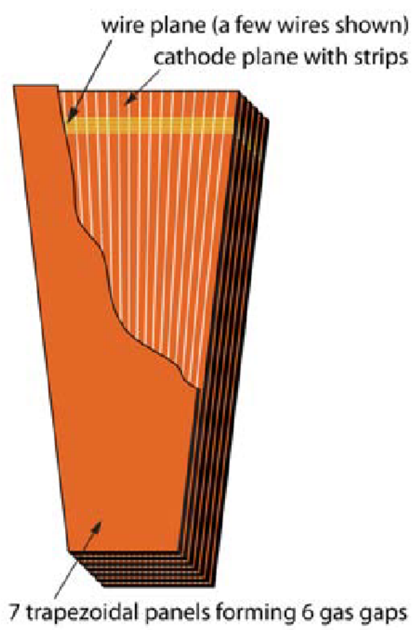

Figure 3·6: Individual CSC segment, viewed from the z direction. Vertical red stripes each represent a cathode strip. Horizontal yellow stripes indicate the wire layout, but is only illustrated in a small section. Taken from Ref. [35]

RPC systems are woven between the other muon detectors and the layers of the return yoke, extending to a pseudorapidity of $|\eta| < 1.9$. The spatial information provided by the RPCs is mostly redundant with the DTs and CSCs, however, it has superior timing resolution, with a precision as fine as $\sim$1 ns (compared with 2-3 ns for the DTs and CSCs) [44]. The RPCs are gas-based detectors, filled with mostly tetrafluoroethane and butane. Each RPC modules has two 2 mm gas-gaps, stacked on top of each other with the HV of both oriented to push negative charges towards the center plane, as shown in Figure 3·7. The voltage difference across the RPCs is significantly higher than in the DTs and CSCs (approximately 10 kV), causing the freed electrons to collide with other atoms in the gas, knocking away electrons and inducing an avalanche. The avalanche charges drift towards the positive electrode at the center of the device, however, the electrodes are coated in a layer of extremely resistive material, which prevents the charges from flowing out along the +HV line. An array of readout strips, with pitches ranging from $1.74-4.10$ cms, sits on the other side of the resistive bulks, near enough to be capacitively coupled to the gas-gaps. As electrons accumulate on the surface of the electrode, positive charges are drawn to the readout strips, inducing a current that can be read out to determine the presence of a hit in that chamber. [46]. The precise timing information of the RPCs makes them ideal for identifying hits that may be out of time with the main collision.

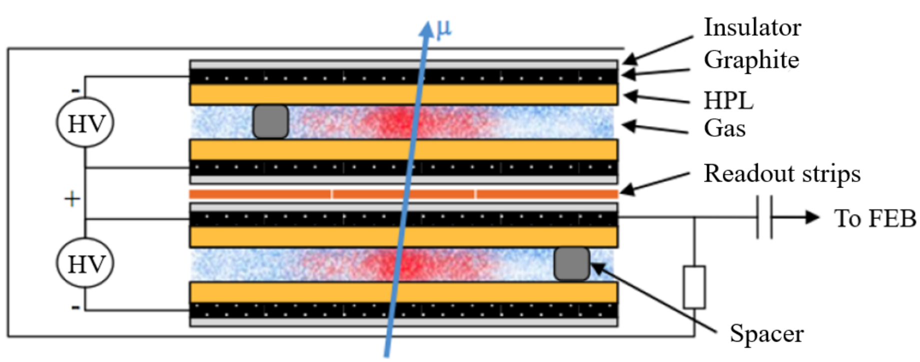

Figure 3·7: Side view of an dual gap RPC. Taken from Ref. [44].

The final component of the muon system is the GEM chambers, newly commissioned for LHC Run-III, to improve coverage in the forward region of the detector ($1.55 < |\eta| < 2.18$). There is currently one GEM station in each endcap, with $z = 5.7$ m. The GEM chamber is filled with a mixture of argon and carbon-dioxide gas. The chamber is ~ 7 cm across, with a 3,200 V drift cathode on one side and a readout board on the other. Spaced within the volume are three gas electron multiplication layers, made of perforated polyimide films with a thin copper coating on each side, held at potential difference of $\Delta V \approx 360$ V. There is an immense electric field within the perforations, drawing in free electrons ionized by incident muons, and funnelling them through the holes. As the electrons pass through, small avalanches are induced due to the strong electric field as the charges advance into the next gas-gap. Because there are three foil layers, this process repeats twice more with multiplicative gains, yielding an amplification in the number of charges up to 10^5 [47]. The charges then drift down to the readout board, so the signal can be recorded, with a spatial resolution of about 100 μms in the azimuthal plane.

A quadrant of the muon system, showing all subdetectors, is shown in Figure 3·8.

3.2.6 Trigger System

The typical size of a CMS event during LHC Run-II, containing the full digitization of the detector readout, is about 2 MBs. LHC collisions happen at a rate of 40 MHz, meaning that the full stream of information produced by the detector is close to 80 TBs of information every second. Storing such a vast quantity of data is not possible, saying nothing of the computing resources that would be required to perform offline reconstruction and analysis. To reduce the fraction of events saved to something more manageable, fast selections are applied via a two-tiered trigger system, which flags events that are more likely to contain hard physics processes [48].

The first of these is the level-1 (L1) trigger, which reduces the event rate from 40 MHz to 100 kHz using low-level reconstructions of tracks in the muon system and

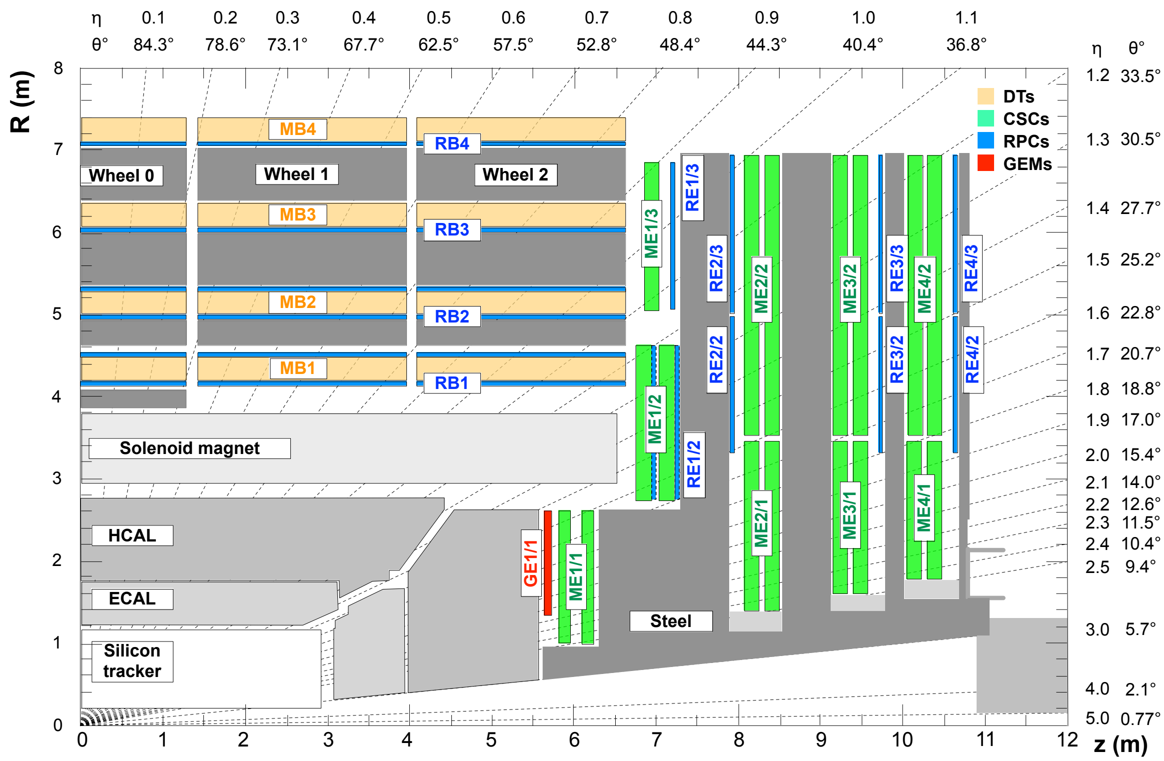

Figure 3·8: Cross section of one quadrant of the CMS muon system. Colored regions correspond to DTs (yellow), CSCs (green), RPCs (blue), and GEMs (red). Taken from Ref. [44].

calorimeter deposits. For the muon trigger, each of the four muon system generate track primitives, reconstructed on or near the front end electronics via fast algorithms to recognize patterns of consecutive hits in that system. These primitives are used by three separate track finding systems to identify likely muon candidates based on the η region of the detector. The "barrel muon track finder" uses primitives from the DTs and barrel RPCs to search for tracks with $|\eta| < 0.83$, the "overlap muon track finder" uses information DTs, RPCs, and CSCs in the region $0.83 < |\eta| < 1.2$, and the "endcap muon track finder" uses the CSCs and endcap RPCs in the range $1.2 < |\eta| < 2.4$. For Run-III, GEM information is also included in the endcap track finder. These reconstructed track candidates are then forwarded to the global muon trigger, which merges track candidates that are likely to be duplicates, and then to the global trigger, which compares them with a predetermined selection of desirable event criteria (L1 seeds), to judge if it is worth promoting to the next trigger stage. The calorimeter triggers follow a similar principle, where the trigger primitives are clusters of deposited energy in both HCAL and ECAL. These are evaluated through multiple layers of processing; first the primitives from ECAL and HCAL are joined into combined regions of interest (towers). Towers are then reconstructed into candidate physics objects, such as photons/electrons, jets, or τ leptons. These primitive objects are passed through a demultiplexer, which reorders the event information such that it can be submitted to the global trigger [49]. Figure 3·9 illustrates the flow of information from the muon and calorimeter front ends to the global trigger in Run-II.

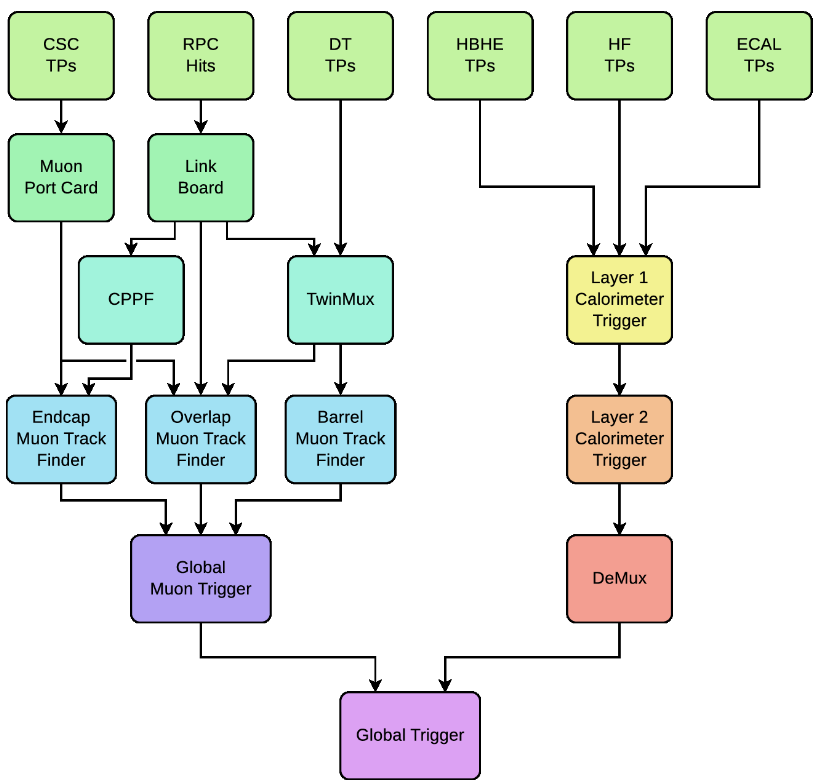

Figure 3·9: Flow chart of L1 trigger information in LHC Run-II. Run-III would also include GEM information pointing to the endcap track finder. Taken from Ref. [49].

Events that satisfy one of the L1 seeds in the global trigger are promoted to the high-level trigger (HLT), which will reduce the 100 kHz L1 event rate to only 1 (2) kHz to be saved offline in Run-II (Run-III). The HLT runs on a farm of CPU nodes, which performs a detailed reconstruction of the physics objects in the event at nearly the level of the full offline reconstruction. The Run-II HLT used exclusively CPUs, but Run-III has implemented “heterogeneous computing,” modifying the reconstruction algorithms to also run on commercial GPUs, granting a roughly 40% decrease in average reconstruction time [44, 50]. After reconstruction, the physics objects in the events are compared with sets of different criteria (paths) included in the current HLT menu. Events which satisfy one or more of these paths are saved to be subjected to a full offline-reconstruction and made available for data analysis. The offline event rate is strictly constrained by the available data transfer and computing resources, however, novel strategies to circumvent these rate limitations have been employed during Run-II and Run-III. These include the B parking and data-scouting programs [51, 52], the latter of which is employed in the analysis portion of this thesis.

Chapter 4

Search for Excess Dimuon Production at Low Mass

The goal of this analysis is to search for direct resonant production of a BSM particle in a range of possible masses between 1 and 8 GeV, which then decays immediately into a pair of opposite sign muons. Given the surplus of possible models containing candidate particles for such a process, it is best to not target a specific model or BSM hypothesis, and instead optimize the search for sensitivity to any anomalous dimuon production. Any resulting constraints on the rate of $pp \to X \to \mu^+\mu^-$ into the experimental fiducial volume will be model-independent and can be utilized by the theory community to set limits on model parameters that depend on this process.

4.1 Bump Hunt Searches

The style of this search is a bump hunt: a common technique to search for new particles in scenarios where the particle decays promptly and the mass can be calculated from the reconstructed final state objects. By curating a dataset that is rich in the production of dimuon pairs and histogramming the events by the mass of the dimuon parent particle, one constructs a continuum of dimuon events that is smooth and mostly comprised of SM Drell-Yan (DY) processes. One expects that if a particle of a given mass exists and decays to muons it will appear in this histogram as excess production (a bump) on top of the smooth background at the value of its true mass. By fitting the data with a smooth, continuous function to model the background, and simultaneously fitting a Gaussian-like signal shape to gauge the presence or absence of an excess on top of the background, we can either detect or exclude the presence of a BSM particle at that mass within some confidence.

Bump hunts are a historically well tested approach, being utilized for many significant particle discoveries. The method was employed for the first observation of the charmonium state J/Ψ by the E598 collaboration [53] and SLAC [54], as well as bottomonium Υ by E288 [55]. A more contemporary example of this approach was the discovery of the Higgs boson in 2012, by CMS and ATLAS, utilizing di-photon and diboson resonances [26, 27]. A notable difference between these two cases is the extent to which irreducible backgrounds are included in the data. Figure 4·1 shows the di-electron mass distribution for the J/Ψ observation from E598 and the di-photon mass distribution from the Higgs observation from CMS. For the former, the resonance is clearly visible to the naked eye and an obvious indication of new physics, while the later is more subtle, requiring intricate signal and background modelling to resolve the excess from the continuum background. For this analysis, we do not expect our peak to be plainly visible and significant effort is spent in developing proper models to control the SM background.

4.2 Benchmark Signal Models

Though the main result of this analysis is model-independent, the results can be reinterpreted as constraints on parameters in common benchmark models, to allow for easier comparison with other experiments. Two models that are considered for this search are a minimal dark photon case, and a pseudoscalar in a two Higgs doublet model (2HDM) scenario with an accessory scalar field.

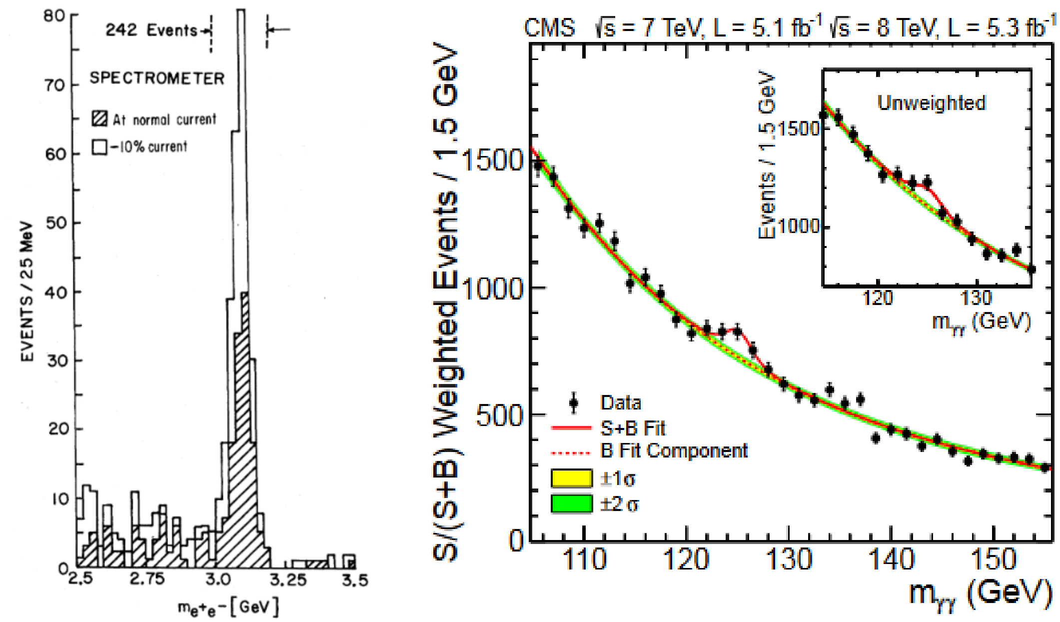

Figure 4·1: Di-electron mass distribution from the E598 discovery of the J/Ψ (left) and di-photon mass distribution from the CMS discovery of the Higgs boson (right). Taken from Refs. [53] and [27].

4.2.1 Dark Photon

Many BSM hypotheses contain some sort of "dark photon" or Z_D; a particle which shares most of the properties of the SM photon, but not the magnitude of its coupling to electric charge. It may or may not be massive and, depending on the model under consideration, it may be a stand-alone particle or a small part of a much larger dark sector. The most minimal construction of a massive dark photon requires extending the SM gauge groups to include an additional $\mathrm{U}(1)_D$ group [56]. A process called kinetic mixing allows for interaction terms between U(1) abelian gauge fields to be included in the expanded Lagrangian:

$$\mathcal{L}_{EM+Dark} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}X_{\mu\nu}X^{\mu\nu} - \frac{\varepsilon}{2}F_{\mu\nu}X^{\mu\nu} + eJ^{\mu}A_{\mu} \tag{4.1}$$

Note the electromagnetic field strength tensor ($F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$), the recognizable SM kinetic term ($-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$), the SM photon field ($A$), and the fermionic current ($J_\mu = \bar{\psi}\gamma_\mu\psi$). The addition of the new gauge group adds a "dark" counterpart to the kinetic term, with a dark photon field (A^D). One could also include a term to allow for the dark photon to interact with a dark fermion current, in an assumption of an extended dark sector, but here we are constructing the minimal case of a lone dark photon. The remaining term of interest is the kinetic mixing term ($-\frac{\varepsilon}{2}F_{\mu\nu}X^{\mu\nu}$) whose strength is proportional to a free parameter ε, which is assumed to be small.

The kinetic mixing term by itself is an opaque feature of this Lagrangian. To understand its phenomenological implications, it is ideal to perform a redefinition of the gauge fields A and A^D into an orthogonal basis of mass eigenstates. By this operation, the kinetic mixing term will become zero, and we can understand the ramifications of possible BSM couplings through the transformed interaction term ($eJ_\mu A^\mu$)[1]. This transformation has the form:

$$\begin{pmatrix} A^D_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1-\varepsilon^2}} & 0 \\ \frac{\varepsilon}{\sqrt{1-\varepsilon^2}} & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} A'^D_\mu \\ A'_\mu \end{pmatrix} \tag{4.2}$$

By taking θ to be small, the right matrix reduces to the identity:

$$\begin{pmatrix} A^D_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1-\varepsilon^2}} & 0 \\ \frac{\varepsilon}{\sqrt{1-\varepsilon^2}} & 1 \end{pmatrix} \begin{pmatrix} A'^D_\mu \\ A'_\mu \end{pmatrix} = \begin{pmatrix} \frac{A'^D_\mu}{\sqrt{1-\varepsilon^2}} \\ \frac{\varepsilon A'^D_\mu}{\sqrt{1-\varepsilon^2}} + A'_\mu \end{pmatrix} \tag{4.3}$$

Given that past experiments have not been sensitive to any evidence of kinetic mixing, it is a safe assumption that $\varepsilon \ll 1$. Therefore we can simplify this equation by neglecting any terms that scale like ε^2:

$$\begin{pmatrix} A^D_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} A'^D_\mu \\ \varepsilon A'^D_\mu + A'_\mu \end{pmatrix} \tag{4.4}$$

We can see that, to first order in ε, the description of the dark photon field is un-

[1]The steps of the following operations were closely followed from references [57] and [58]

changed, but the SM photon field has gained a term that includes the dark photon, with linear dependance on ε. We now recognize A' and A'^D as the "real" photon and dark photon fields and substitute them into the SM interaction term:

$$eJ^\mu A_\mu \rightarrow eJ^\mu(A'_\mu + \varepsilon A'^D_\mu) \tag{4.5}$$

With this interaction term, the dark photon can couple to charged fermions, albeit at a strength suppressed by the small parameter ε. A simple tree-level process that can be drawn from this interaction is $q\bar{q}$ annihilation into a dark photon, followed by a decay back into opposite sign fermions, shown in Figure 4·2. This Drell-Yan process is the anticipated production mode for dark photons produced in the LHC. If the dark photon has a mass, which it can acquire through an Abelian Higgs mechanism as described in Section 1.1.4, then it will be produced resonantly in events with the correct partonic momentum exchange and can be observed in a bump-hunt search. The matrix element to produce the dark photons via the previously mentioned process is suppressed by one factor of the coupling from the $q\bar{q}Z_D$ vertex. Squaring the matrix element indicates that the production rate is proportional to the coupling squared. The branching fraction into muons is the decay width $\Gamma(Z_D \rightarrow \mu\mu)$ divided by the total decay width $\Gamma(Z_D \rightarrow \mathrm{SM})$. Because all SM widths have the same dependance on the kinetic mixing parameter, it cancels out in the branching fraction, leaving it independent of ε [59]. The resonant yield for a dark photon decaying into muons is the product of these two values, which will be directly proportional to ε^2; a common metric that will allow us to benchmark our result relative to other experiments probing this same process.

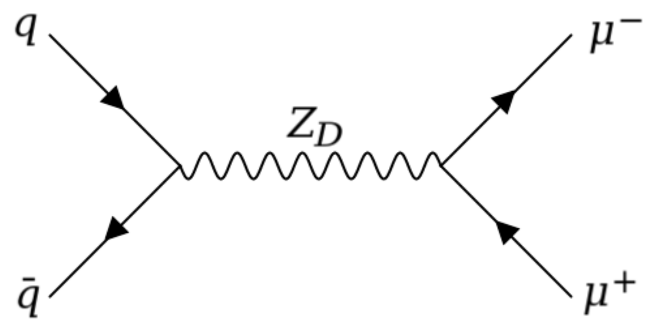

Figure 4·2: Example Feynman diagram of a DY process mediated by a dark photon, which decays into muons.

4.2.2 Light Pseudoscalar

Another possible BSM situation which can produce a dimuon resonance, is a 2HDM scenario where an additional scalar field is included (2HDM+S) [60]. Generic 2HDMs involve extending the SM Higgs field to include an additional SU(2) doublet, similar to the first in structure:

$$\phi_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_1\end{pmatrix} \qquad\qquad \phi_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_2\end{pmatrix} \tag{4.6}$$

As described in Section 1, in the process of electroweak symmetry breaking with a single Higgs doublet, the SU(2) $\otimes$ U(1) gauge fields absorb 3 of the 4 degrees of freedom required for the Higgs field to manifest particles. However, by the addition of a second Higgs doublet, superficially similar to the first, 4 additional degrees of freedom are added to the theory (a real and complex value for both entries of the new doublet), while still only 3 can be eaten by the weak bosons. This leaves 5 physical degrees of freedom, each of which will correspond to a real scalar boson. These include:

1. two CP-even neutral scalars (h and H), one of which is the SM Higgs boson

2. a neutral CP odd pseudoscalar (A)

Type	I	II	III (X)	IV (Y)
q_u coupling	ϕ_2	ϕ_2	ϕ_2	ϕ_2
q_d coupling	ϕ_2	ϕ_1	ϕ_2	ϕ_1
ℓ coupling	ϕ_2	ϕ_1	ϕ_1	ϕ_2

Table 4.1: Different 2HDM configurations, for the most common combinations of fermion couplings. Adapted from Ref [62].

3. two oppositely charged Higgses (H^+ and H^-)

There is a lot of freedom in how one can assemble and apply a 2HDM, but a common constraint is the assumption that each doublet is exclusive in its coupling to each type of fermion. This means that up-type quarks (u, c, t) can couple to ϕ_1 or ϕ_2, but not both. Likewise for down type quarks (d, s, b) and the charged leptons. The purpose of this restriction is to avoid tree-level flavor-changing neutral current (FCNC) processes facilitated by these fields, which can emerge if there are shared couplings. Note that there is nothing "wrong" with including FCNCs in the phenomenological implications of a BSM theory, but they can be undesirable given existing experimental constraints on such processes [61]. Non-FCNC 2HDM theories can be classified by which combination of fermions couples to each doublet, which is summarized in Table 4.1.

The phenomenological differences between these combinations is strongly dependant on the ratio of the vacuum expectation values between the Higgs doublets ($\tan\beta = \frac{v_2}{v_1}$), which is a free parameter of the model. Experiments probing similar channels when targeting these models will often consolidate on a fixed value of $\tan\beta$, to allow for easier comparison of results.

Of interest to this analysis is a further extension of 2HDM, wherein the universe includes not just an additional Higgs doublet, but also a complex singlet scalar field (S)[63]. The scalar couples only to the two Higgses and does not interact directly with the other SM fields. One can construct a massive pseudoscalar particle state (a) via mixing of the pseudoscalar A with the imaginary portion of S:

$$a = \cos(\theta_{\mathrm{H}})\mathrm{Im}(S) + \sin(\theta_{\mathrm{H}})A \tag{4.7}$$

This state has a Yukawa coupling to the SM fermions, proportional to the fraction of the state that is attributed to A as opposed to S, denoted as $\sin(\theta_{\mathrm{H}})$. For a given fermion, f, this coupling has the form.

$$\mathcal{L}_{\mathrm{Yuk}} = -i\frac{m_f}{v}\sin(\theta_{\mathrm{H}})\kappa_f \bar{f} f a \tag{4.8}$$

This coupling is similar to other SM Yukawa terms, except that it is additionally dependant on κ_f and $\sin(\theta_{\mathrm{H}})$. κ_f is either equal to $\pm\tan\beta$ or $\pm\frac{1}{\tan\beta}$, depending on which fermion and type of 2HDM are being considered. The possible combinations are summarized in Table 4.2. Unless $\tan\beta$ is equal to 1, it is clear that certain decay modes will be favored in different scenarios, as the strength of the Yukawa coupling will be enhanced or penalized according to this parameter. The possibility that best favors this analysis is one that has a strong branching fraction to muons in our targeted mass range. Figure 4·4 shows the branching fraction to all SM particles for a variety of 2HDM modes and values of $\tan\beta$. It is found that the best case scenario for a discovery in this channel is a Type IV 2HDM (also known as "Y" or "Flipped"), with $\tan\beta < 1$. Similar analyses use $\tan\beta = 0.5$ as the benchmark value for setting constraints on this process, facilitating easy comparison between different results.

With $\tan\beta$ fixed, the coupling depends only on the pseudoscalar fraction, $\sin(\theta_{\mathrm{H}})$, which scales directly with the rate of producing the particle a. The results of this analysis will be presented as an upper limit of the possible value of $\sin(\theta_{\mathrm{H}})$ in the previously mentioned 2HDM scenario. We expect that in LHC collisions, the most

2HDM Case	I	II	III	IV
q_u	$\frac{1}{\tan\beta}$	$\frac{1}{\tan\beta}$	$\frac{1}{\tan\beta}$	$\frac{1}{\tan\beta}$
q_d	$-\frac{1}{\tan\beta}$	$\tan\beta$	$-\frac{1}{\tan\beta}$	$\tan\beta$
ℓ	$-\frac{1}{\tan\beta}$	$\tan\beta$	$\tan\beta$	$-\frac{1}{\tan\beta}$

Table 4.2: Dependance of pseudoscalar Yukawa term on ratio of Higgs vacuum expectations for all combinations of fermion and 2HDM type. Taken from Ref. [64].

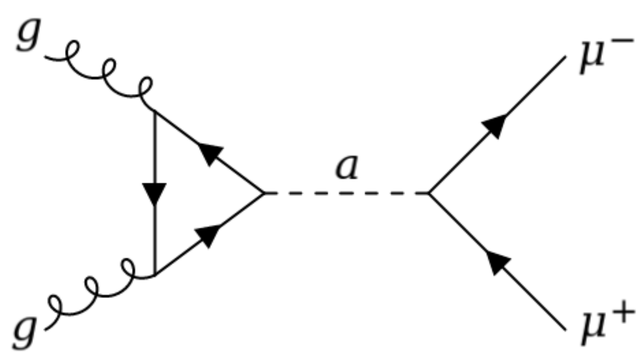

Figure 4·3: Example Feynman diagram of a ggF process producing a light pseudoscalar, which decays into muons.

likely process to facilitate production of this particle is gluon-gluon fusion (ggF), as shown in Figure 4·3, similar to the SM Higgs. This production mode is enhanced by the Type IV model choice, via the top-quark (up-type) contributions to the loop.

4.3 Low Mass Searches

CMS searches for resonant BSM production are often probing significantly higher masses than this analysis ($m_X > 100$ GeV). These searches are well motivated, as the LHC is the first collider to give access to events with partonic momentum exchange in the multi-TeV range. New physics signatures could be plainly obvious, having strong couplings to the SM fields, and were only previously unobserved due to the high CM energies required to create them. Unfortunately, these searches have to date not recorded conclusive evidence of any BSM particles or mechanisms [65, 66], so it is logical to expand searches to lower masses. It is possible that previous experiments may have "missed" the observation of a new particle due to its feeble coupling to the SM fields. This analysis seeks to surpass the sensitivity of previous measurements and probe for new physics of exceptionally weak couplings, by leveraging the incredible amount of data provided by the LHC.

A major challenge in attempting to use CMS data to probe low masses is the inefficiency of the main physics triggers in recording dimuon events with masses below about 45 GeV [52]. The standard HLT trigger threshold for dimuon events in Run-II required the leading (subleading) muon to have a transverse momentum of at least 17(8) GeV. For muons of such momentum to have originated from a common parent particle of lower mass than the momenta of the final state particles, implies that the parent is strongly boosted in the transverse plane. This greatly reduces the fiducial acceptance for these low mass dimuon decays, making it nearly impossible accrue the required statistics for a competitive analysis in this mass range. To account for this, this analysis makes use of a specialized readout stream called "data-scouting", which operates in parallel with the main physics readout, albeit with significantly reduced event content and custom triggers with lowered momentum thresholds. The scouting datasets were previously used in a similar CMS dimuon resonance search [65], targeting comparatively higher masses ($11 < m_X < 45$ GeV), but still low enough that the scouting triggers provided significant improvements in sensitivity relative to the standard triggers. Figure 4·5 shows the model independent limit on BSM dimuon production from this analysis. More details about data-scouting will be provided in Section 4.4.1.

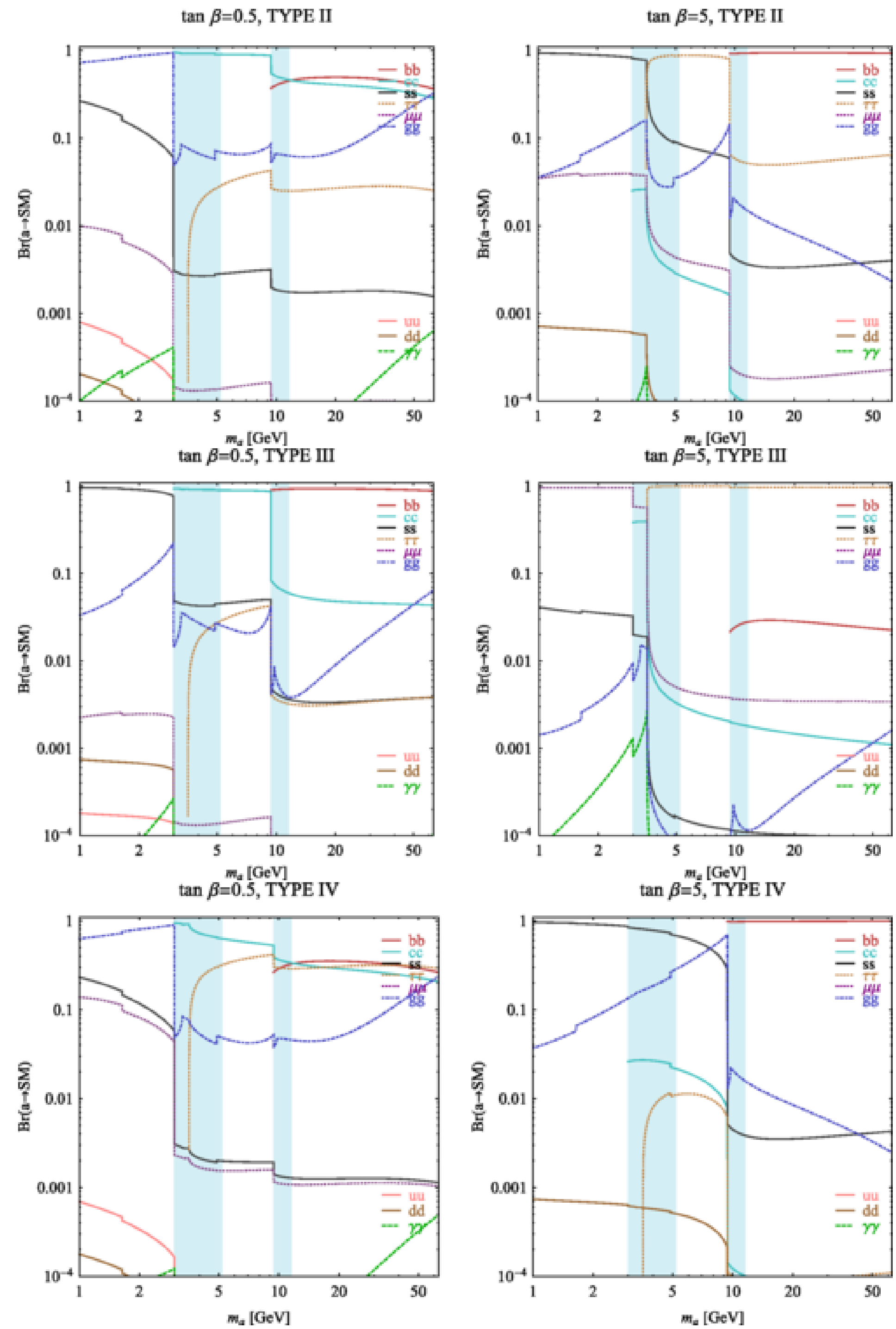

Figure 4·4: The predicted branching fraction for the pseudoscalar a to decay to any SM particle. The rows correspond to a type II, III, and IV 2HDM case respectively. The left (right) column shows the case of $\tan\beta < 1$ ($\tan\beta > 1$). The scenario considered for this analysis is the bottom left; note the relative enhancement of the muon branching fraction (magenta line). Taken from Ref. [63].

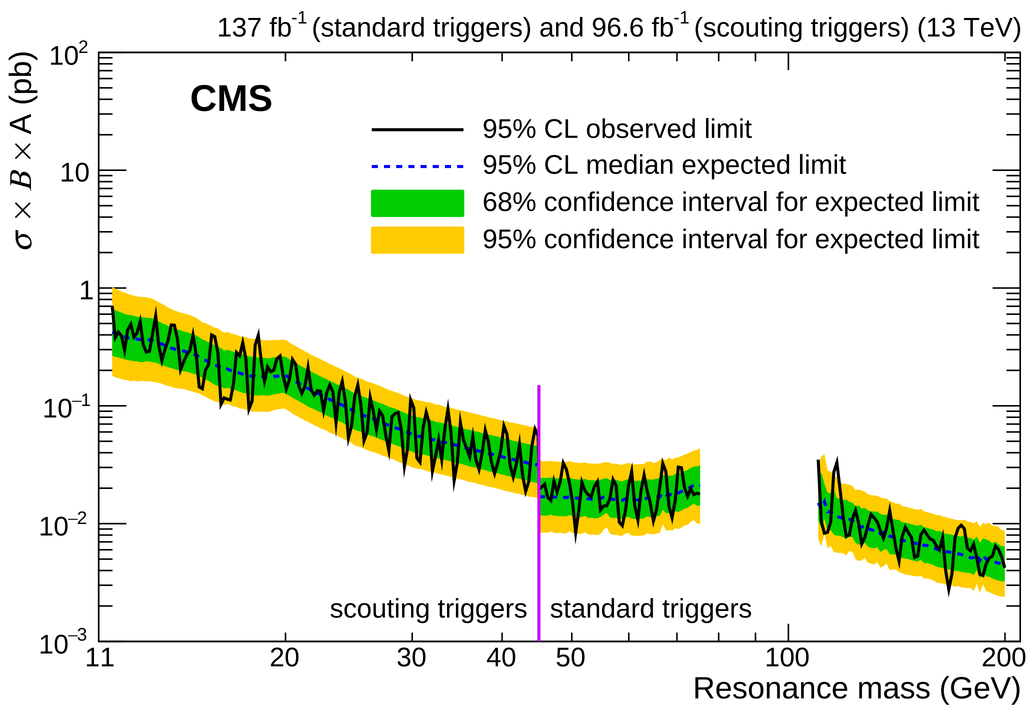

Figure 4·5: Model independent limit on excess dimuon production between 11 and 200 GeV. The purple line marks the point where the standard triggers are no longer efficient and lower masses are more optimally probed with data-scouting. Taken from Ref. [65].

4.4 Dataset Construction

The first step in this analysis is to build a sample of dimuon input events that maximally includes events we deem to be signal-like and rejects those that are background-like. This is done both by selecting triggers that have physics requirements that closely match our desired signal kinematics, as well as by constructing offline requirements to select events that most closely resemble resonant production.

4.4.1 Scouting Datasets

The relevant CMS datasets for this analysis are the Run-II dimuon scouting datasets. These recorded 35.3 fb^{-1} of pp collisions at 13 TeV in 2017 and an additional 61.3 fb^{-1} in 2018. As previously stated, the triggers that correspond to these datasets have dramatically lowered momentum thresholds on the candidate physics objects, significantly enhancing the trigger rate for low energy processes. To avoid overburdening the dedicated computing and storage resources for the nominal physics readout stream, these "scouting events" are not subjected to the full CMS offline reconstruction process, and the full readout of the detector is discarded before saving the event. The only information that is retained for these events is the reconstructed physics objects at the level of the HLT. In Run-II, the scouting datasets recorded events at rates up to $\sim$ 5kHz at peak luminosity, roughly 5 times faster than the standard triggers, while requiring only a fraction of the offline computing and storage resources. Figure 4·6 shows the dimuon mass distribution collected by the Run-II scouting triggers and the standard single and double muon triggers, normalized to the same integrated luminosity, which illustrates the masses where the scouting triggers are most impactful in increasing dimuon acceptance.

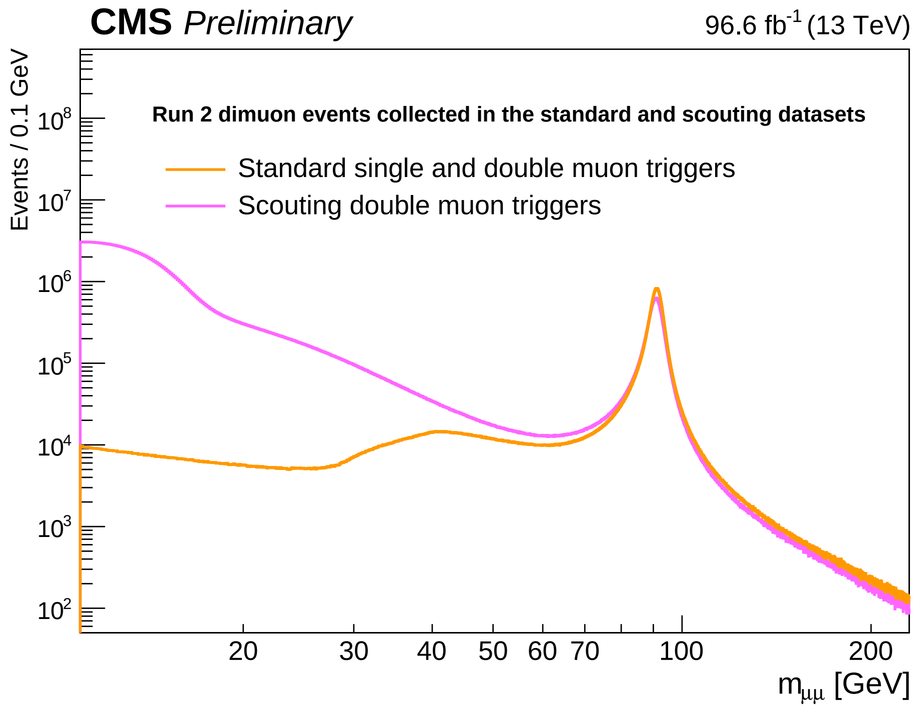

Figure 4·6: Comparison of the Run-II recorded dimuon invariant mass distribution between the standard and scouting triggers, normalized to 96.6 fb^{-1}. Taken from Ref. [67]

Scouting L1 Seeds

The dimuon scouting trigger has a set of specialized L1 seeds to target low p_{T} muon pairs. This analysis makes use of four of these. The first requires only the presence of two muons with the leading muon having $p_{\mathrm{T}} > 15$ GeV and the subleading muon having $p_{\mathrm{T}} > 7$ GeV. The second requires that the both muons have $p_{\mathrm{T}} > 4.5$ and lie within the pseudorapidity range $|\eta| < 2.0$. It additionally requires that the muons be of opposite charge and have an invariant mass between 7 and 18 GeV. The third requires that both muons are of opposite charge, with $p_{\mathrm{T}} > 4$ (4.5) GeV in 2017 (2018) and an angular separation $\Delta R < 1.2$. The final trigger has no momentum requirement, needing only two opposite sign muons within the pseudorapidity range $|\eta| < 1.5$ and an angular separation $\Delta R < 1.4$. These requirements are summarized in

L1 Trigger	p_{T} [GeV]	$\|\eta\|$	ΔR	$m_{\mu\mu}$	Charge
I	>15/7	—	—	—	—
II	>4.5	<2.0	—	7–8	Opposite
III	>4.0 (4.5)	—	<1.2	—	Opposite
IV	—	<1.5	<1.4	—	Opposite

Table 4.3: The physics requirements for an event to pass the L1 scouting triggers used by this analysis. The p_{T} requirements for trigger III refer to 2017(2018) respectively.

Table 4.3. This analysis uses an inclusive OR of all of these trigger paths, and they are henceforth treated as an amalgamation "L1 scouting trigger".

HLT Dimuon Scouting Trigger

Dimuon events that pass these L1 selections are forwarded to the HLT CPU farm, and are subjected to a fast version of the CMS event reconstruction software. While many of these scouting candidate events will contain particles of insufficient momenta to satisfy the standard dimuon HLT paths, a special scouting path is employed to record them. This trigger requires only the presence of two muons, with $p_{\mathrm{T}} > 3$ GeV and the activation of one or more of the L1 scouting seeds. Events that meet these requirements are immediately saved to disk with a small subset of the total event information, the rest being discarded. The saved information includes the hits in the tracking system and muon chambers, reconstructed energy clusters from the calorimeters, and the physics objects included in the event at the level of HLT reconstruction. The result of this is a significantly reduced event size of around 5 kB per event, in contrast with a standard CMS event being roughly 1 MB. This alleviates both the burden of storing the full size of these events, and the offline computing resources needed to do a full event reconstruction allowing for a significantly higher rate of events to be saved. Fig. 4·7 shows the dimuon mass distribution of events saved by the scouting trigger in 2017 and 2018, broken down by which of the L1 scouting seeds were activated.

While the lowered momentum thresholds provided by the scouting triggers are a significant advantage to this analysis, there are some additional difficulties posed by using them. The standard CMS datasets are the subject of rigorous study by dedicated working groups within the collaboration. These groups release performance measurements for all of the standard triggers and define common benchmark working points to describe physics objects of various qualities. The scouting datasets, being a comparatively newer program, do not benefit from this centralized support and analyzers hoping to utilize them are responsible for producing trigger performance studies and offline object definitions, which will be described in later sections.

4.4.2 Simulation

Many searches for BSM signatures rely on robust simulated datasets to model expected background contributions for certain processes and probe for excesses. For this search, our prediction of the background yield is encoded entirely in the assumption that the dimuon mass continuum is smoothly distributed around our mass hypotheses for each window we will fit. As long as our modelling for this smooth continuum is accurate, we can constrain our background yield using the data sidebands and do not need to produce full simulation of our background processes.

However, simulated samples are still needed to support our understanding of the triggers, offline event selections, and fiducial acceptance for different signal modes. Two SM samples are produced with PYTHIA 8.230 [69] to corroborate our data-driven efficiency measurements; a continuum DY sample containing dimuon pairs between 1 and 10 GeV, and a resonant sample containing $J/\Psi \to \mu\mu$ and $\Upsilon \to \mu\mu$ events. Both samples are produced at leading order (LO) in perturbative QCD. Two BSM signal samples are also produced to compare the efficiency of these objects to pass our analysis selections relative to the SM process in data that are used as proxies when defining them. The first is a dark photon sample is produced using MADGRAPH5_aMC@NLO v3.4.1 [70] at next-to-LO in perturbative QCD. This sample uses a hidden Abelian Higgs model (HAHM), HAHM_variableMW_UFO, which is effectively the BSM scenario described in Section 4.2.1, but is so named for the additional Higgs state that it contains. The stand-in value for the dark photon mixing strength in this sample is $\varepsilon = 0.02$. The final simulated dataset is a pseudoscalar sample to emulate ggF production for the 2HDM+S limit, produced with PYTHIA 8.230.

Parton showering and hadronization for all samples is done with PYTHIA 8.230, using the CP5 tune [71]. While accurate showering and hadronization modelling is important for any simulated pp data, these samples are particularly dependant on the production of initial state radiation (ISR) jets to boost the muons above the scouting trigger momentum thresholds, since a low mass parent particle ($m_X < 6$ GeV) decaying at rest will not produce muons with enough momentum to satisfy the scouting trigger thresholds. Simulation of the detector is done using the GEANT4 software [72], to model hardware effects and particle reconstruction.

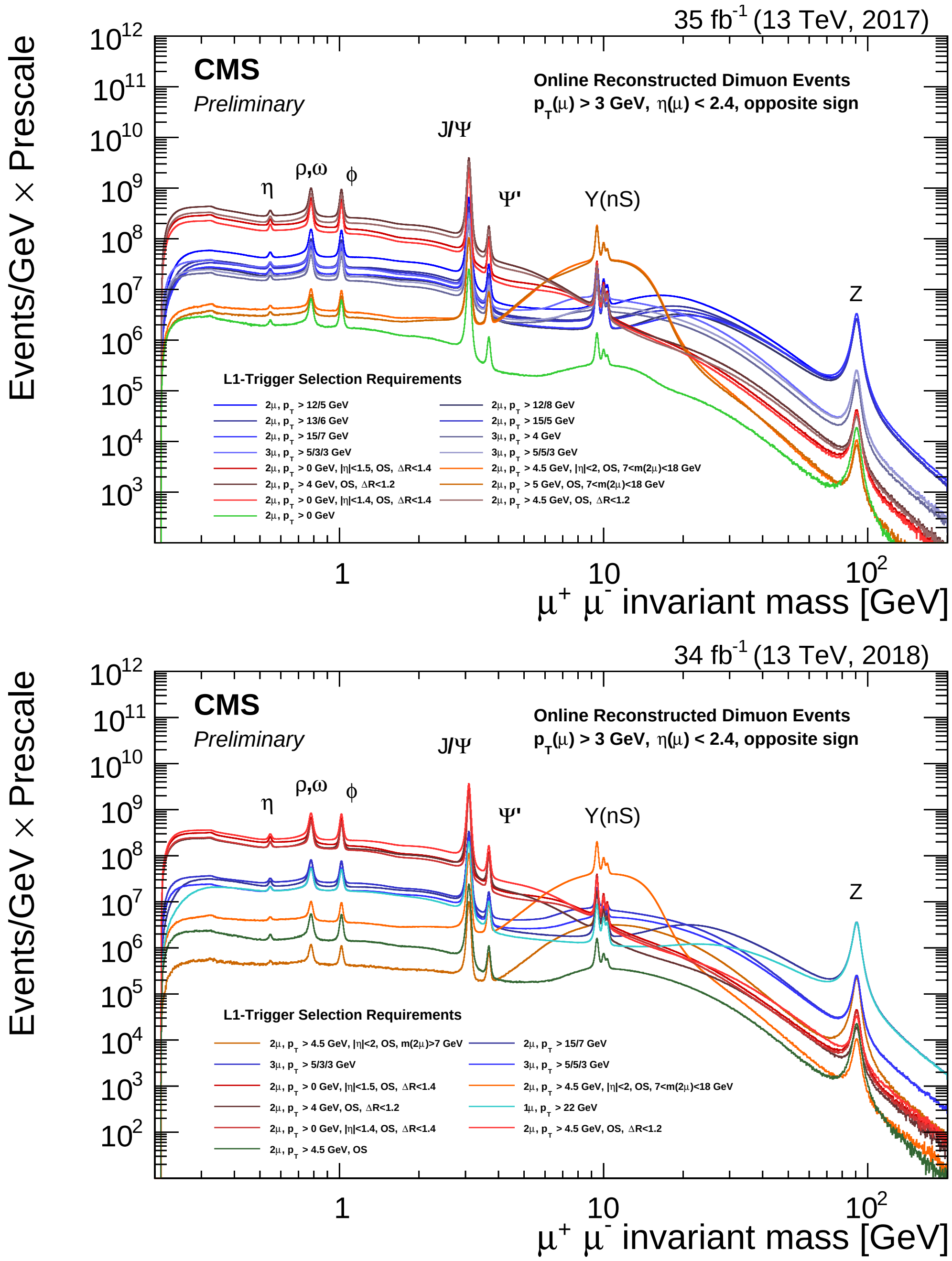

Figure 4·7: Dimuon mass distribution as reconstructed in the HLT and saved in the scouting stream for the full 2017 integrated luminosity (left) and a subset of the 2018 luminosity (right). Different colored lines correspond to the different L1 seeds that may have set off the scouting trigger. Taken from Ref. [68]

4.5 Event Selection

To further enhance the sensitivity of this search, we make use of the available offline information in the scouting data to select for events that are kinematically most like to resonant production, and reject those that are most like continuum or combinatoric background. The simplest of these selections are that both muons must have $p_{\mathrm{T}} >$ 4 GeV, and $|\eta| < 1.9$. The momentum requirement is to ensure that we do not experience turn on effects from the 3 GeV dimuon scouting trigger. The cut on pseudorapidity is to constrain the analysis to parts of the detector where momentum resolution is optimal. It is also required that the event contains two muons of opposite charge that share a PV within 0.2 cm of the beamspot. This ensures that we are searching for a neutral BSM particle that decays promptly into muons.

The previously mentioned CMS scouting search, which targeted masses above 10 GeV, made use of a set of flat cuts on variables relating to the tracks of the muons. These include a minimum on the number of hits required in the tracker and pixel detectors, a maximum on the fitted χ^2 value for each of the reconstructed muon tracks, and the relative isolation of each track, defined as the summed momenta of all tracks within a radius $\Delta R < 0.3$, divided by the momentum of the primary track. While this could be reimplemented for this analysis, it does not perform as well for the case of invariant masses below 10 GeV. In particular, the requirement on the isolation of the two tracks becomes very inefficient at very low masses, as any pair of muons with momentum sufficient to overcome the trigger threshold, originating from a low mass parent, requires the pair to be boosted and thus closer together. Figure 4·8 shows the efficiency of this selection when applied to lower mass J/Ψ events versus higher mass Z boson events. The inefficiency of this selection for low p_{T} muons is a significant tax on the acceptance of our desired signal events. This identifier is used as a benchmark to compare the performance of alternate selections and is referred to as the "legacy cut-based ID."

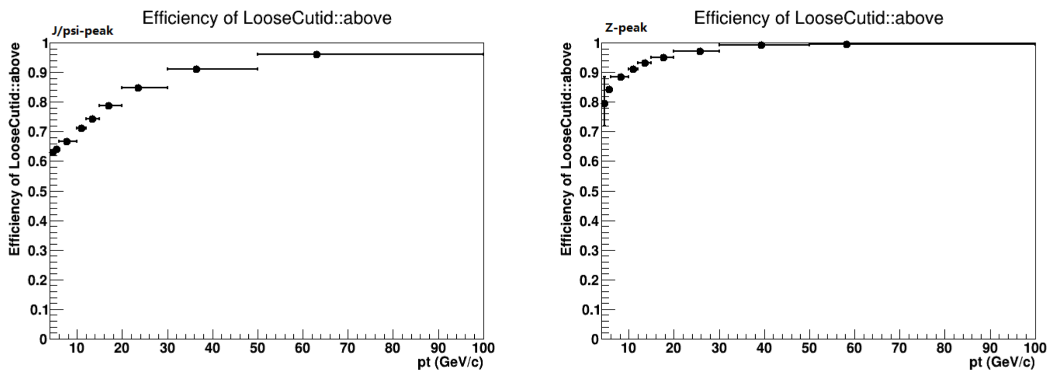

Figure 4·8: Efficiency of a scouting muon to pass the legacy cut-based selection from the higher mass search. The left plot includes low mass events sampled from the J/Ψ peak in data as a function of p_{T}, and the right plot shows events taken from the Z peak.

4.5.1 BDT Classifier

To replace this flat-cut approach, this analysis implements a boosted decision tree (BDT) [73] identifier trained on the SM resonances present in the data, making use of the relevant offline event variables in the scouting dataset. A BDT is a machine-learning technique that can aid in sorting objects into different categories, based on a variety of features. A normal decision tree resembles a flow chart, with a series of nodes that pose a binary selection on some variable in the data, such as "does this event contain two muons with more than 3 pixel hits each?" Depending on how an event meets this classification, it will be directed through the tree on different branching paths, subjecting it to further selections until it is eventually binned as signal-like or background-like. Figure 4·9 (left) shows a toy example of a simple decision tree that might apply to an analysis like this. The precise architecture of the tree, such as the total number of nodes and the threshold values for each classifier can vary to maximize its performance at accurately assigning pre-labelled training data which is known to be signal or background. To accurately incorporate all aspects of the data that lend themselves to classification, a single decision tree will often need to be quite complicated and include many layers of nodes, increasingly subdividing the training data. Complex models are prone to over-training and the architecture of a single tree can easily become finely tuned to specific aspects of the training data, making it hard to assign out-of-sample data. Because of this, a single decision tree is considered to be a weak-learner [74], having high variance and a tendency to misclassify our data. This effect can be mitigated by considering an ensemble of these weak-learners, where the contribution of each to the final decision is weighted by its performance relative to the other learners. This method is called "boosting," and an aggregate classifier that employs it is called a boosted decision tree. This analysis makes use of a particular boosting technique called AdaBoost to construct its selection.

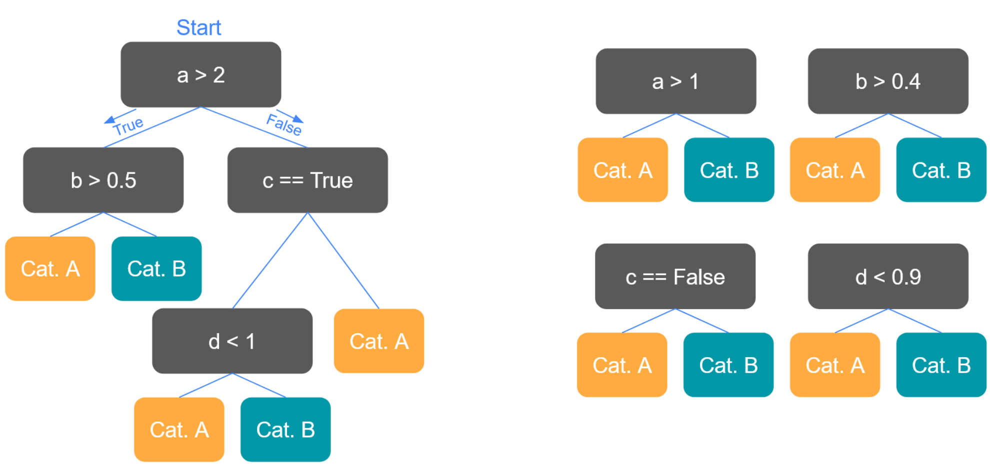

Figure 4·9: Toy examples of decision trees, made to sort data with 4 variables {a, b, c, d} into 2 categories {A, B}. The left element is an ordinary decision tree with 3 layers. The cluster on the right is an ensemble of stumps, akin to those created by the AdaBoost algorithm

AdaBoost, short for "adaptive boost," is a method of assembling a robust aggregate predictor, where weak learners are applied to the model iteratively, with the goal of recovering events that were misclassified by the previous learner [75]. In this approach, the weak learners are single binary decision nodes with no branching paths called "stumps". Figure 4·9 (right) shows an ensemble of such stumps. At the start, the training events are all weighted evenly and the first weak learner to be used is the one that simply minimizes the number of misclassifications. The data is then subjected to a reweighting, where the events that were properly categorized by the first learner are weighted down and those that were improperly categorized are weighted up. The next identifier will minimize the weighted error of this adjusted dataset, then the weights will then be further adjusted based on the events it failed to categorize. This continues ad nauseam, until the improvement in the classifiers performance plateaus. This method acts as a form of self-regularization, and even in scenarios with a high number of weak learners, the model complexity remains low and out of sample performance does not degrade.

Training Samples

The samples used to train our BDT identifier are extracted from the scouting data in regions that will not be used for the analysis. For the signal training sample, we use events with two muons of opposite charge from under the J/Ψ and Υ peaks to model the anticipated features of resonant production, which we expect to be similar to our BSM dimuon production. The SM meson peaks are not in the search region of this analysis, so these training events have no overlap with the actual data. For the combinatoric background training, we select events with two muons of the same charge that have an invariant mass between 1 and 10 GeV. This search does not include signatures for doubly-charged particles, so these events are assumed to be background and are also uncorrelated with the analysis data. We refer to these regions as opposite sign (OS) and same sign (SS) respectively.

To construct these training samples, we use a technique called tag-and-probe

(TnP) [76], commonly used in cases where one wants to measure the performance of some selection criteria by exploiting a two-body decay of a known particle in the data. This is done by scanning the data for events containing a high-quality muon, called a tag, with a selection stringent enough that it is nearly guaranteed to be perfectly efficient for the whatever additional selection we are constructing. In this case the tag requirement is that the muon must have $p_{\mathrm{T}} > 12$ GeV, $|\eta| < 2.4$, and pass the inefficient legacy cut-based ID. Events containing a valid tag muon are then checked to see if they contain exactly one additional muon, the "probe", which is only required to satisfy the analysis' fiducial cuts, $p_{\mathrm{T}} > 4$ GeV, $|\eta| < 1.9$. If the tag and probe muons have the same charge, the event is grouped into the SS sample, and if they have opposite charge, it is grouped into the OS sample. The dimuon mass distributions for both of these samples are shown in Figure 4·10.

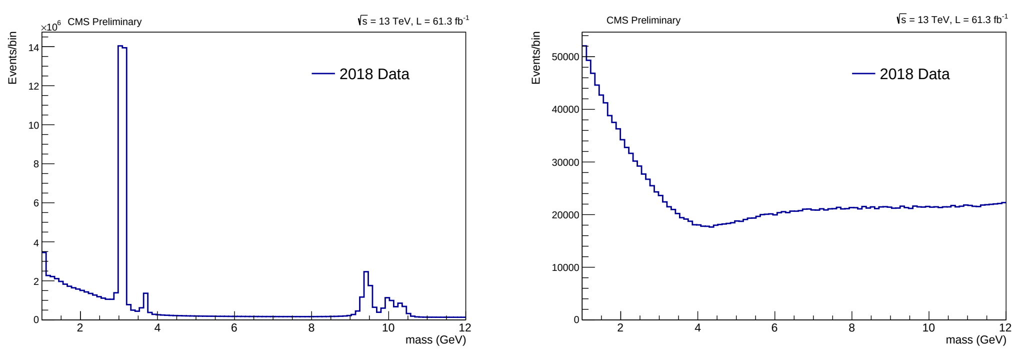

Figure 4·10: Dimuon mass distributions for the opposite sign, signal-like TnP sample (left) and the same sign, background-like TnP sample.

Two different sets of signal events are extracted from the OS data, which will ultimately be used to train a separate BDT for the higher and lower mass ranges we consider. The lower mass training sample is taken from events under the J/Ψ peak in the OS sample ($3.0 < m_{\mu\mu} < 3.2$ GeV). The higher mass sample is taken from events under the Υ peak ($9.2 < m_{\mu\mu} < 9.7$ GeV). The set of probe muons in these events are ideal for modeling our signal when training the BDT, given that they are gathered with minimal bias, having no underlying requirements on their quality, but can nevertheless be inferred to be real signal muons, by their association with a higher-quality muon in the same event and a dimuon invariant mass under a SM peak. Likewise, the SS probes are collected without bias, but are guaranteed to represent undesirable combinatoric background processes by nature of their charge being the same as the tag.

4.5.2 BDT Input Variables

The BDT identifiers will be provided with several of the probe scouting muon variables during training, to maximize its ability to resolve signal events from background. The same variables from the legacy cut-based ID are included, these being the number of hits in the pixel detector (N_{Pixel}), the number of hits in the strip tracking layers (N_{Strip}), the goodness of fit for the muon track (χ^2_μ), and the relative isolation of the muon track (*TrkIso*). Information on the number of hits in the muon system is a potentially useful handle for the BDT, but is excluded due to being improperly saved in the 2017 data, and having ultimately having a small effect on BDT performance in measurements with 2018 data.

Some further variables describing the quality of the dimuon vertex are included, such as the goodness of fit for the vertex (χ^2_{vtx}), the absolute displacement of the vertex from the beamline in the transverse plane (L_{xy}), and the value of the absolute transverse displacement divided by the uncertainty on the measured position ($\sigma_{L_{\text{xy}}}$). Figure 4·11 shows that these variables are mostly uncorrelated with one another in the data, except for some measurable correlation between the number of hits in the pixel and strip trackers, and the near total degeneracy of L_{xy} and $\sigma_{L_{\text{xy}}}$. Because of this strong redundancy, only one of these displacement variables in included in the BDT at a time.

Distributions for all input variables are shown in Figure 4·12 for SM resonances in

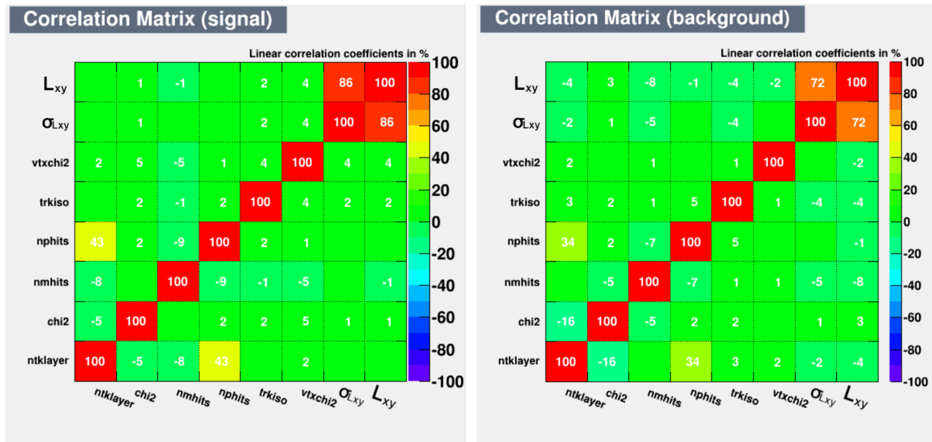

Figure 4·11: Correlation matrix of BDT input variables, taken from the signal sample Υ events (left) and background sample same sign continuum (right).

the data, the SS background sample, as well as the DY simulated sample. A long tail can be seen in the L_{xy} distribution for the J/Ψ, due to its tendency to be produced in displaced vertices from B-meson decays [77]. Because of this, we restrict models trained on the J/Ψ to have no access to this displacement variable, as it is likely to have outsized influence in identifying J/Ψs specifically rather than a generic prompt resonance. Figure 4·13, shows the same variables in the DY sample, subdivided into different mass ranges.

4.5.3 Selection Optimization

For the high mass BDT, many possible configurations are explored to find the selection criteria with the best possible performance in identifying signal events. These include experimenting with various combinations of the training samples and which of the displacement variables (L_{xy} or $\sigma_{L_{\mathrm{xy}}}$) are included in the training. Additional improvement can be found by placing a requirement on the maximum value of L_{xy}, both as a preselection on the signal training sample and a flat cut to be applied in parallel with the BDT. The first round of identifiers consist of training three BDTs

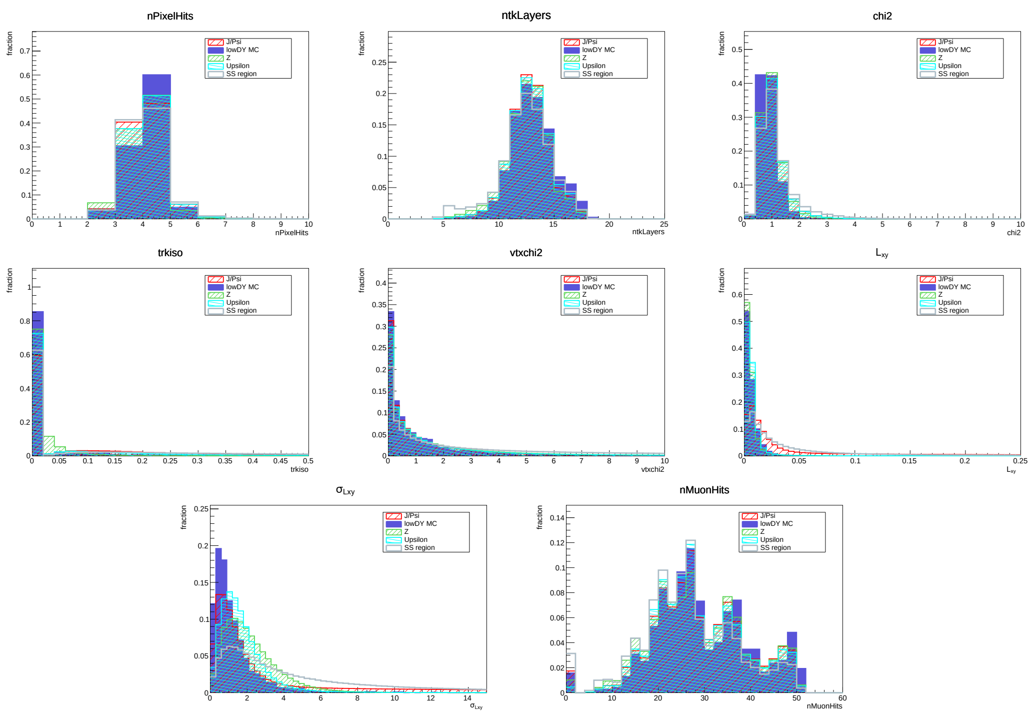

Figure 4·12: Normalized distributions of BDT input variables for the J/Ψ, Υ, and Z peaks in the scouting data, the same sign background training data, and the low mass DY MC sample.

using the Υ signal sample, one which includes neither displacement variable in the training (`Ups1`), one which includes $\sigma_{L_{\mathrm{xy}}}$ (`Ups2`), and one of which includes L_{xy} (`Ups3`). All three configurations are subjected to a flat cut of $L_{\mathrm{xy}} < 0.2$ cm prior to training. Distributions of the BDT discriminant (R_{BDT}) for each of these are shown in Figure 4·14.

The nominal best performance point for each of these identifiers is to require that R_{BDT} is greater than zero for a given muon. However, the sensitivity of the analysis can be further improved by applying another cut on L_{xy} post training, to enhance the purity of prompt signal events in our data. To find the ideal working point for this cut, we define a figure of merit that captures the statistical significance of our signal acceptance versus background rejection, $S/\sqrt{B}$, where S is the number of allowed signal events from the the Υ peak in data and B is the number of background

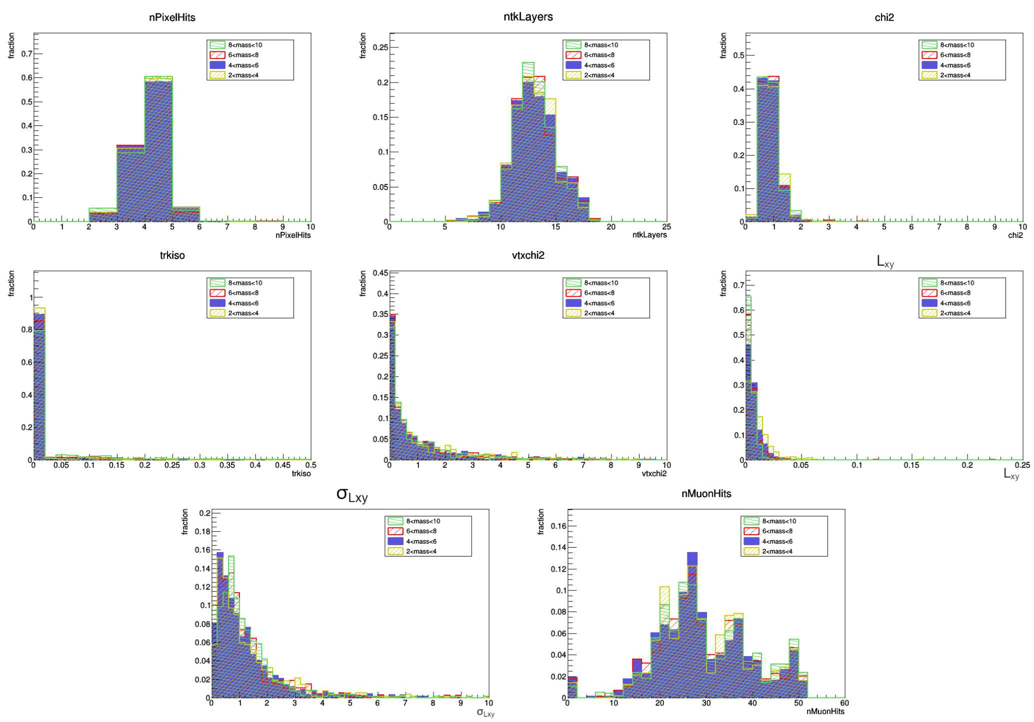

Figure 4·13: Normalized distributions of BDT input variables for different mass ranges of the low mass DY MC sample.

events accepted from the dimuon continuum in the data. This sensitivity metric will enable direct comparisons in the performance of different working points and allow us to gauge of the final enhancement of our new selection relative to the legacy cut-based ID. The optimal performance point for each of these identifiers is found by measuring this sensitivity across a two-dimensional scan of different cuts on L_{xy} and BDT response, summarized in Figure 4·15. The overall best performing of these is `Ups3` with the requirement $R_{\mathrm{BDT}} > 0.0$ and $L_{\mathrm{xy}} < 0.015$, which has $\sim 75\%$ acceptance for Υ signal events, and only $\sim 20\%$ acceptance for sideband background events.

Though this selection already provides a significant enhancement in sensitivity relative to the cut based ID, still further improvements can be found by tightening the displacement cut applied before the BDT training. Two additional BDTs are trained and optimized using the same inputs as `Ups3` with stronger displacement preselection,

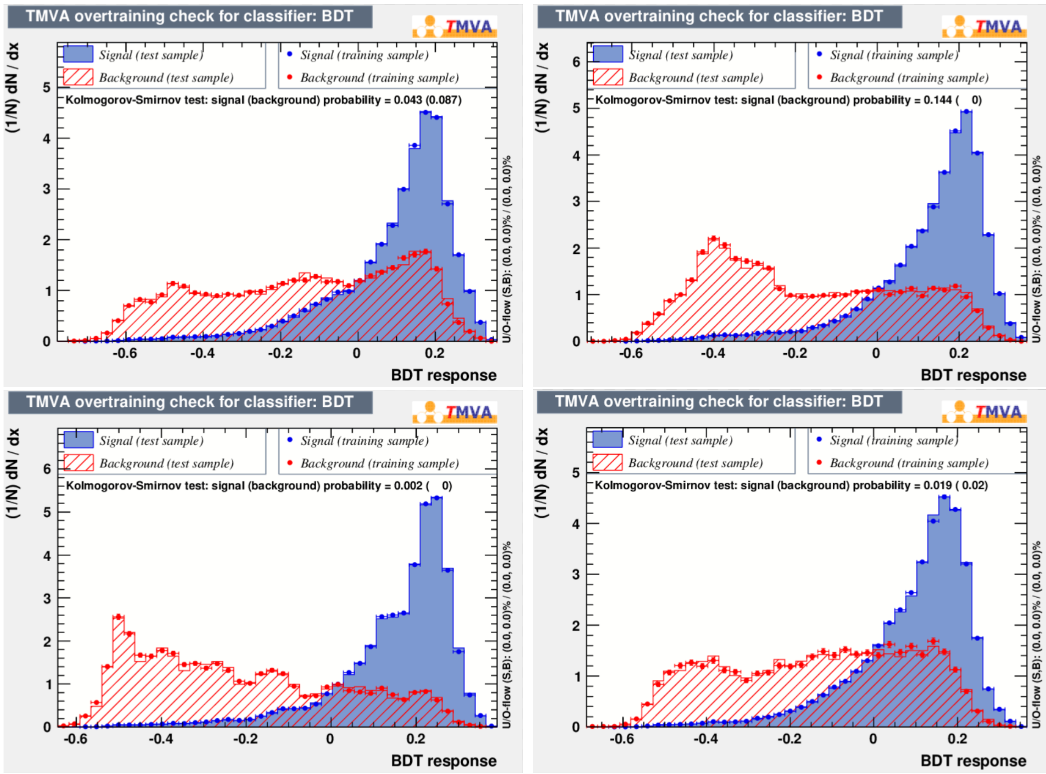

Figure 4·14: Distributions of R_{BDT} for the Υ-trained identifiers `Ups1` (top left), `Ups2` (top right), `Ups3` (bottom left) and `Ups5` (bottom right).

$L_{\mathrm{xy}} < 0.05$ (`Ups4`) and $L_{\mathrm{xy}} < 0.03$ (`Ups5`). Of all the considered possibilities, the highest average sensitivity is found with `Ups5` with $R_{\mathrm{BDT}} > 0$ and $L_{\mathrm{xy}} < 0.015$ as the post training working points. This combination, henceforth referred to as simply `Ups5`, is the new benchmark offline muon requirement employed for this analysis, with an Υ signal acceptance of $\sim 68\%$ and a background acceptance of only $\sim 16\%$. Figure 4·14 (lower right) shows the BDT output distribution for this ID and Figure 4·15 shows the parameter scan optimization.

Additional BDT configurations are were trained and optimized using alternate signal samples for the training. The OS J/Ψ sample was used in two configurations, one with no displacement variables used in the training (`JPsi1`), and one with $\sigma_{L_{\mathrm{xy}}}$ included (`JPsi2`). The DY signal MC sample was also used to train two configura-

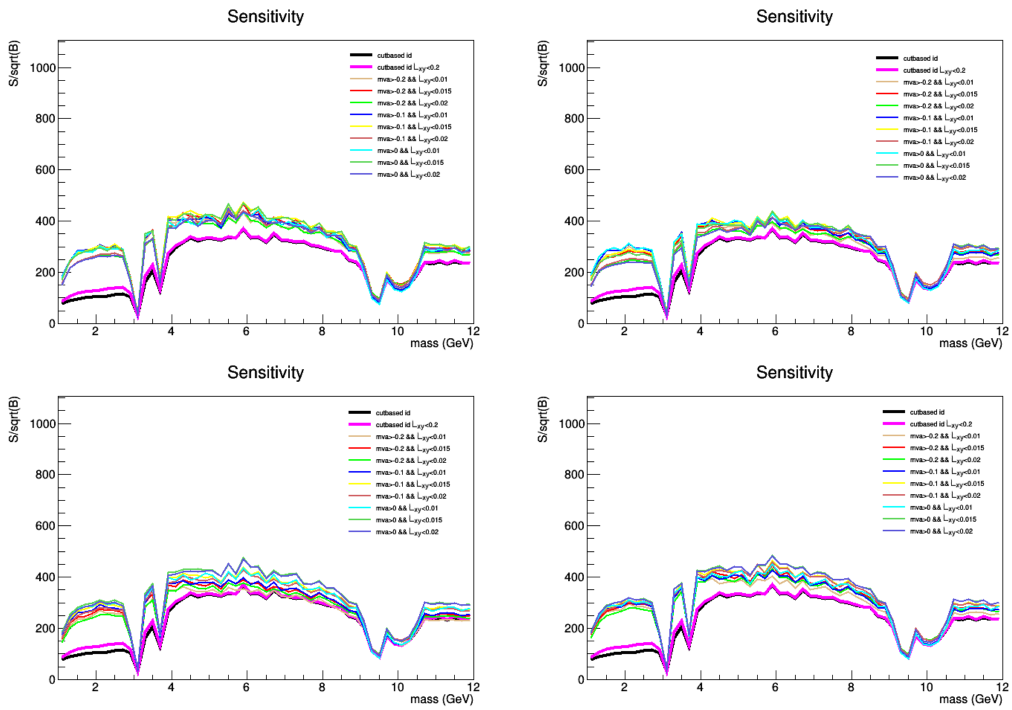

Figure 4·15: Sensitivity scans of L_{xy} and BDT response for `Ups1` (top left), `Ups2` (top right), `Ups3` (bottom left), and the best performing identifier `Ups5` (bottom right). The black line shows the performance of the legacy cut-based ID. The magenta shows the same selection with the displacement cut $L_{\mathrm{xy}} < 0.2$ cm, which does not substantially improve its performance.

tions, one without displacement information (`DY1`) and one with $\sigma_{L_{\mathrm{xy}}}$ (`DY2`). Even after optimizing the 2D scan optimization, none of these selections surpassed the sensitivity of `Ups5`, as shown in Figure 4·16. The inputs and best selections for all of the discussed muon identifications are summarized in Table 4.4.

4.5.4 Low Mass Optimization

Though this selection has a substantial improvement in performance across all mass hypotheses relative to the cut-based ID, its performance is comparatively degraded below ~ 4 GeV due to a broadening distribution of L_{xy} in the data, as can be seen on the yellow line of the L_{xy} panel in Figure 4·13 (center row, right). One expects that

Name	Signal input	Vertex preselection	Disp. variable	Optimized working point
`Ups1`	Υ-data	$L_{\mathrm{xy}} < 0.2$ cm	—	$R_{\mathrm{BDT}} > -0.1$, $L_{\mathrm{xy}} < 0.01$ cm
`Ups2`	Υ-data	$L_{\mathrm{xy}} < 0.2$ cm	$\sigma_{L_{\mathrm{xy}}}$	$R_{\mathrm{BDT}} > 0.0$, $L_{\mathrm{xy}} < 0.01$ cm
`Ups3`	Υ-data	$L_{\mathrm{xy}} < 0.2$ cm	L_{xy}	$R_{\mathrm{BDT}} > 0.0$, $L_{\mathrm{xy}} < 0.015$ cm
`Ups4`	Υ-data	$L_{\mathrm{xy}} < 0.05$ cm	L_{xy}	$R_{\mathrm{BDT}} > 0.0$, $L_{\mathrm{xy}} < 0.015$ cm
$\star$ `Ups5`	Υ-data	$L_{\mathrm{xy}} < 0.03$ cm	L_{xy}	$R_{\mathrm{BDT}} > 0.0$, $L_{\mathrm{xy}} < 0.015$ cm
`JPsi1`	J/Ψ-data	$L_{\mathrm{xy}} < 0.2$ cm	—	$R_{\mathrm{BDT}} > -0.1$, $L_{\mathrm{xy}} < 0.01$ cm
`JPsi2`	J/Ψ-data	$L_{\mathrm{xy}} < 0.2$ cm	$\sigma_{L_{\mathrm{xy}}}$	$R_{\mathrm{BDT}} > -0.1$, $L_{\mathrm{xy}} < 0.01$ cm
$\star$ `JPsi3`	J/Ψ-data	$L_{\mathrm{xy}} < 0.015$ cm	—	$R_{\mathrm{BDT}} > -0.1$, $\sigma_{L_{\mathrm{xy}}} < 3.5$
`DY1`	DY MC	$L_{\mathrm{xy}} < 0.2$ cm	—	$R_{\mathrm{BDT}} > -0.1$, $L_{\mathrm{xy}} < 0.01$ cm
`DY2`	DY MC	$L_{\mathrm{xy}} < 0.2$ cm	$\sigma_{L_{\mathrm{xy}}}$	$R_{\mathrm{BDT}} > -0.2$, $L_{\mathrm{xy}} < 0.01$ cm

Table 4.4: Summary of BDT inputs and best post training working points for all considered classifiers. Stars indicate best performing IDs.

the experiments vertex resolution will worsen at lower masses, as the dimuon pairs are more boosted and the intersection of their tracks becomes harder to identify. The tight requirements on L_{xy} included in the `Ups5` selection becomes very inefficient, and removes a high number of desirable signal candidate events in the search region lower than the J/Ψ mass in the data.

To improve the sensitivity further, a third BDT is trained with the J/Ψ sample as the signal input, using neither displacement variable, with a very tight preselection, $L_{\mathrm{xy}} < 0.015$ cm (`JPsi3`). This is to ensure that the BDT will learn only features associated with prompt signal production, even though it cannot use displacement information as a discriminant. The effect of this cut on the J/Ψ input is shown in Figure 4·17. The displaced contribution to the training sample is well fit to an exponential, and the majority of desirable prompt events fall below the $L_{\mathrm{xy}} < 0.015$ cm requirement. Post training, this BDT is optimized similar to the others except

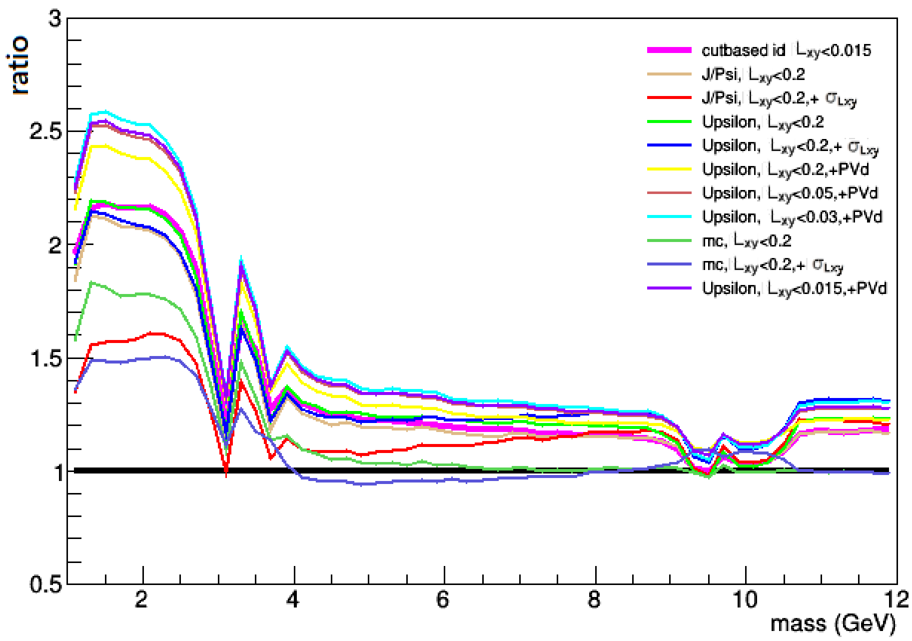

Figure 4·16: Sensitivity comparison of all optimized muon identifications. All curves are rescaled to show their improvement in sensitivity relative to the legacy cut-based ID, which has been scaled to 1 at all mass points. The magenta line shows the sensitivity improvement of the cut-based ID when subjected to an additional $L_{\mathrm{xy}} < 0.2$ cm requirement. `Ups5` (cyan) visibly provides the best sensitivity at all masses of interest below 10 GeV.

that it is scanned over different possible values of $\sigma_{L_{\mathrm{xy}}}$, rather than Lxy. Cuts on $\sigma_{L_{\mathrm{xy}}}$ are not as inefficient as on Lxy at low masses, but can still help to improve sensitivity. The best performing working point is found to be $R_{\mathrm{BDT}} > -0.1$ and $\sigma_{L_{\mathrm{xy}}} < 3.5$.

The relative sensitivity of both `Ups5` and `JPsi3` to the cut-based ID is shown in Figure 4·18. Both identifiers provide substantial improvements over the legacy working point for all masses, and have further enhanced strength in complementary mass regions above and below ~ 4 GeV. Applying these selections to the full scouting dataset gives the complete dimuon mass distributions that will serve as the inputs for this search. Figure 4·19 shows these distributions for both classifiers. To maximize

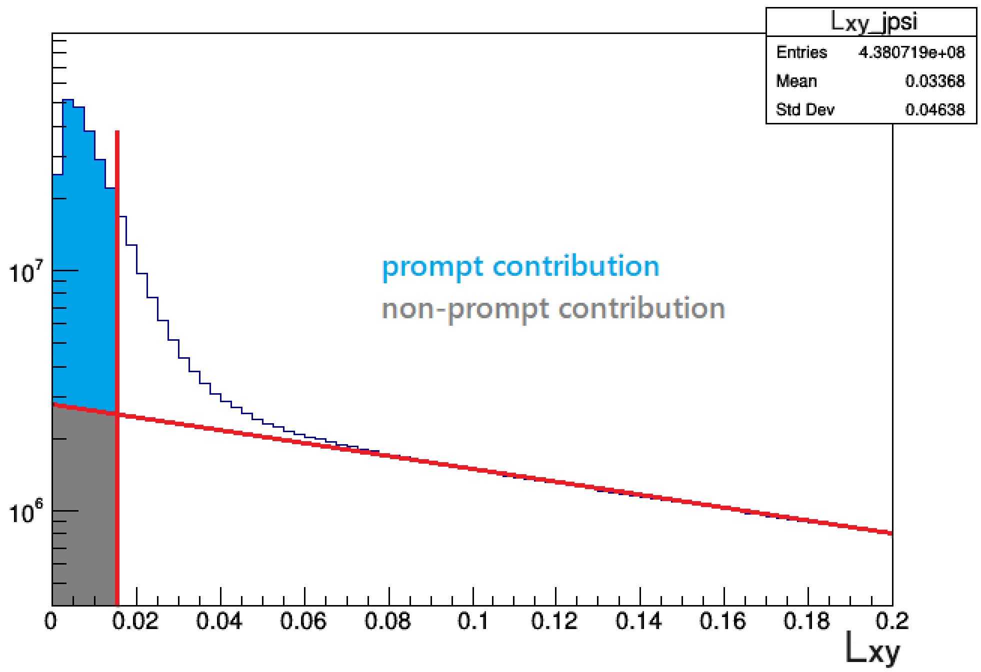

Figure 4·17: Distribution of L_{xy} for events with dimuon mass beneath the J/Ψ peak in the OS sample. The slanted red line shows the exponential fit that models the fraction of J/Ψs from non-prompt sources, and the vertical red line corresponds to the preselection applied to the sample prior to training the `JPsi3` BDT. The non-prompt fraction remaining in the training sample is $< 10\%$ (shaded gray).

our sensitivity to new physics signatures, the analysis is broken into two categories where the distribution derived from `JPsi3` is used as the input below the J/Ψ peak, and `Ups5` is used above it. Because no limit is set on the SM resonances, the J/Ψ and Ψ' peaks in the data form a natural break in the analysis, so no consideration needs to be paid to the precise mass point where the optimal selection changes.

4.5.5 Scalar Signal Optimization

The main result of this search is a model independent limit on BSM production, with as broad of a fiducial acceptance as is practical. However, additional sensitivity can be found for by assuming the particle is likely to be produced by some specific

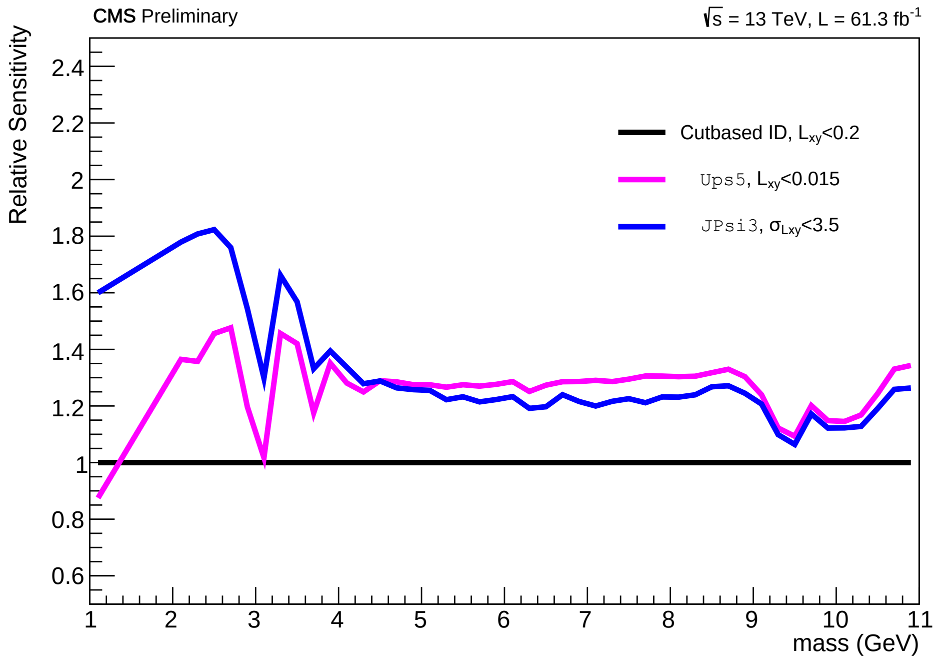

Figure 4·18: Sensitivity comparison of two best performing muon classifiers. Curves are rescaled to show their improvement in sensitivity relative to the legacy cut-based ID, which has been scaled to 1 at all mass points.

mechanism and refining the analysis cuts to purify the signal acceptance of that particular event topology. For the case of observing a pseudoscalar in a 2HDM+S scenario, we expect the leading production process to be ggF. Gluons carry twice the color charge of quarks, making it significantly more likely that one of the incoming partons will emit an ISR jet, causing the pseudoscalar and resulting dimuon pair to recoil away from the jet with potentially significant momentum. This enhanced likelihood of boosting is not mirrored in the $q\bar{q}$ mediated continuum background, so it is possible to reject a large fraction of background events without significantly taxing our signal acceptance, by applying a cut on the net transverse momentum of the combined dimuon pair ($p_{\mathrm{T}_{\mu\mu}}$). Figure 4·20 compares the $p_{\mathrm{T}_{\mu\mu}}$ distributions between data that passes our nominal event selections, and pseudoscalar signal Monte-Carlo

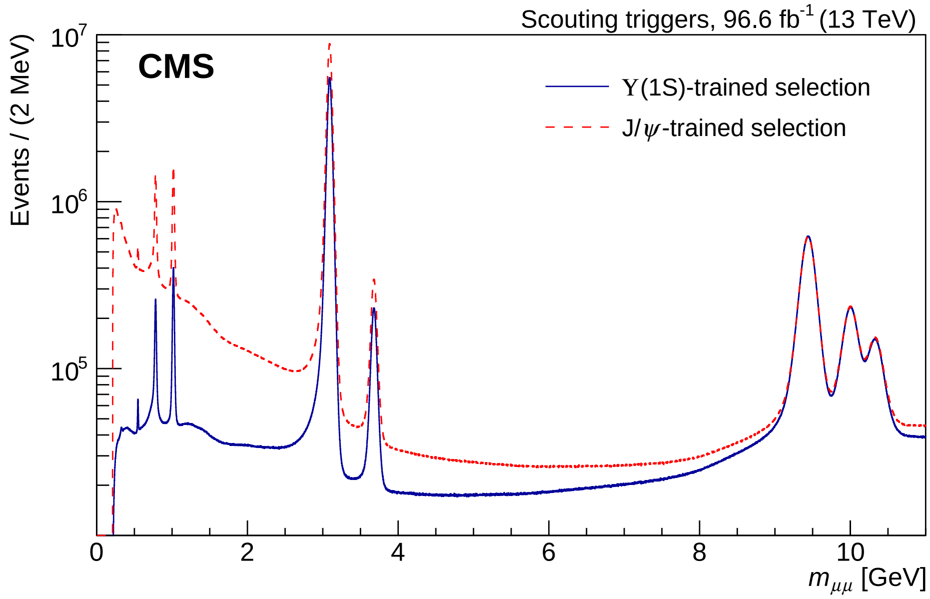

Figure 4·19: Histogram of all dimuon events passing the dimuon scouting trigger and the best Υ-trained selection `Ups5` (blue), and the best J/Ψtrained selection `JPsi3` (red).

(MC) at a range of different masses, and illustrates the strong tendency for the dimuon pairs originating from a pseudoscalar to be boosted.

This shift in the dimuon pair momentum also effects the momentum distributions of the individual muons. Figure 4·21 shows the momentum distribution for the subleading muons in data and pseudoscalar MC. Again the distribution is broadened relative to the data, indicating the potential to enhance our sensitivity by tightening the p_{T} requirement. From here the analysis is split for into two categories; the first is called the "inclusive selection," which is formed by all the cuts and selections mentioned prior to this section and will be applied for the constraints on dark photon production. The other is referred to as the "boosted selection," which will be further refined by cuts on the muon and dimuon momenta, and will be used to best constrain the 2HDM+S pseudoscalar production.

The optimal working point for the p_{T} and $p_{\mathrm{T}_{\mu\mu}}$ cuts in the boosted category

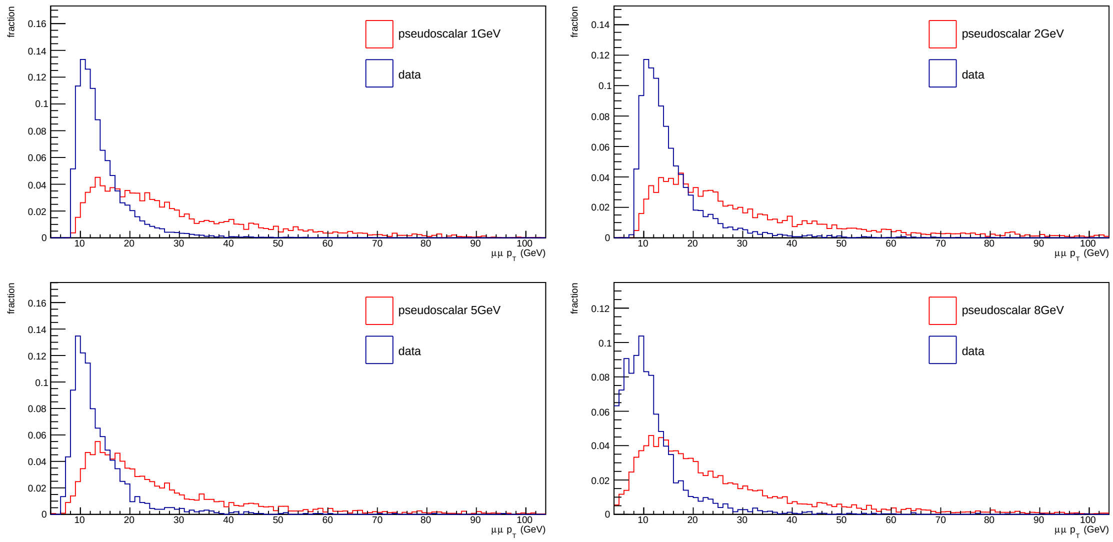

Figure 4·20: Normalized distributions of $p_{\mathrm{T}_{\mu\mu}}$ for pseudoscalar signal samples with a mass of 1 GeV (top left), 2 GeV (top right), 5 GeV (bottom left), 8 GeV (bottom right), compared with the same variable in data sampled from a mass range close to the pseudoscalar mass.

are again determined by a two-dimensional scan across both variables, to determine which values provide the maximum signal sensitivity $S/\sqrt{B}$. For this measurement, the signal events considered are the pseudoscalar samples shown in the muon and dimuon momentum distributions, and the background is the data from the inclusive selection that they were contrasted against. These two-dimensional scans are shown in Figure 4·22 for representative mass points at 1, 2, 5, and 8 GeV. The bottom left corner of each of these plots is close to the baseline sensitivity of the inclusive selection, and the z-axis shows the factor by which that combination of cuts enhanced the pseudoscalar sensitivity relative to that benchmark value. The chosen working points for the boosted category are: $p_{\mathrm{T}} > 5$ GeV for both muons, and $p_{\mathrm{T}_{\mu\mu}} > 35$ (20) GeV for masses below (above) 4 GeV.

Though these cuts successfully improve the sensitivity for the pseudoscalar result at all masses, the underlying selection used for the higher mass hypotheses ($\texttt{Ups5} + L_{\mathrm{xy}} < 0.015$) becomes less efficient when applied to events with the boosted

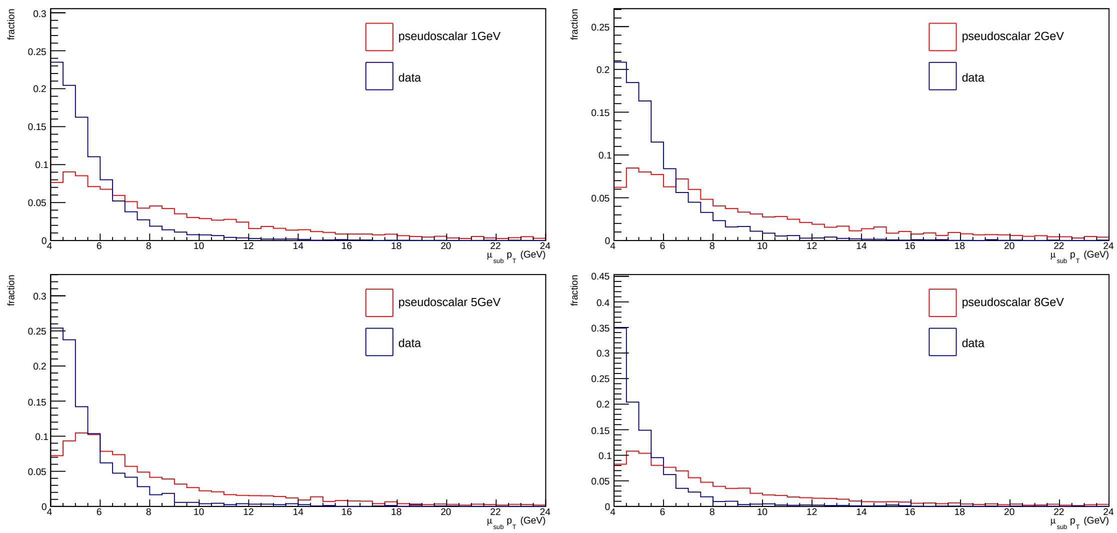

Figure 4·21: Normalized distributions of subleading muon p_{T} for pseudoscalar signal samples with a mass of 1 GeV (top left), 2 GeV (top right), 5 GeV (bottom left), 8 GeV (bottom right), compared with the same variable in data sampled from a mass range close to the pseudoscalar mass.

requirements. Boosted dimuons have worse vertex reconstruction due to the tracks being more parallel, and the L_{xy} cut becomes suboptimal. This is the same effect that motivated the creation of the low mass ID ($\texttt{JPsi3} + \sigma_{L_{\mathrm{xy}}} < 3.5$) in the inclusive selection. Since the J/Ψ-trained selection favors boosted topologies, the sensitivity above 4 GeV is further improved by applying this selection for all masses in the boosted category, in lieu of the Υ-trained ID and flat displacement cut.

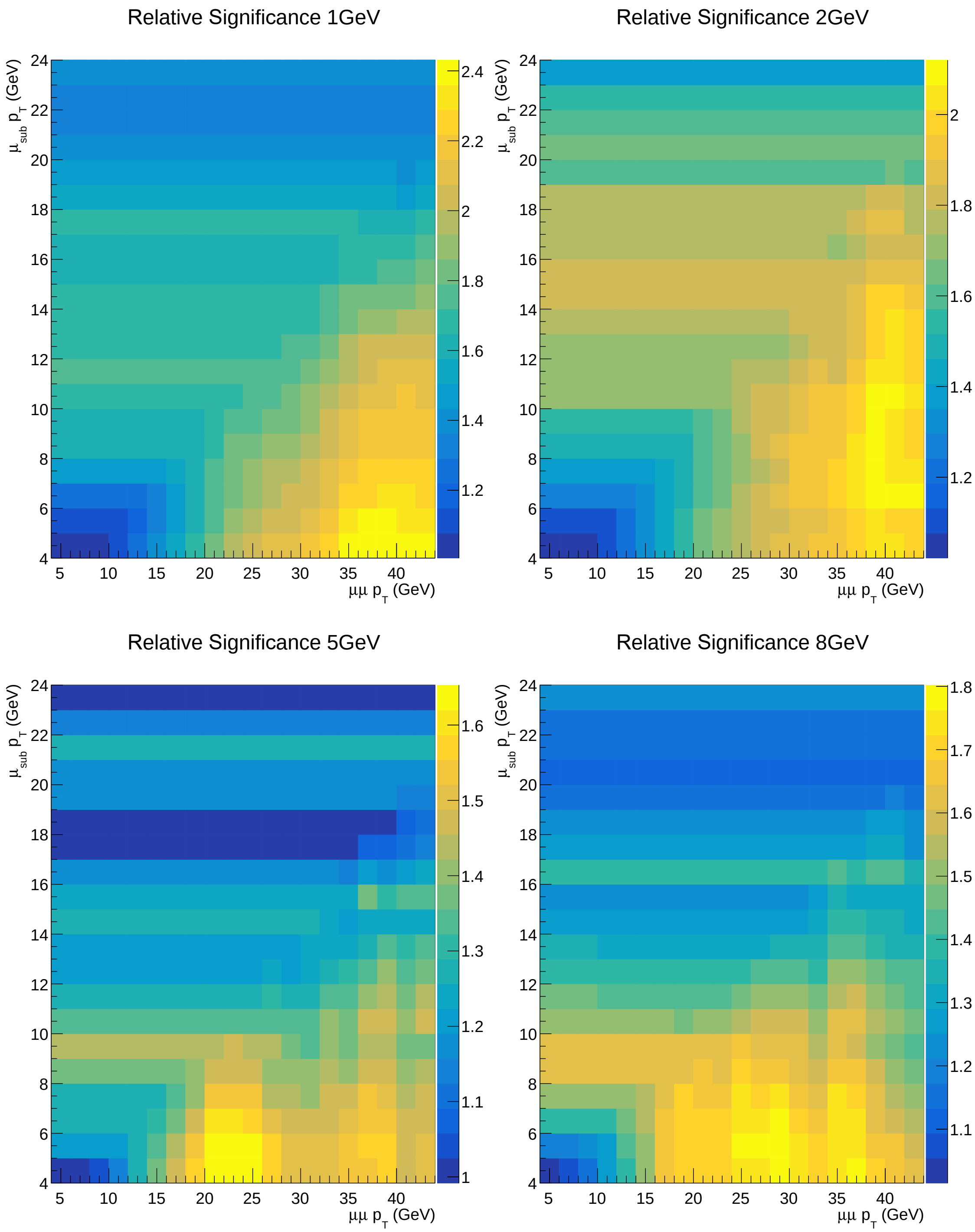

Figure 4·22: Relative sensitivity enhancement for pseudoscalar signals at masses of 1 GeV (top left), 2 GeV (top right), 5 GeV (bottom left), 8 GeV (bottom right), as a function of $p_{\mathrm{T}_{\mu\mu}}$ and subleading muon p_{T} requirement.

All event selection criteria described in this section are summarized in Table 4.5. The full analysis now has 4 distinct input categories; the boosted and inclusive selections, each of which is subdivided into a high-mass and low-mass selection.

Preselection	$L_{\text{xy}} < 0.2$ cm, $\lvert\eta_\mu\rvert < 1.9$, $N_\mu \geq 2$, Opposite sign			
Signal	Inclusive		Boosted	
Selection	$m_{\mu\mu} < 4$ GeV	$m_{\mu\mu} > 4$ GeV	$m_{\mu\mu} < 4$GeV	$m_{\mu\mu} > 4$ GeV
$p_{\text{T}\mu}$	> 4 GeV	> 4 GeV	> 5 GeV	> 5 GeV
BDT Score	`JPsi3` > -0.1	`Ups5` > 0	`JPsi3` > -0.1	`JPsi3` > -0.1
Vertex cut	$\sigma_{L_{\text{xy}}} < 3.5$	$L_{\text{xy}} < 0.015$	$\sigma_{L_{\text{xy}}} < 3.5$	$\sigma_{L_{\text{xy}}} < 3.5$
$p_{\text{T}\mu\mu}$	—	—	> 35 GeV	> 20 GeV

Table 4.5: Summary of optimized selections for an event which had passed dimuon scouting trigger to be included in the analysis.

4.6 Efficiency Measurements

We will ultimately report our limit in terms of the model independent cross section to produce a BSM mediator (X), by relating the number of measured signal events in the data to the total integrated luminosity included in the analysis. The most naive form of this relationship, assuming 100% branching fraction to muons, is:

$$\sigma_{pp \to X} = N/\mathcal{L}, \tag{4.9}$$

where N is the number of events under the post-fit signal shape, and $\mathcal{L}$ is the integrated luminosity. This expression does not account for the fact that some fraction of BSM dimuon events will not be successfully identified by our offline selection, recorded by the scouting trigger, or even land within in the experiment volume of the detector. It is also likely that the BSM mediator will decay to other particles in addition to muons, further taxing the fraction of BSM production that we can measure in this search. Each of these effects weakens the extent to which we can identify and constrain BSM production, because for any measured value of the signal yield we must assume a much higher production cross-section to offset the new physics events that we did not record or select for. The corrected form of the model independent value of the cross section, which includes all of these factors is:

$$\sigma_{pp \to X} = \frac{N}{\mathcal{B}\ A\ \epsilon\ \mathcal{L}}, \tag{4.10}$$

where $\mathcal{B}$ is the branching fraction for the BSM particle to decay to muons, A (acceptance) is the fraction of final state muons that land within our effective experimental volume, and ϵ is the fraction of muon pairs within the fiducial volume which will also successfully activate the scouting trigger and pass the offline requirements. When considering our measurement, the first two of these cannot be determined without making some sort of assumption about the underlying model creating the new particle; branching fraction calculations require definite statements on the field structure and couplings of the BSM sector, and acceptance measurements require studies of the final state muon kinematics which will vary greatly depending of the type of mediator being considered. To maintain model independence, we move these values to the left of this expression, treating them as unknown parameters alongside the production cross section:

$$\sigma_{pp\to X}\ \mathcal{B}\ A = \frac{N}{\epsilon\ \mathcal{L}} \tag{4.11}$$

The remaining factor in the right hand side is the efficiency of events which satisfy the baseline fiducial requirements to enter the analysis ($p_{\mathrm{T}} > 4 GeV$ and $|\eta| < 1.9$ for the inclusive category; tightened to $p_{\mathrm{T}} > 5 GeV$ and $p_{\mathrm{T}_{\mu\mu}} > 20(35) GeV$ for the boosted category), to also pass both the dimuon scouting trigger and the offline BDT based selections. This can be determined through a variety of data-driven methods, which will be detailed in this section.

4.6.1 Measurement Strategy

Because the trigger and offline selections are applied to the data sequentially, it is reasonable to interpret the efficiency of a good signal event to pass both criteria as the efficiency to the product of the efficiency of the signal muon to pass the scouting trigger, times the efficiency of an event passing the trigger to then also satisfy the offline selection:

$$\epsilon(\text{Trigger+Selection}|\mathit{signal}) = \epsilon(\text{Trigger}|\mathit{signal}) \times \epsilon(\text{Selection}|\text{Trigger}), \tag{4.12}$$

where *signal* represents a hypothetical set of dimuon events that can be known to satisfy our fiducial cuts with near certainty. This approach allows us to measure these two values separately, by studying the behavior of the trigger relative to a sample of unbiased dimuon events in the data, and measuring the effect of the offline selection using events that have already passed the scouting trigger.

4.6.2 Trigger Efficiency

To measure the efficiency of the trigger, we must first identify a sample of dimuon events in the data that we can use as a proxy for "true" signal events. One strategy to do this is the orthogonal trigger method, wherein we exploit a different CMS dataset, built of events that do not trigger on dimuon pairs, or any muon information at all. Though not explicitly triggered for, this dataset will incidentally include some number of events which contain two muons emerging from a process similar to the one we are hoping to study, in addition to the primary triggering objects. Further, a CMS event is saved with the full list of L1 seeds and HLT paths that the event satisfied, including triggers entirely unrelated to the dataset it is saved to. By selecting one of these datasets, and probing for events that happen to contain two reconstructed muons that meet the baseline fiducial requirements, we can count the number of these events that activate and do not activate the dimuon scouting trigger. Because the fiducial cuts require the offline muons to have momentum above the trigger threshold, the only reason one of these events would not have activated the trigger, is that the trigger path in question failed to identify event. We can therefore define the efficiency of the trigger as the ratio of these unbiased dimuon events that passed the trigger, to the total number of dimuon events in the sample:

$$\epsilon(\text{Trigger}|signal) = \frac{N_{pass}}{N_{pass} + N_{fail}}. \tag{4.13}$$

We expect that the efficiency of the trigger will vary at different dimuon masses, and bin this measurement as a function of mass so that it can be applied to the appropriate mass hypotheses. To claim that that this approach is truly unbiased and that events in this sample could only have failed to activate the scouting path due to the intrinsic inefficiency of the trigger, requires near perfect confidence in the CMS offline software to reconstruct and identify a muon in the data. If the definition used to select these muons is itself inefficient, then our measurement will become a convolved efficiency, including both trigger and offline effects. Dedicated CMS working groups define benchmark working points to label muons in the data according to the quality of the track, and provide performance studies for each of these criteria. Figure 4·23 shows the efficiency of muons to pass an offline working point called "Medium ID," [78] which represents a middle point between higher quality yet less efficient muon criteria, and lower quality but more inclusive criteria. The measurement shows that even at the lowest momentum considered for this search, the efficiency of this criteria is upwards of 90%, therefore, using it to identify muons for our efficiency measurement will not significantly bias our input.

The typical physics processes creating the muons in the orthogonal dataset may not be the same as those contributing to the dimuon scouting data, and it is possible that the measured efficiency at a given mass value will be a poor descriptor of the actual behavior of the scouting trigger. This would be the case if, for example, the orthogonal dataset characteristically produced dimuons with a larger angular separation than an unbiased sample of DY events. To offset this possibility, we can bin our efficiency in multiple kinematic parameters and weigh the contribution of each bin to the total efficiency at a given mass according to the distribution of that variable in a DY simulation, thereby mapping the kinematics of the orthogonal dataset onto our desired signal process.

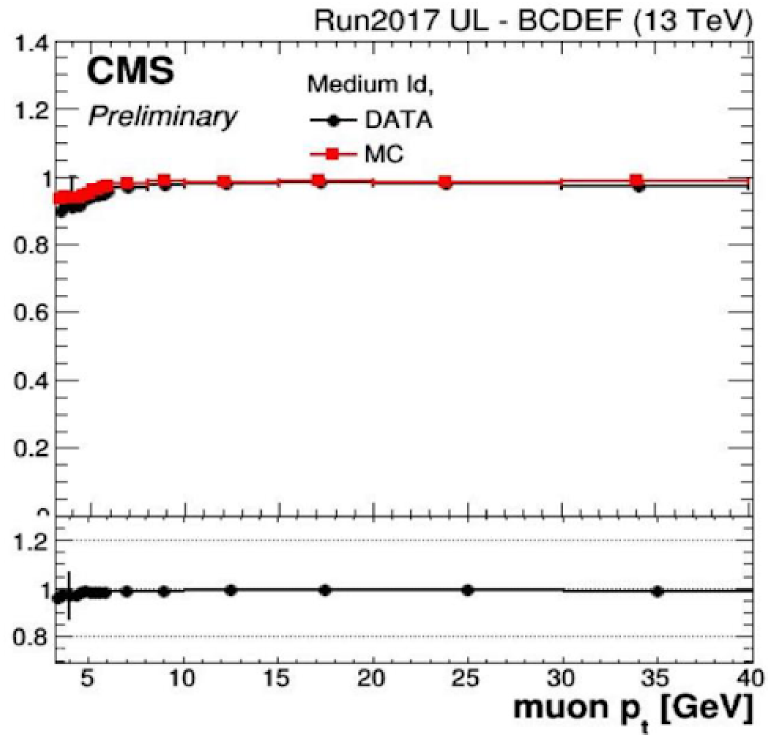

Figure 4·23: Efficiency of CMS Medium ID working point as a function of muon p_{T} in data and simulation.

When considering which CMS dataset to use for sampling these unbiased events, the most important quality is that it contains many dimuon events in our kinematic region of interest. Datasets filled with events that trigger on jets or large missing energy in the detector include many useable muons, but the p_{T} threshold for the primary physics objects are in the range of many hundreds of GeV. If the dimuon pairs we are hoping to find in these datasets have recoiled off of such a hard physics object, then they will likely be extremely boosted, to the point that there will be few available dimuon events representing the kinematics of our desired process. Datasets which are filled by triggers of large deposits in the ECAL have significantly lower momentum thresholds, and are therefore more likely to contain muons with similar kinematics to our signal. Any lingering discrepancy due to the particular physics processes that produce the events in these datasets, will be addressed by the previously mentioned reweighting. The names and momentum thresholds of the chosen CMS datasets are `SingleElectron` ($p_{\mathrm{T}} > 30$ GeV) and `DoubleEG` ($p_{\mathrm{T}} > 20/10$ GeV) for 2017 data and `EGamma` ($p_{\mathrm{T}} > 30$ GeV) for 2018 data.

The first measurement to be performed is a validation of the Medium ID working point as a proxy for true signal muons in the data. This can be done by comparing the efficiency of the scouting triggers in data with respect to Medium ID, to the efficiency in generator level MC which has not been subjected to a Medium ID requirement. If Medium ID is in fact a low bias way of identifying signal muons in the data, these efficiencies should be very similar. We expect some mismodeling of the trigger behavior in the MC sample, which is accounted for by correcting the MC efficiency according to the scale factor:

$$S.F. = \frac{\epsilon(\text{Data}|\text{Medium ID})}{\epsilon(\text{MC}|\text{Medium ID})}. \tag{4.14}$$

Because the Medium ID requirement is applied in both the numerator and denominator, any underlying inefficiency of the selection should cancel out and the MC efficiency is scaled to what is effectively a truth level representation of the efficiency in data. Any remaining disagreement between the data and rescaled truth MC efficiencies can be attributed to the inefficiency of the Medium ID cut applied to the data. Figure 4·24 shows the comparison of these two measurements using 2018 data and the low mass DY signal sample. The two efficiencies are in very close agreement for most bins, with predictable deviations in the lower p_{T} bins where Medium ID is less efficient. These effects are negligibly small compared to the to uncertainty range that will eventually be prescribed to the efficiency measurement.

The analysis will employ the trigger efficiency as measured in the data only, but understanding the relationship between the efficiency measured in data and non-rescaled MC is a good handle to gauge the reliability of this measurement technique.

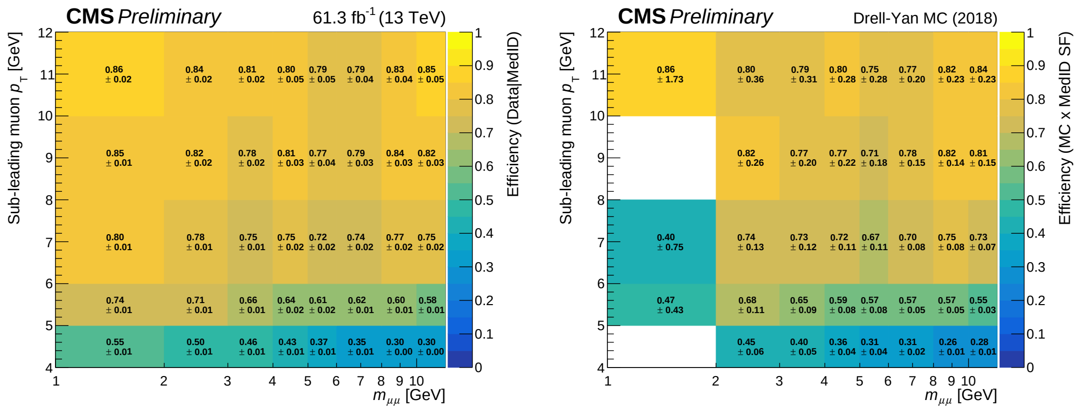

Figure 4·24: Efficiency of dimuon scouting trigger in 2018 data (left) and rescaled DY simulation (right), binned by dimuon invariant mass and sub-leading muon p_{T}.

We will define the uncertainty on the trigger efficiency at a given mass to be percent difference between the data and MC efficiencies. This is a very conservative estimate for the uncertainty on this quantity, but accounts for the "worst-case" scenario, where the events in the orthogonal dataset are not at all representative of the actual signal process we are trying to measure. Figure 4·25 shows that the difference in efficiency between data and simulation is nearly flat as a function of ΔR, indicating that whatever mismodeling in the simulation causes the discrepancy is independent of the muon separation. Limited events available in the 2017 MC causes large statistical error bars and sparsely populated bins, which propagates to all future plots that include this sample.

We can extend this plot to a second dimension to understand how the behavior of the trigger depends on the mass of the dimuon parent particle. Figure 4·26 shows the efficiency of the trigger in data and MC, for both years, plotted with respect to ΔR and dimuon mass. Predictably, there are no events in the low mass bins with large ΔR, as this would imply a close to decay at rest process, where the final state muons would not have sufficient momentum to pass the trigger cuts. Figure 4·27 shows the

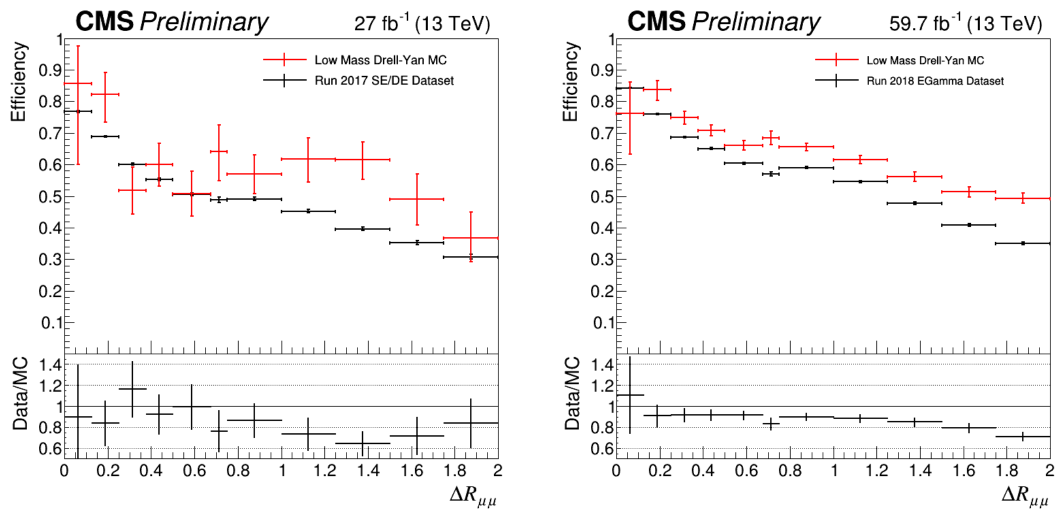

Figure 4·25: Efficiency of dimuon scouting trigger in 2017 (left) and 2018 (right), as a function of ΔR. The 2017 MC sample has fewer available events, which causes the larger statistical error bars.

ratio of the data efficiency to the MC efficiency.

Because Figure 4·25 shows that there is no large systemic variation in the data/MC scale factor as a function of ΔR, our confidence of the distribution of this variable in the data, relative to our desired signal process, is improved and we choose to explore other variables which might exhibit more dramatic mismodeling effects. Differences in the p_{T} spectrum of the individual muons is another possible source of discrepancy, which we study using the sub-leading (lower energy) muon as it is definitionally softer and more likely to be the determining factor in whether or not the trigger criteria are satisfied. Normalized distributions of sub-leading muon p_{T} in data and simulation are shown in Figure 4·28 for both years, which illustrates a non-negligible difference between signal muons and muons in our proxy dataset. To accommodate for this, the previously described kinematic reweighting procedure will be applied to the subleading muon momentum, when determining the efficiency at a given mass.

Figure 4·29 shows the efficiency in data and MC as a function of dimuon mass and subleading muon p_{T}. The weighted average of each column is the trigger efficiency at

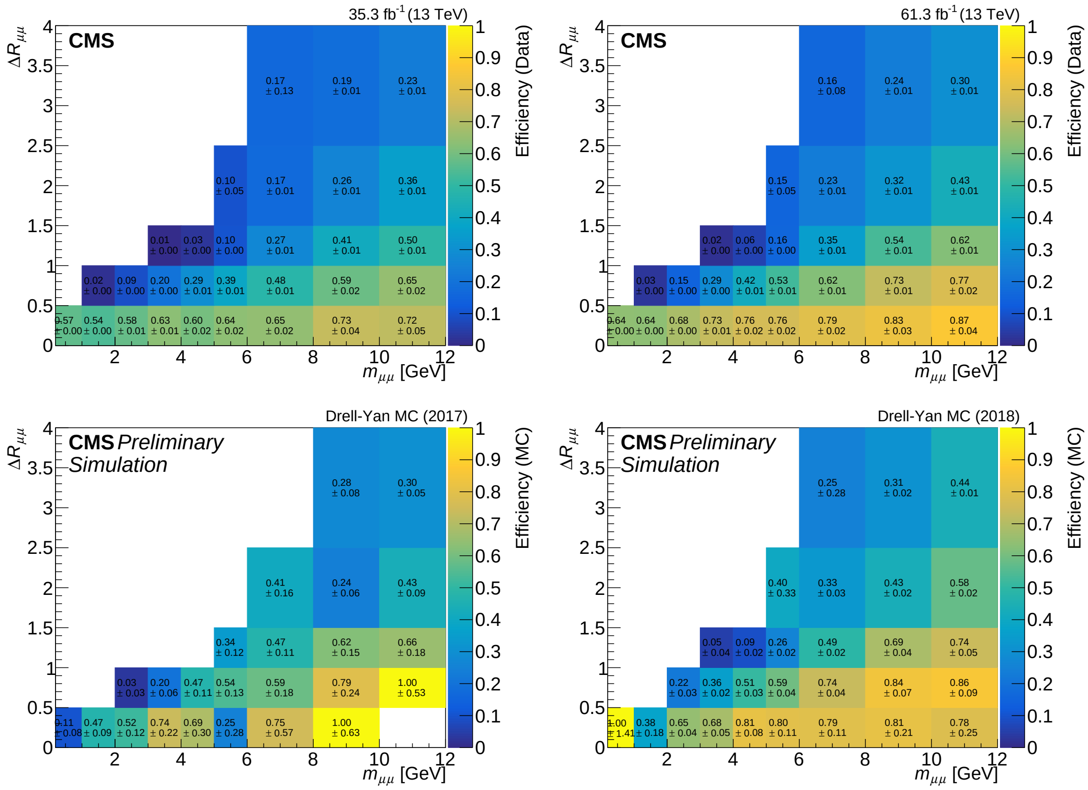

Figure 4·26: Two-dimensional efficiency of dimuon scouting trigger in 2017 (left) and 2018 (right), with respect to data (top) and MC (bottom). Plotted as a function of dimuon mass and ΔR. The 2017 MC sample has fewer available events, which causes larger statistical error, and some sparsely populated bins.

that mass, and is factored in when calculating the production cross section limit for BSM production. Figure 4·30 shows the data/MC scale factors for the same variables, from which we estimate the uncertainty to be the value, $\left|1 - \frac{eff(\text{Data})}{eff(\text{MC})}\right| \times 100\%$.

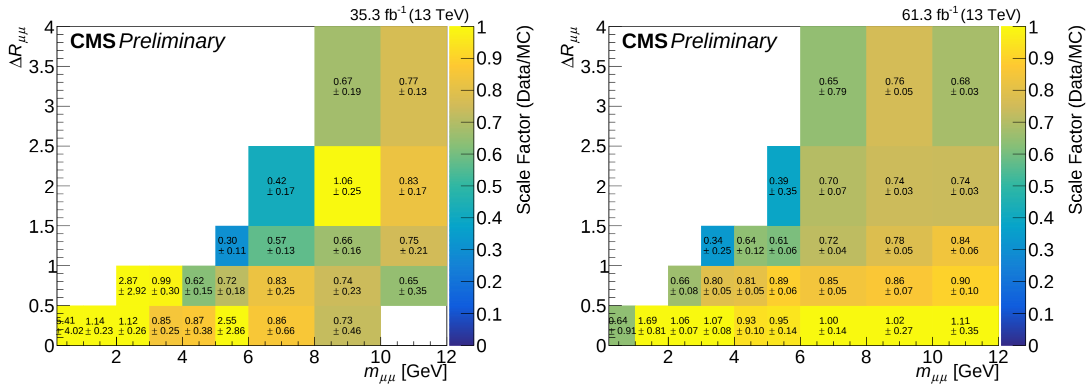

Figure 4·27: Two-dimensional data/MC scale factors of the dimuon scouting trigger in 2017 (left) and 2018 (right), with respect to data (top) and MC (bottom). Plotted as a function of dimuon mass and ΔR.

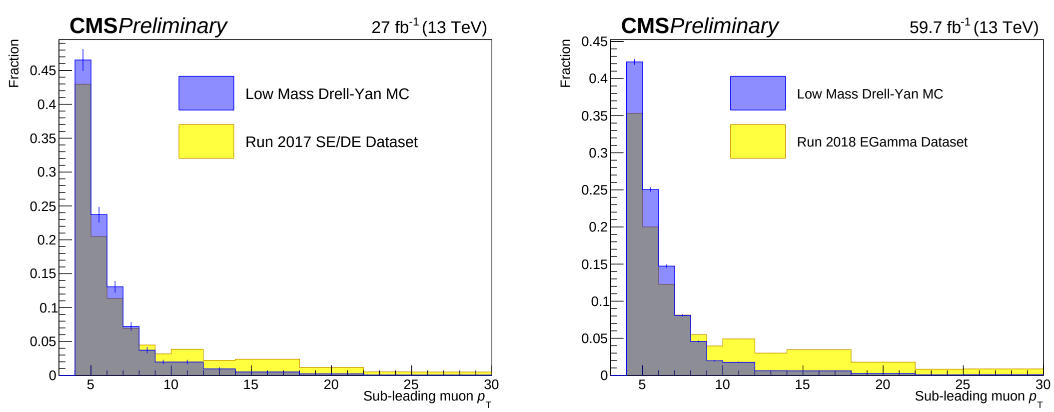

Figure 4·28: Distributions of sub-leading muon momentum in ECAL triggered data vs. MC for 2017 (left) and 2018 (right).

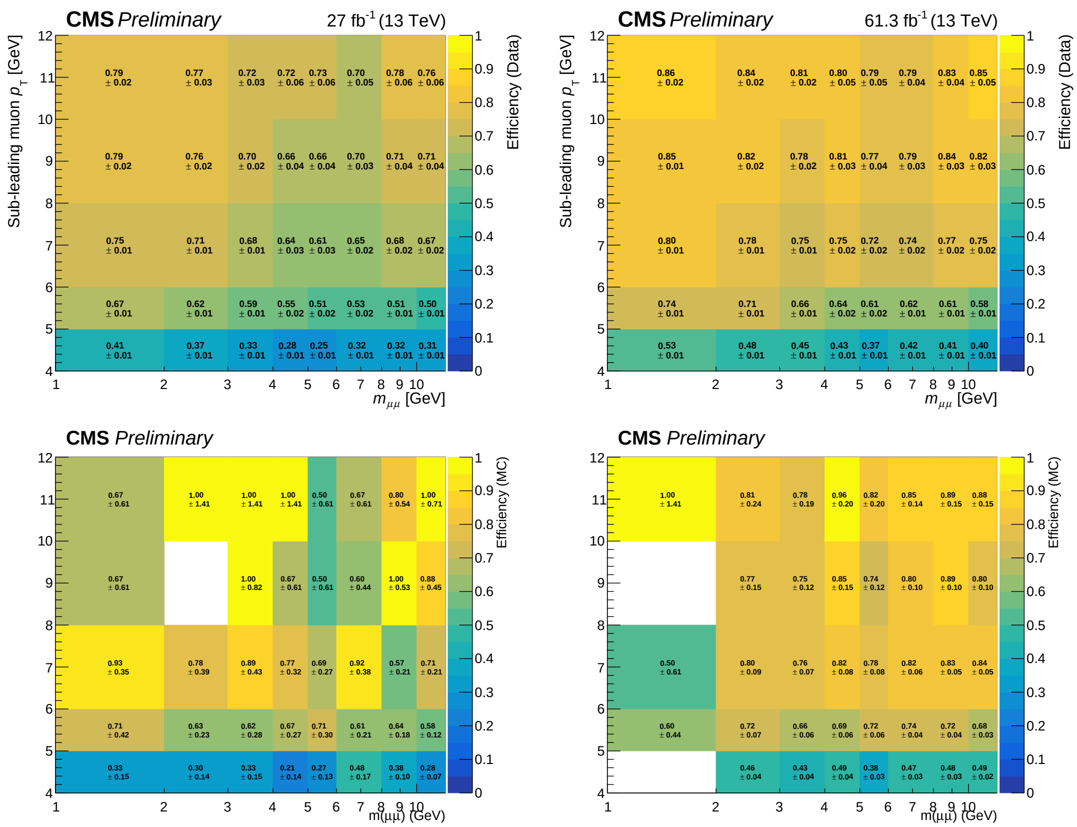

Figure 4·29: Two-dimensional efficiency of dimuon scouting trigger in 2017 (left) and 2018 (right), with respect to data (top) and MC (bottom). Plotted as a function of dimuon mass and p_{T} of the sub-leading muon.

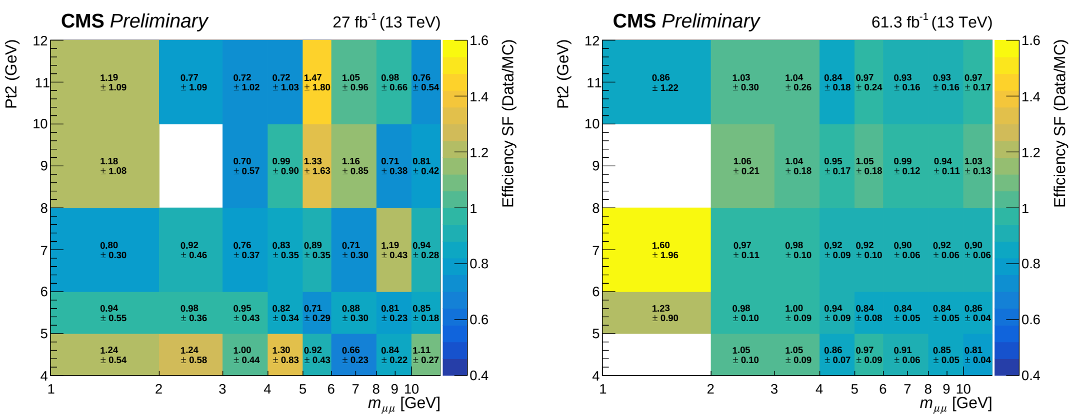

Figure 4·30: Two-dimensional data/MC scale factors of the dimuon scouting trigger in 2017 (left) and 2018 (right), with respect to data (top) and MC (bottom). Plotted as a function of dimuon mass and p_{T} of the sub-leading muon.

4.6.3 Offline Selection Efficiencies

The efficiencies for the high-mass and low-mass selections are studied separately. For each of these, the efficiency of the displacement cut are BDT are measured sequentially; the vertex displacement cut for both IDs are measured with respect to events in the data that have passed the trigger, and the BDT selections are measured with events that have passed the trigger and then the relevant displacement cut.

Vertex Cut Efficiency

We first measure the displacement cut for the low-mass selection ($\sigma_{L_{\mathrm{xy}}} < 3.5$), by exploiting the known signal events under the J/Ψ peak in the scouting data. Normalized distributions of $\sigma_{L_{\mathrm{xy}}}$ are shown in Figure 4·31 for J/Ψ signal MC and events extracted from the data with $m_{\mu\mu} \in [3.04, 3.14]$ GeV. The left plot includes a noticeable contribution from backgrounds in the data that are not attributed to prompt J/Ψ decays, creating the long tail. In the background pure region of the plot, where the resonant contribution has fully fallen off, we can see that these background events are nearly perfectly described by a falling exponential. We model this by fitting the region $8 < \sigma_{L_{\mathrm{xy}}} < 20$ with an exponential, shown in blue, and subtracting it from the data distribution. The right plot shows the background subtracted, renormalized distribution, which has near perfect agreement with the signal MC. The fraction of events below 3.5 in this distribution is the efficiency of the $\sigma_{L_{\mathrm{xy}}}$ displacement cut, which is approximately 97% in both data and MC. As was done with the triggers, we assign the uncertainty for this measurement by measuring the percent disagreement between the efficiency in data and MC. In this case, the disagreement is less than a percent, and we treat the uncertainty as negligible.

The same procedure is followed for the $L_{\mathrm{xy}} < 0.015$ cm cut included in the high-mass selection. This time, the cut is studied using both J/Ψ and Υ events, to gauge

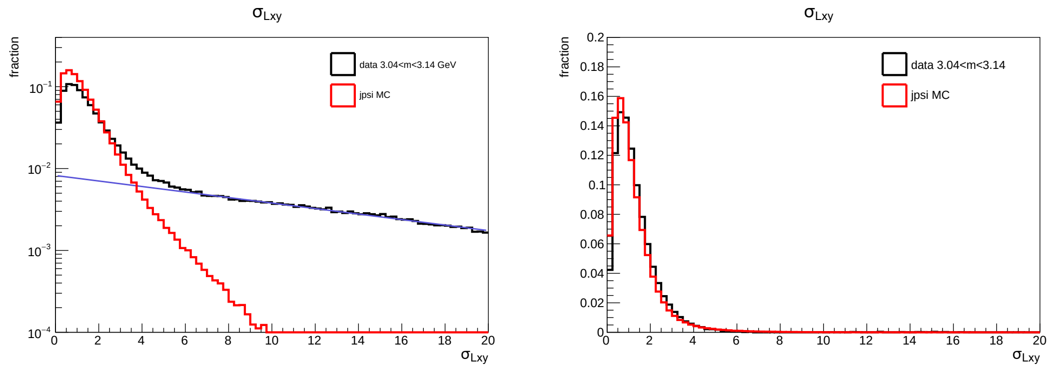

Figure 4·31: Comparison of $\sigma_{L_{\rm xy}}$ in J/Ψ events from data and signal MC. The left shows the distribution before the non-resonant background is subtracted off, with the fit indicated in blue. The right shows the same distributions with the estimated background yield subtracted from the data.

how the efficiency of the cut might vary as a function of mass. Figure 4·32 shows the distribution of $L_{\rm xy}$ in Υ signal MC and data events with $m_{\mu\mu} \in [9.2, 9.7]$ GeV. Again the data is contaminated by an exponentially decaying non-prompt contribution, which is fit in the range $0.06 < L_{\rm xy} < 0.2$ cm and subtracted out. The efficiency of the $L_{\rm xy}$ requirement in these events is 92.5% in the data and 99.0% in the signal MC (ratio = 0.93). This is done for J/Ψ events as well as shown in Figure 4·33, where the efficiency of the cut is 80% in data and 89% in signal MC (ratio = 0.90). We treat the efficiency of Υ events in the data as a flat efficiency for the high-mass category, but assign an uncertainty of 3%, which corresponds to the difference in relative performance between data and MC, when using the Υ versus J/Ψ samples. The results of these measurements are summarized in Table 4.6.

Selection	Signal samples	Data eff.	MC eff.	Ratio
$\sigma_{L_{\mathrm{xy}}} < 3.5$	J/Ψ events	97.1%	97.0%	1.00
$L_{\mathrm{xy}} < 0.015$ cm	Υ events	92.5%	99.0%	0.93
$L_{\mathrm{xy}} < 0.015$ cm	J/Ψ events	80.0%	89.0%	0.90

Table 4.6: All measured efficiencies for L_{xy} and $\sigma_{L_{\mathrm{xy}}}$ displacement cuts.

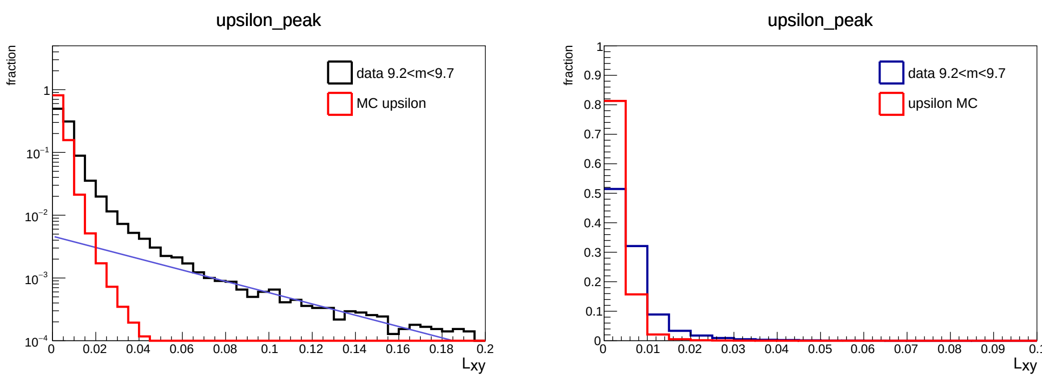

Figure 4·32: Comparison of L_{xy} in Υ events from data and signal MC. The left shows the distribution before the non-resonant background is subtracted off, with the fit indicated in blue. The right shows the same distributions with the estimated background yield subtracted from the data.

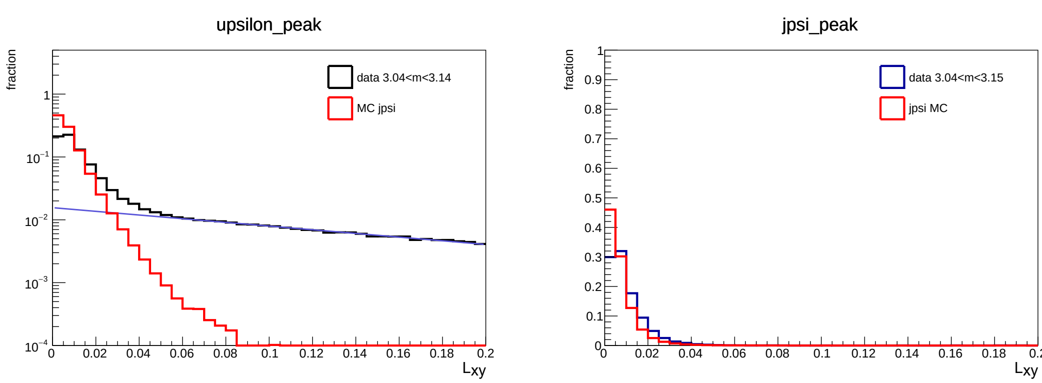

Figure 4·33: Comparison of L_{xy} in J/Ψ events from data and signal MC. The left shows the distribution before the non-resonant background is subtracted off, with the fit indicated in blue. The right shows the same distributions with the estimated background yield subtracted from the data.

BDT Efficiency

To measure the efficiency of the BDT categorizers, `JPsi3` and `Ups5`, we use the TnP approach with resonances in the data comprised of events that have passed the trigger and the corresponding vertex selection. Similar to when we used TnP to build the sample of unbiased muons for the BDT training, the tag muon definition is $p_{\mathrm{T}} > 12$ GeV and it must pass the legacy cut-based ID. For this measurement it must additionally satisfy the vertex selection. The probe muon must pass the baseline analysis cuts of $p_{\mathrm{T}} > 4$ GeV and $|\eta| < 1.9$ and also satisfy the relevant vertex selection.

For studying the efficiency of the high-mass BDT, `Ups5`, we use events from a range around the Υ-peak, between 8 and 12 GeV. These events are sorted into bins depending on the p_{T} of the probe muon, and the ΔR of the muon pair, where the efficiency will be measured separately. For each of these bins, the events are further divided into a "passing" and "failing" category, depending on whether or not the probe muons satisfies the `Ups5` BDT criteria. The dimuon mass spectrum for both of these categories shows the peaks of the $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$, which are our proxy signal process for this measurement. To negate the contribution from the continuum backgrounds, we fit the signal with a sum of three Voigtian functions and the background to a falling exponential, which can be subtracted off from the passing and failing yields. Figure 4·34 shows the mass spectra of the TnP samples for a selection of p_{T} bins, as well as the signal and background fits. These, and all other fits considered converged well, with the same signal and background model.

The efficiency of the BDT to identify a signal muon is defined as the number of Υ events where the probe passed, divided by the total number of Υ events in a given bin. Figure 4·35 shows the efficiency of `Ups5` as a function of ΔR and sub-leading muon p_{T} for Υ events in both years of data.

A similar measurement is performed using J/Ψ events to determine the perfor-

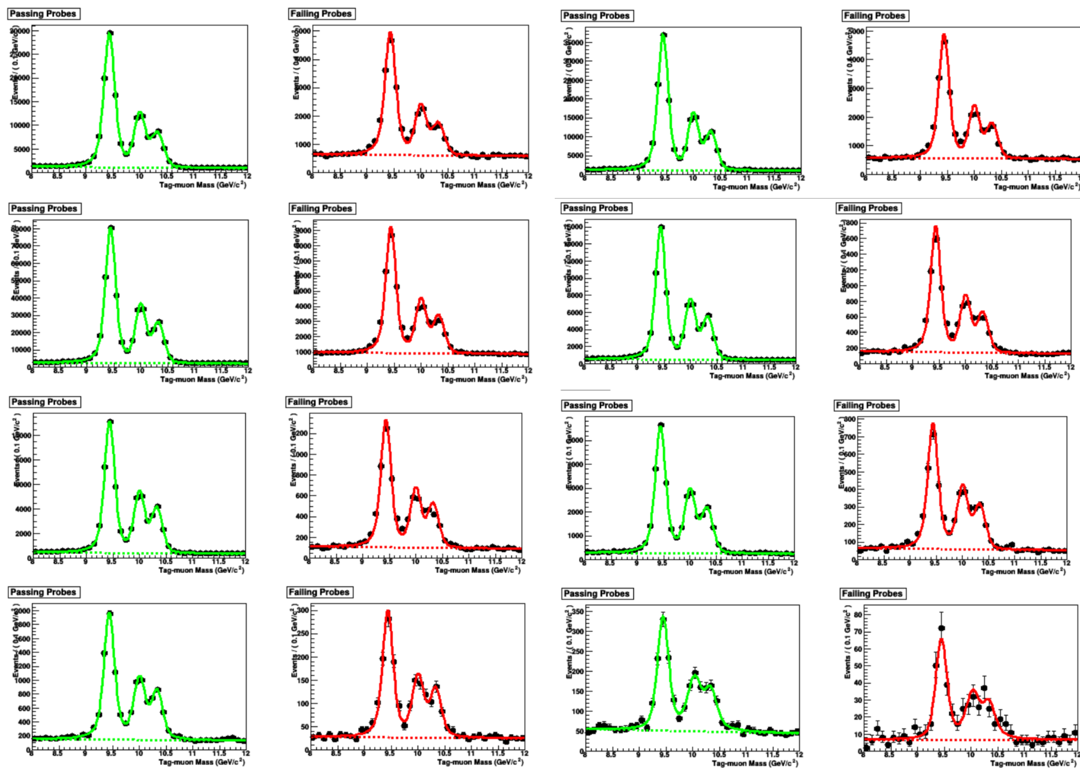

Figure 4·34: TnP event distributions and signal fits for passing and failing probe muons. Each pair of plots represent events in the ΔR bin [0.5,0.7] and subleading muon p_T ranges of [4,5], [5,6], [6,10], [10,12], [12, 15], [15, 20], [20,30], and [30,50] GeV (beginning from the top left and reading to the right, row-by-row).

mance of the low-mass BDT, `JPsi3`. These events are collected and organized into passing and failing samples by the same procedure as the high-mass selection. Events in these samples are subjected to the low-mass displacement cut, $\sigma_{L_\mathrm{xy}} < 3.5$, and are fit with a single Voigtian function plus exponential background, in the dimuon mass range between 2.8 and 3.4 GeV. To reflect the more boosted kinematics of low mass processes, the ΔR range considered is tightened, and binned more finely. Figure 4·36 shows the efficiency of a J/Ψ signal event in 2017 and 2018 data to pass the `JPsi3` identifier, again expressed as the ratio of passing probe muons to total probe muons in each bin.

To estimate the uncertainty on both of these measurements, we rely again on

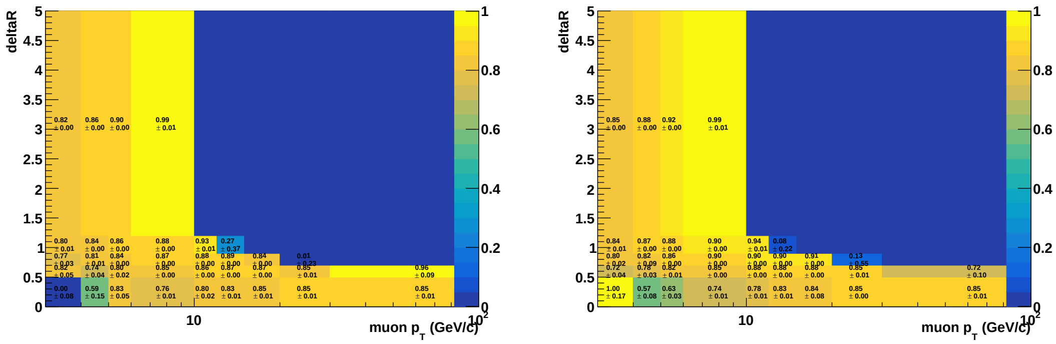

Figure 4·35: Efficiency of BDT ID `Ups5` to identify signal muons from Υ decays in 2017 (left) and 2018 (right) data.

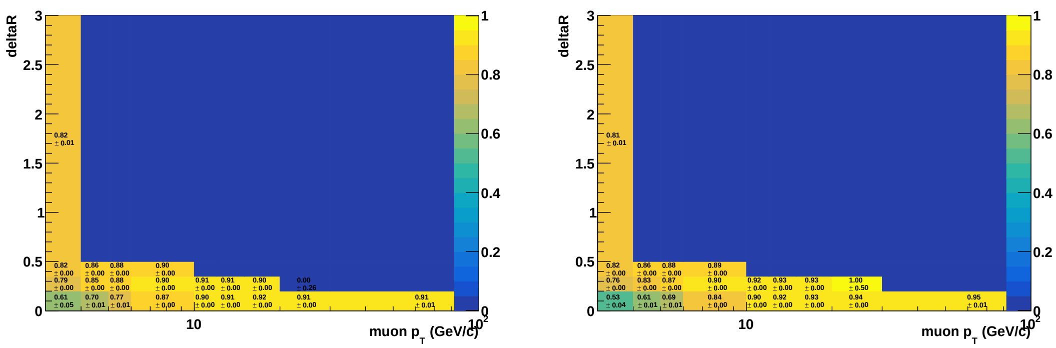

Figure 4·36: Efficiency of BDT ID `JPsi3` to identify signal muons from J/Ψ decays in 2017 (left) and 2018 (right) data.

comparison with the BDTs performance in signal MC. For these studies, we use two samples of resonant SM production; one of J/Ψs and one of $\Upsilon(1s)$s, both decaying promptly into muons. Figure 4·37 shows the efficiency of both identifiers relative to their matching signal sample. Figure 4·38 shows the ratio of these values to the corresponding measurement in data, for both years.

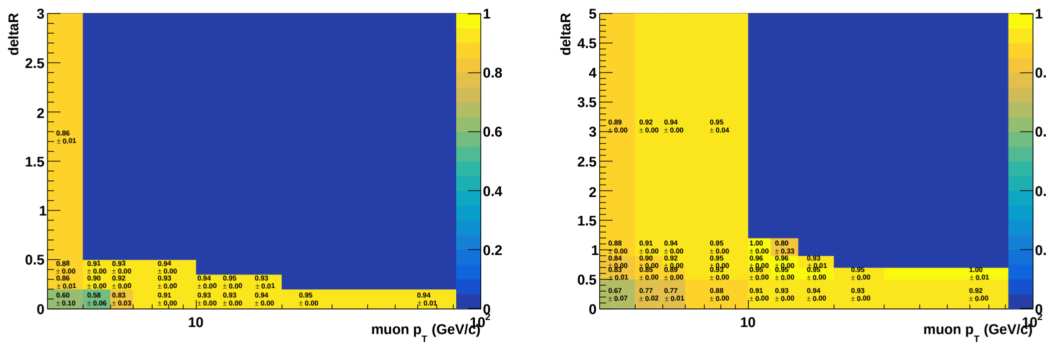

Figure 4·37: Efficiency of BDT IDs with respect to their relevant MC signal samples. `JPsi3` is measured using SM $J/\Psi \to \mu\mu$, and `Ups5` is measured using $\Upsilon \to \mu\mu$.

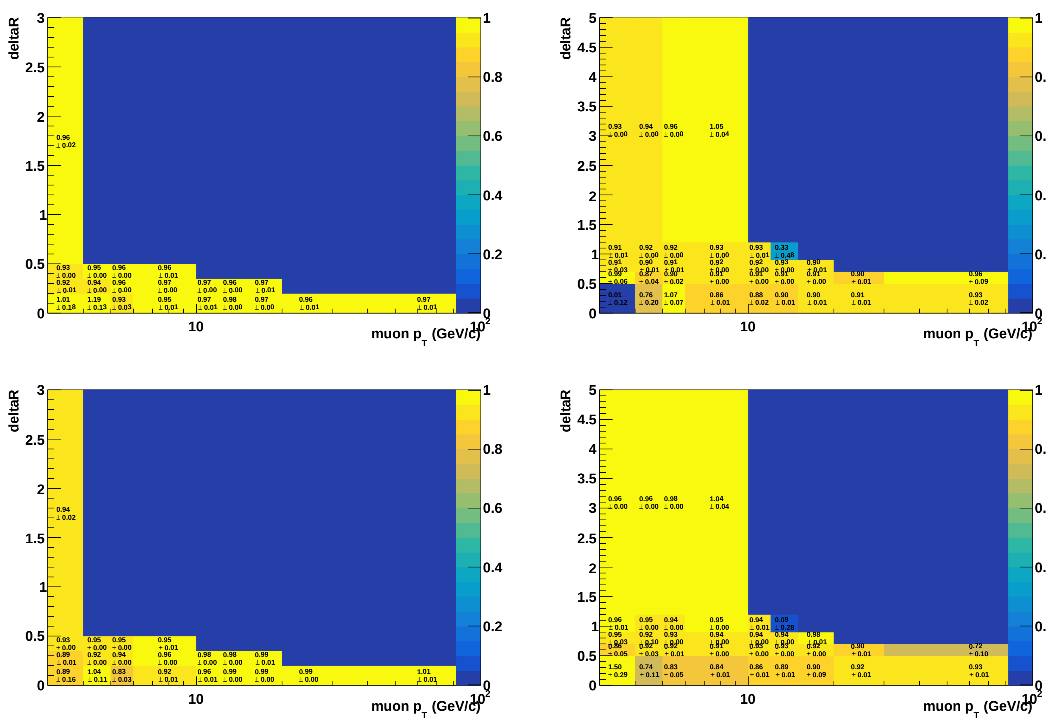

Figure 4·38: Data/MC scale factors for ID efficiency of `JPsi3` (left) and `Ups5` (right), using data from 2017 (top) and 2018 (bottom).

ID Application Closure Tests

When considering our methodology for measuring the trigger efficiency, one of the first checks to validate our approach was to compare the data-driven efficiency values to those in signal MC, rescaled so as to cancel out effects related to reconstruction in data versus simulation. This gave us confidence that the events we were studying in data were not hiding pathological traits or kinematic properties that would cause them to be a poor model for our actual signal process. Similarly, there is a possibility that the some variables not studied in our BDT efficiency measurements may be highly impactful, causing our data-driven efficiency to not match that of the signal processes. To check this, we examine the efficiency of both BDTs as a function of two possibly confounding variables in DY and pseudoscalar signal MC; the number of pile-up interactions in the event and the η value of the probe muon. These measurements are shown in Figure 4·39, and do not exhibit any significant dependance of the efficiency on either variable.

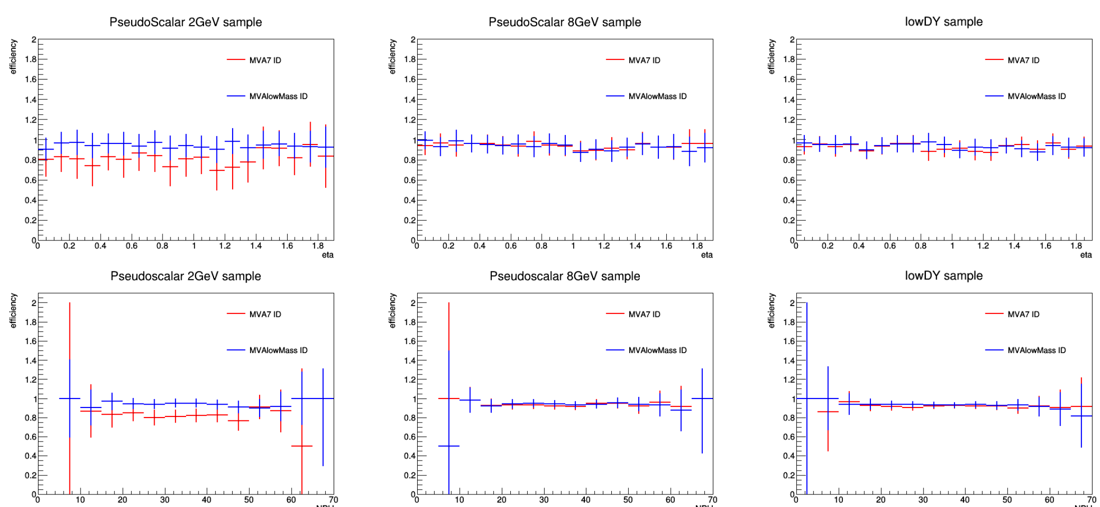

Figure 4·39: Efficiency of BDT IDs `JPsi3` and `Ups5` as a function of probe muon η (top) and number of pile-up interactions (bottom). Measurement is performed on a 2 GeV pseudoscalar sample (left) and 8 GeV pseudoscalar sample (center) and the low-mass DY sample (right).

A final check is performed by examining the dimuon yield in the DY MC sample, scaled down by the efficiency as measured in data. For this operation, the efficiency used at each mass point is determined using the 2D distributions shown in Figures 4·36 and 4·35, weighted according to the expected ΔR and muon p_{T} distribution at the mass value in question. This rescaled distribution is then compared with the same DY sample, now with the BDT actually being applied to the events. This effectively scales the sample down by the BDT efficiency in MC. This line is then multiplied by the data/MC scale factor determined from the TnP measurements. Figure 4·40 shows that this operation recovers the distribution with the data-driven efficiency applied almost exactly. This gives us confidence that the BDT efficiency does not vary due to any latent variables which might be correlated with the dimuon mass, except for ΔR, which is simple to model.

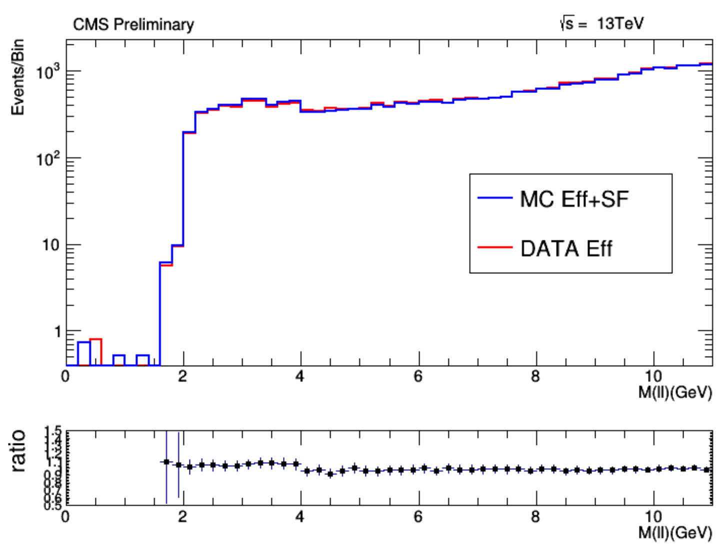

Figure 4·40: Dimuon mass spectrum from the low-mass DY signal sample when rescaled by the selection efficiency in data (DATA Eff), versus the same distribution with the BDT applied, rescaled by the data/MC scale factor. `JPsi3` and the corresponding J/Ψ TnP measurement is used below 4 GeV. `Ups5` and the corresponding Υ measurement is used above 4 GeV.

The combined efficiency of the vertex and BDT offline selections is shown as a function of dimuon mass in Figure 4·41, determined using DY signal MC. Below 2 GeV it is difficult to generate DY events due to modelling of the quark parton-distribution function (pdf) for very low energy processes, and no events are available in the sample. An alternative sample, consisting of dimuons from a 1 GeV pseudoscalar is included on this plot. A linear extrapolation is drawn between the DY and pseudoscalar points, which matches the falling trend that can be seen above the red line.

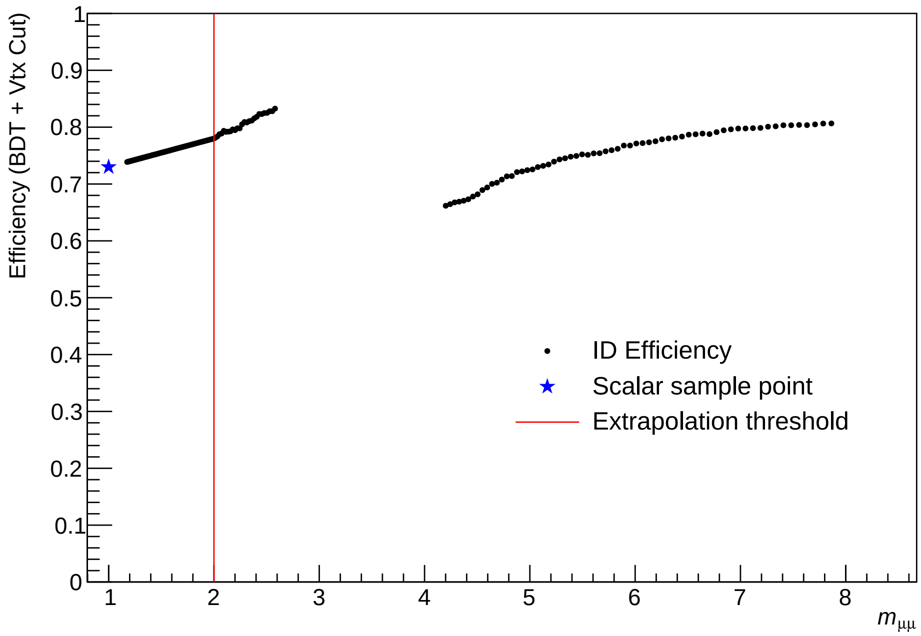

Figure 4·41: Efficiency of a muon to pass both the BDT ID and vertex selection for its relevant mass range. The vertical red line shows the point where the DY sample ends, and a linear extrapolation to the pseudoscalar measurement begins.

Efficiency Measurement Summary

The combined efficiency for an event to pass the dimuon scouting trigger and our offline selection is shown in Figure 4·42, for both analysis categories. The offline efficiency measurements for the boosted category are not exhaustively described, but follow an identical methodology to the inclusive measurement with the addition of the tightened fiducial requirement. Because of the sequential structure of our efficiency measurements, i.e.:

$$\epsilon(\text{Total}|\textit{signal}) \approx \epsilon(\text{Sel.}|\text{Vtx.}) \times \epsilon(\text{Vtx.}|\text{Trigger}) \times \epsilon(\text{Trigger}|\textbf{MediumID}), \quad (4.15)$$

the total efficiency can be factorized into the components we have measured, and recovered by simply multiplying them together, as shown in the figure. Because the TnP strategy used for the BDT gives the efficiency of an individual muon to satisfy the BDT cuts, two factors of the BDT efficiency are applied, representing the odds of two muons with nominally uncorrelated quality needing to satisfy the selection to be a good signal candidate. The accuracy of this method as a proxy for the "true dimuon BDT efficiency" is investigated in Section 4.8.

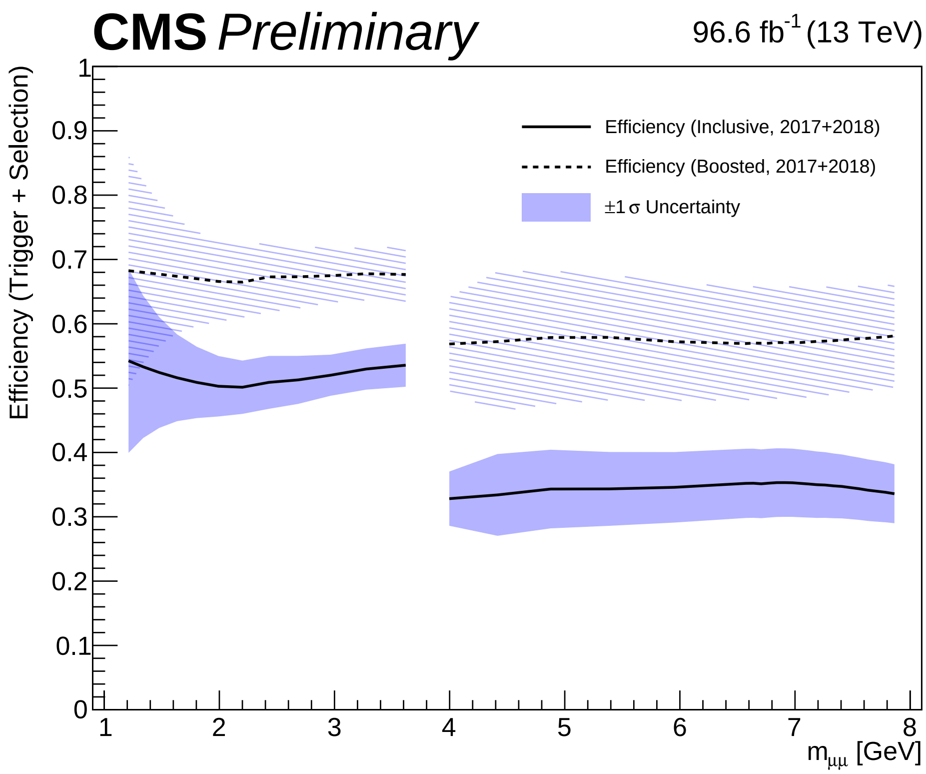

Figure 4·42: Efficiency of a dimuon event to pass both the trigger and offline event selections as a function of dimuon mass. The solid line and shaded band corresponds to the inclusive category, and the dashed line corresponds to the boosted category. The efficiency is expressed as a weighted average of both years, according to the recorded luminosity from each. Shaded bands correspond to the estimated standard deviation on this measurement, determined by quadrature summing the uncertainties of the underlying measured efficiencies.

4.7 Signal and Background Modeling

With the input criteria for the analysis thoroughly defined, the remaining component of this search is to develop the shapes that will be used to model the continuum background and peaking signal contributions. These shapes will undergo a simultaneous maximum-likelihood (ML) fit in a small window of the dimuon mass distribution, with the signal shape fixed directly in the center of the window. The post fit amplitude of the signal shape will be converted to an upper limit on BSM production at the central mass value of that window, then the window will increment to a slightly higher mass range, and the process will repeat.

4.7.1 Signal Modeling

The true underlying shape of a resonance in the dimuon mass spectrum can vary wildly depending on the BSM phenomenology one is considering. For the dark photon and scalar production modes that this analysis intends to constrain, the likely physical shape of the resonance will match a relativistic Breit-Wigner function [79], peaking sharply around the rest mass of the particle. However, in our data the reconstructed mass of any measured pair of muons is subjected to some degree of smearing, due to the imperfect momentum resolution of the measurement of the final state muons. This effect is dramatic enough that the shape of any given resonance in the data is likely to be entirely dominated by the mass smearing, and the underlying signal shape is irrelevant as it cannot be resolved.

While this mass smearing is generally undesirable, as it causes the presence of a signal to be flatter and more difficult to distinguish from background, it does greatly simplify the development of our signal shape. Because the width of the reconstructed resonances is so much wider than the physical resonances we are searching for, the signal shape we develop will be a good fit for any resonant process that is significantly more narrow than the minimum width we can resolve in our data. This means that our model-independent limit on BSM production can be reinterpreted to constrain any dimuon resonance process, so long as it does not include predictions for particles with decay widths similar to or broader than our intrinsic mass resolution.

The signal model we use will be the sum of a Gaussian distribution and a double-sided Crystal Ball (dCB) function [80, 81], the latter of which is suited to capturing long tails on either side of the signal distribution in data. To predict the best values for the many free parameters included in this combined function, we again exploit the sideband meson resonances in data. This time they are used to perform a series of calibration fits to the Gaussian+dCB shape, from which we will extract the best-fit function parameters. With our assumption that the signal shape in data is dominated by resolution effects, they should be passable descriptors for any narrow BSM resonance. These fits are shown for the J/Ψ, $\Psi(2S)$, and the $\Upsilon(1S)$), (2S), and (3S) peaks in Figure 4·43, for both years of data in the inclusive category. The background shape for these fits is modeled by a fourth order Bernstein polynomial, which will be subjected to further study below, as a candidate for the background functional form in this search. The post-fit parameters from these studies, except for amplitude, are averaged and treated as frozen parameters for the actual ML fit.

The width of the signal shape is excepted to vary as a function of the dimuon mass, but the relative width of the peak ($\sigma/m_{\mu\mu}$) should be roughly constant across all the masses we consider. Figure 4·44 shows the relative width of all resonances from the above fits, plus fits to the Φ and ω mesons. Variation between these fits is very small except for the ω meson, most likely due to its partial overlap with the ρ meson peak. Disregarding this highly discrepant point, the remaining variation lies within a range of $\pm 20\%$ about the chosen working value of 0.013. Resonance width is the only, non-amplitude, signal parameter that is allowed to vary during the ML fit, and

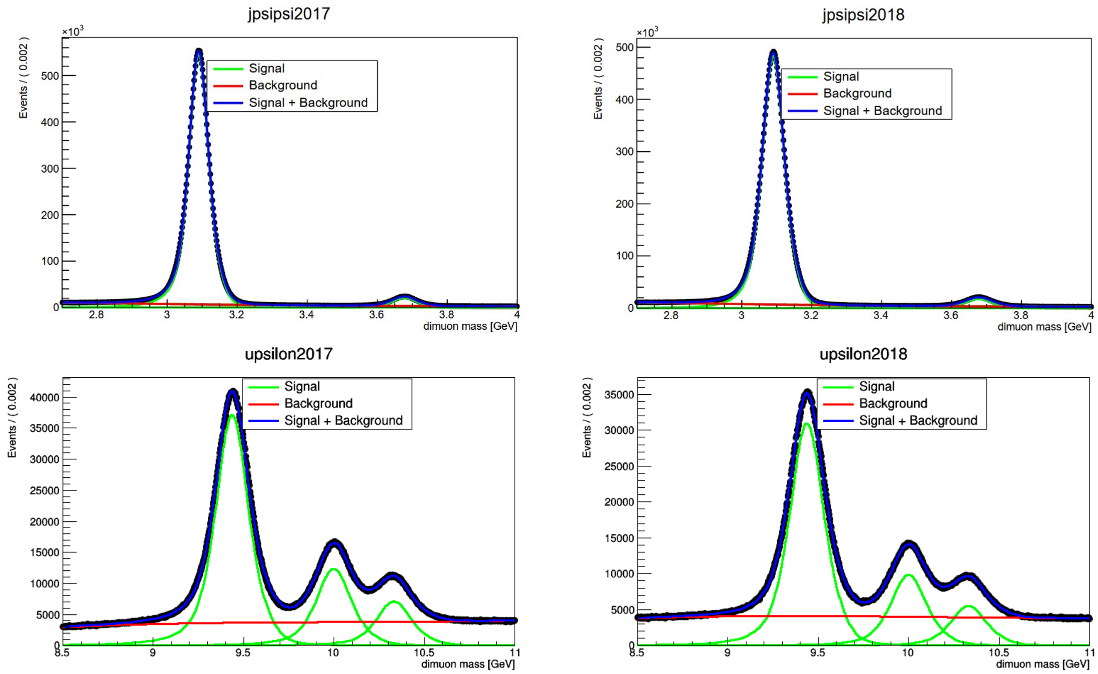

Figure 4·43: Signal plus background fits to known SM resonances in 2017 (left) and 2018 (right) data. The upper plots show the fits to the J/Ψ, and $\Psi(2S)$ peaks. The lower plots show the fits to the Υ peaks.

it is constrained by the stated uncertainty on relative width. It is possible that the tightened p_{T} requirements for the boosted category will impact the relative width, so these fits are repeated with the mesons as they appear in the boosted dimuon mass distribution, shown in Figure 4·45. A small systemic increase in relative width can be observed, but it remains within the allotted 20% uncertainty, therefore we choose to use the same function parameterization for both categories.

4.7.2 Background Modeling

The signal is fit on top of a polynomial-like background shape, meant to constrain the contribution from the smooth SM DY continuum. This shape extends to the edges of the sliding mass window, giving it access to some data far away from the signal shape. The size of the window is defined relative to the width of the signal shape, to ensure that there is a sufficient sideband to stabilize the background fit. The chosen

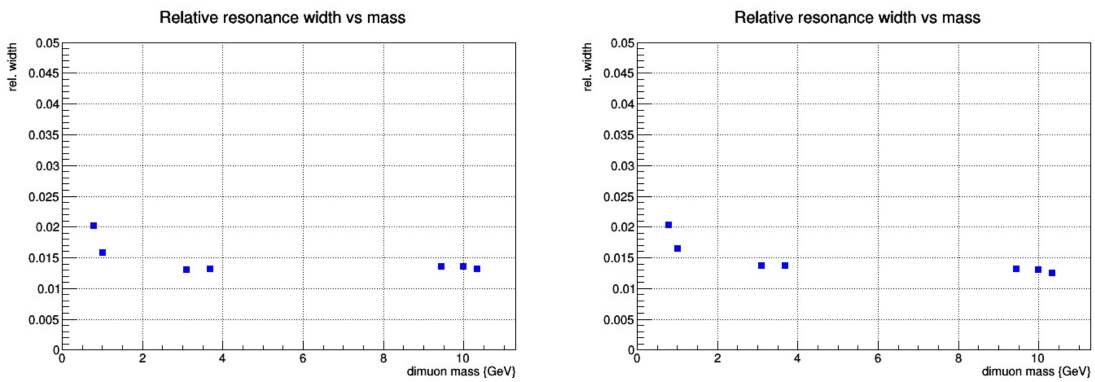

Figure 4·44: Relative width of SM peaks in the inclusive analysis category in 2017 (left) and 2018 (right).

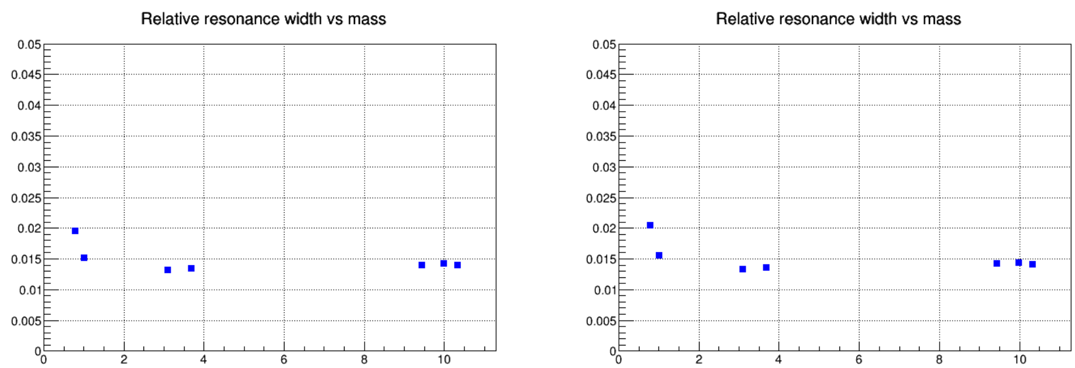

Figure 4·45: Relative width of SM peaks in the boosted analysis category in 2017 (left) and 2018 (right).

window size for this analysis is plus or minus 5 (8) times the expected signal width for the inclusive (boosted) category. The decision to use a larger window size for the boosted category was made very late into the development of this analysis, after most techniques had been frozen and the data had been fully unblinded. The main body of this thesis will present only the methods and results of this analysis as it exists in its current published form, however, an accounting of the discussion and studies that motivated this late-stage adjustment is provided in Appendix B.

Given the contributions from the tails of the SM peaks in the data, as well some sculpting of the dimuon mass distribution due to trigger turn-on effects, it is extremely difficult to pose a well motivated analytical function to model the shape of the background in a given window. To circumvent this, we instead assemble a set of purely mathematical functions that can conform to the data in all mass windows and gauge their fitness as a background model via empirical metrics, such as background only goodness-of-fit and bias towards certain signal strength measurements. To explore a wide variety of possible forms, we derive these candidate models from four different families of functions:

- Bernstein polynomial: $B_n(x) = \sum_{\nu=0}^{n} \beta_\nu b_{\nu,n}(x)$, where $b_{\nu,n} = \binom{n}{\nu} x^\nu (1-x)^{n-\nu}$
- Polynomial times exponential: $P_n(x) = e^{c \cdot x} \sum_{n=1}^{n} \beta_n x^n$
- Sum of exponentials: $E_n(x) = \sum_{n=1}^{n} a_n e^{c_n \cdot x}$
- Bernstein polynomial plus power law: $Bp_n(x) = f B_n(x) + (1-f) x^a$

The strategy for the background model is to pick one function from each of these families, identify the optimal order that will allow them to accurately describe the dimuon mass distribution within most mass windows, and include each of these candidates in the ML fit. This is implemented using the "discrete profiling" approach [82], wherein the fit has access to multiple background functions which it may switch between as a discrete nuisance parameter. This gestalt function will be then be thoroughly studied to ensure that it does not have a systemic preference to produce spurious signals in the fit, or hide real ones.

Function Order Selection

The choice of which order of function from each of the above families should be included in the background profile is determined via a Fisher test (F-test) [83]. Generically, an F-test is used to gauge the likelihood that two normal populations are drawn from underlying distributions with the same or different variance. An “F statistic” is formulated, which encodes the relative in-group variance of each population, and it is compared to some desired confidence level for the study in question. If the value of the F statistic is lower than the confidence threshold, then the null hypothesis of equal underlying distributions is rejected. The F-test can be reformulated to estimate the relative performance of two different models in describing the same sample of data. In this version of the F-test, one of the models is assumed to be of higher complexity than the other. The lower complexity function has n free parameters (p_n) and a summed least squared error metric when fit to the data (χ^2_n). The higher complexity function has m free parameters (p_m), where $m > n$, and a corresponding error metric (χ^2_m). The F statistic for this study compares the goodness of fit between the two models and has the form:

$$F = \frac{\frac{\chi^2_n - \chi^2_m}{p_m - p_n}}{\frac{\chi^2_m}{n_{data} - p_m}}. \tag{4.16}$$

n_{data} is the number of points in the data; in our case the number of bins per mass window, which is set to 100. For this test, the null hypothesis supposes that the more complex function does not provide a significant improvement in modeling the data over the simpler function. We begin with the lowest order function that can be constructed within a given family and compare to the next highest order; for example, a first and second order Bernstein polynomial [84] ($B_1(x)$ and $B_2(x)$). We perform background only fits of both functions to the dimuon mass distribution in every window where we intend to set a limit and calculate the F statistic for each window. If the F statistic is large for a given window, that indicates that the data there is complex enough that the additional parameters in the higher order function appreciably improve the fit, and that the complex function may be more favored for the analysis. We treat each window as an independent dataset, but assume that the underlying complexity of the dimuon mass distribution is roughly flat across all mass ranges we will fit, and seek a single function that best models all of them.

For any comparison of functions with different numbers of parameters there is an underlying F-distribution, the shape of which is determined by the number of data-points being fit as well as the number of free parameters of both models. The F-distribution is the expected distribution of the F statistic in the case of random samples being drawn from a world that truly matches the null hypothesis of the F-test in question. Agreement between this distribution and the set of measured F statistics from our study is how we determine when we have arrived at an ideal function to model the background, i.e. the null hypothesis is preferred and additional parameters no longer improve the modeling. To quantify this agreement between our studies and the F-distribution, we choose a desired false-rejection confidence level (α), which corresponds to some "critical value" of the F statistic (F_{crit}), above which the area under the F-distribution is equal to α. In our case, we chose a confidence level of $\alpha = 0.05$. This means that for a comparison of two functions, we might observe up to 5% of our window fits to have an F statistic above the value of F_{crit}, and we would interpret that as the expected number of Type 1 errors under the null hypothesis. In that case we would conclude that we had reached a point of diminishing returns on the background function complexity and select the lower order function. Conversely, if significantly more than 5% of the fits have an F statistic above F_{crit}, then we infer that the null hypothesis is rejected and we stand to improve our fit by selecting the higher order function. If that occurs, this function is paired up with the next higher order function in the same family, and the study is repeated.

For all of our studies, we are comparing functions that differ by only 1 free parameter and are fitting them to data with significantly more bins than model parameters. This results in a very similar F-distribution for all of the function comparisons, and

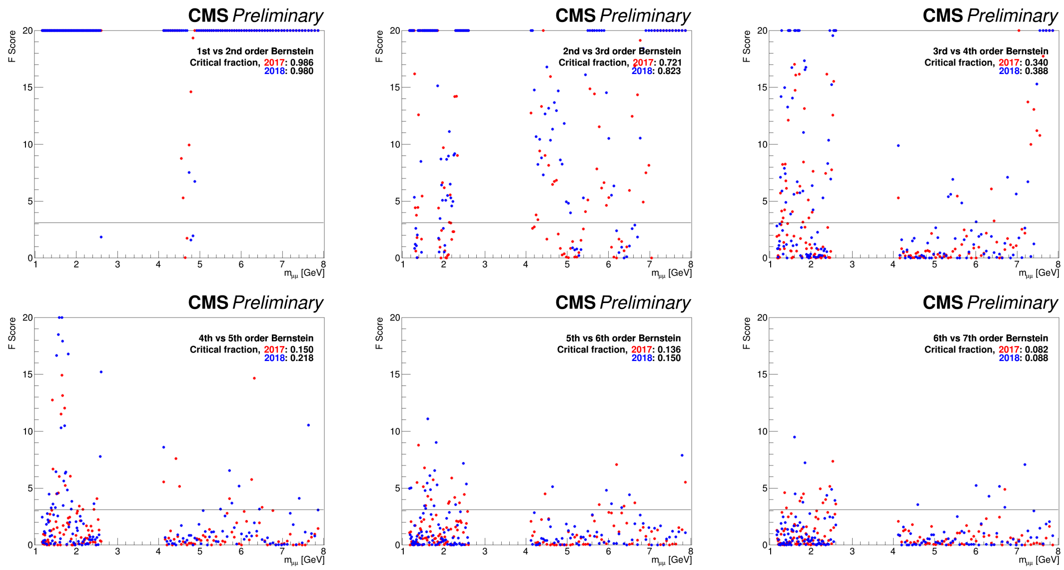

Figure 4·46: Scatter plots of measured F statistics for comparisons of successively higher order Bernstein polynomials fit to the inclusive selection data. Horizontal line is the critical value of the F statistic for these F tests at a false-rejection probability of 0.05.

roughly the same critical value, $F_{crit} \approx 4$, for the chosen confidence level of $\alpha = 0.05$. Figures 4·46 and 4·47 shows a set of F-tests to identify the optimal order from the family of Bernstein polynomials for the inclusive and boosted categories respectively. Though the number of outlier F statistics above 4 never reaches exactly 5%, the improvement in both years sharply plateaus after the comparison of 4th vs 5th order. Further, most of the outlier values for the F statistic in this measurement are localized at an area of the low mass distribution that does not abide the assumption of flat complexity, due to the presence of small peaking SM background shapes which are described in Section 4.7.3. Given this, we choose 4th order Bernstein polynomial as the representative order from this family and repeat these studies for the other functions.

The final cadre of candidate background function for this measurement, after conducting these F-tests for all function families, is chosen to be: a forth order Bernstein

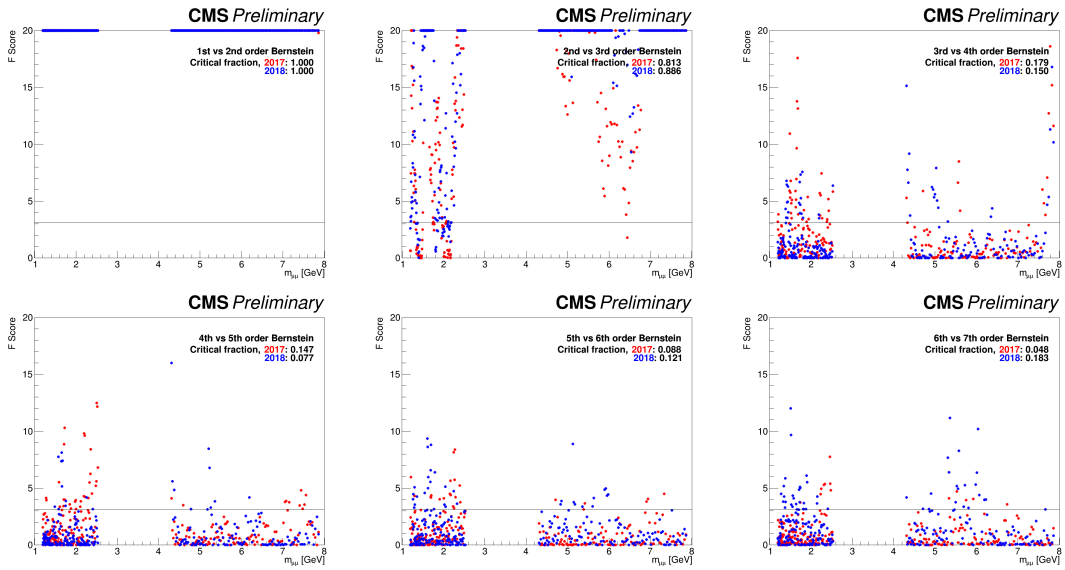

Figure 4·47: Scatter plots of measured F statistics for comparisons of successively higher order Bernstein polynomials fit to the boosted selection data. Horizontal line is the critical value of the F statistic for these F tests at a false-rejection probability of 0.05.

polynomial ($B_4(x)$), a forth order polynomial times an exponential ($P_4(x)$), a sum of 7 exponentials ($E_7(x)$), and a third order Bernstein polynomial plus a single power-law term ($Bp_3(x)$). The goodness-of-fit for each of these functions across all mass windows in both years and categories is shown in Figure 4·48. Generally, the best fitting function is $B_4(x)$, though other functions are occasionally favored.

Bias Studies

When fitting these functions simultaneously with the discrete profiling method, a penalty is applied to the likelihood of each function proportional to the number of parameters it has. This prevents more complicated background functions from over-fitting the data and biasing the measurement. In addition to this, we desire a high degree of confidence that none of our chosen background function forms, when considered together or separately, has a tendency to impart bias on the signal measurement. A background function would be considered biased if it has a tendency to result in

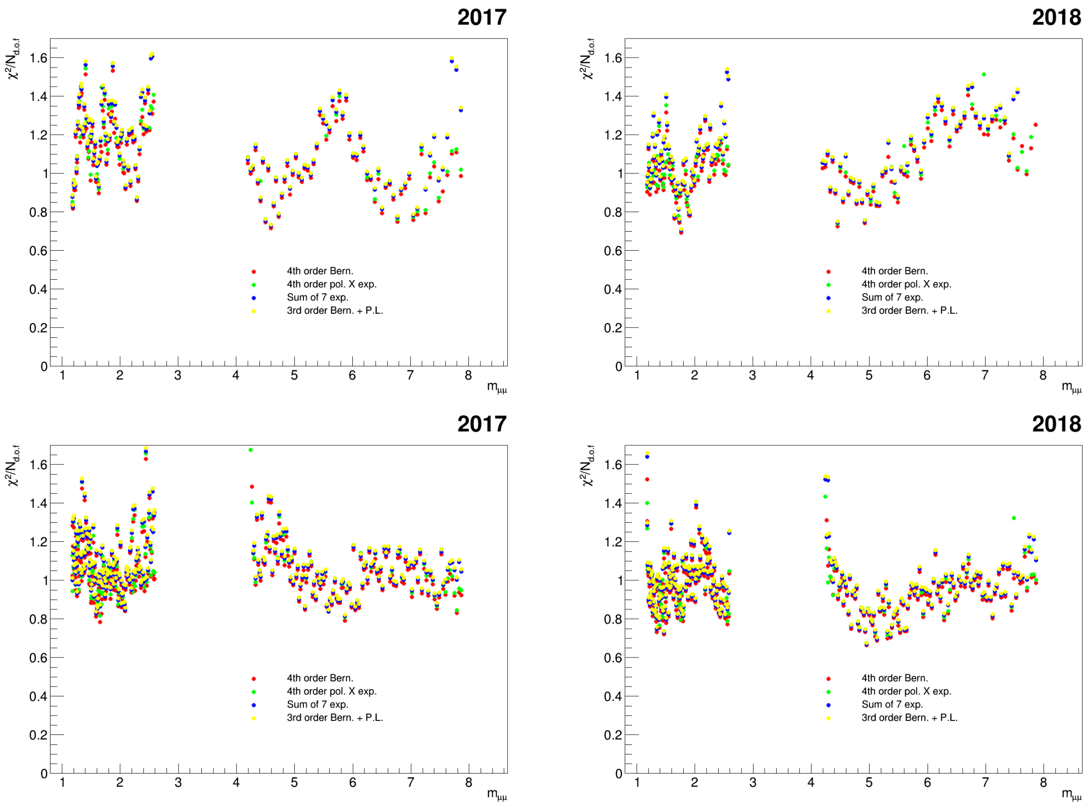

Figure 4·48: Scatter plots of the χ^2 goodness-of-fit for all functions to fit the analysis mass windows, in 2017 (left) and 2018 (right) data. The top (bottom) row corresponds to the inclusive (boosted) category inputs.

a certain measured signal value, positive or negative, regardless of the true signal content of the data.

The strategy to quantify the bias introduced by our chosen functions is to apply them to fits of toy data, where the true contribution from signal is precisely known. To create these toys, we select a sample of mass windows from the data, and subject them to a background only fit with one of the candidate functions. We then sample this post-fit background function to produce a set of 2000 toy histograms, each similar to the original data, but known to have no peaking signal at all. We then perform the ML signal plus background fit on every toy and fill a histogram of the measured signal strengths of each sample. Examples of such histograms are shown in Figure 4·49, for the case of the 4th order Bernstein being fit back to toys generated from itself. The average value of this histogram, determined from a Gaussian fit to the distribution, is the systemic deviation of this fit from the true signal strength in the toy data, which is nominally zero. We call this value the “bias” of that particular function in fitting that mass window.

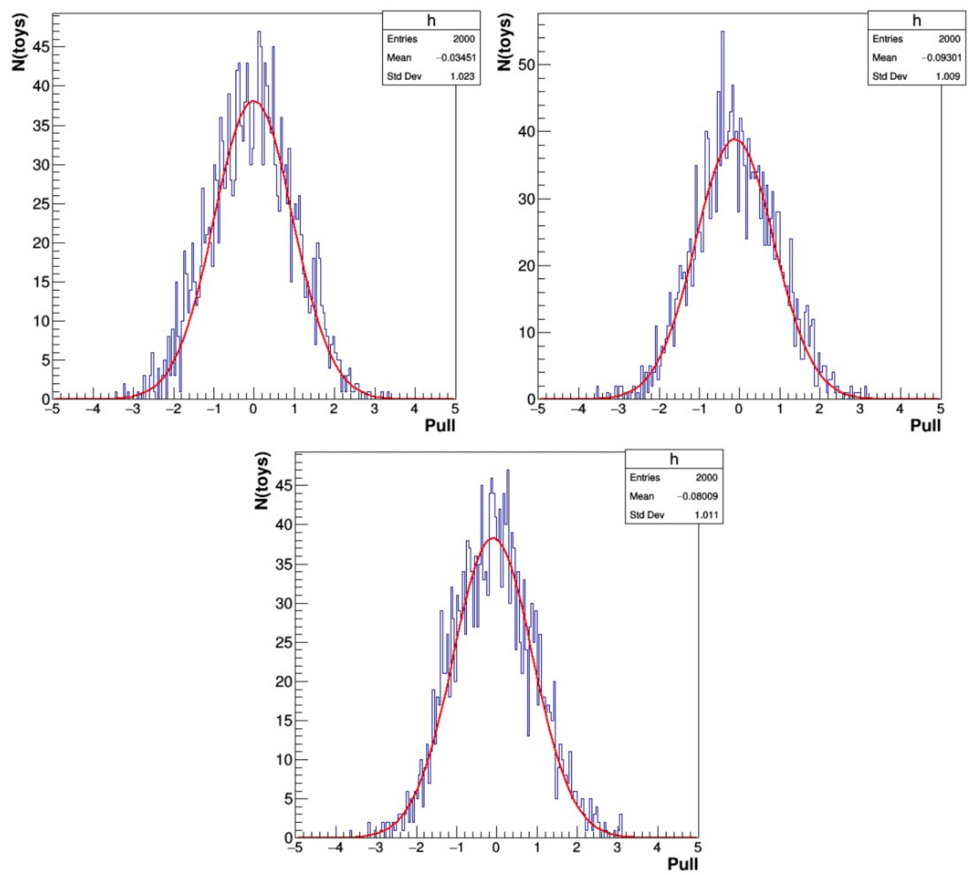

Figure 4·49: Histogram of the measured signal strength for 2000 toy analyses, generated from background only fits of the data at masses of 0.246 GeV (top left), 1.478 GeV (top right), and 7.262 GeV (bottom). Each distribution is fit by a Gaussian function, the mean of which describes the bias of the function used in this fit.

We consider a wide set of possible configurations for the bias measurements to ensure that there are no edge cases where bias may become a significant effect. These include measuring the bias of: a function relative to toys generated from its own background-only fit, a function relative to toys generated from a different functions background-only fit, and the behavior of the full function profile when fit to toys from a variety of different functions. We additionally experiment with injecting a fixed amount of signal into the toys to see if the fit can accurately resolve the true value. The bulk of these studies are included in Appendix C.

Figures 4·50 and 4·51 show the most directly relevant measurements for this analysis, which is the bias of the combined background function profile when fit to toys generated from each of the constituent functions, with and without signal injected, for both analysis categories. The bias is presented as the magnitude of the average deviation in the signal measurements, divided by the statistical uncertainty in the signal measurement. Even in the most unfavorable combinations shown, the projected impact of bias on the signal measurement is never more than $\sim 30\%$ of the statistical uncertainty, which impacts the total uncertainty by less than 5% when the two quantities are quadrature summed. Because of this, we determine that bias from our signal and background is a negligible effect on the analysis, and no mechanisms need to be implemented to model or counteract it.

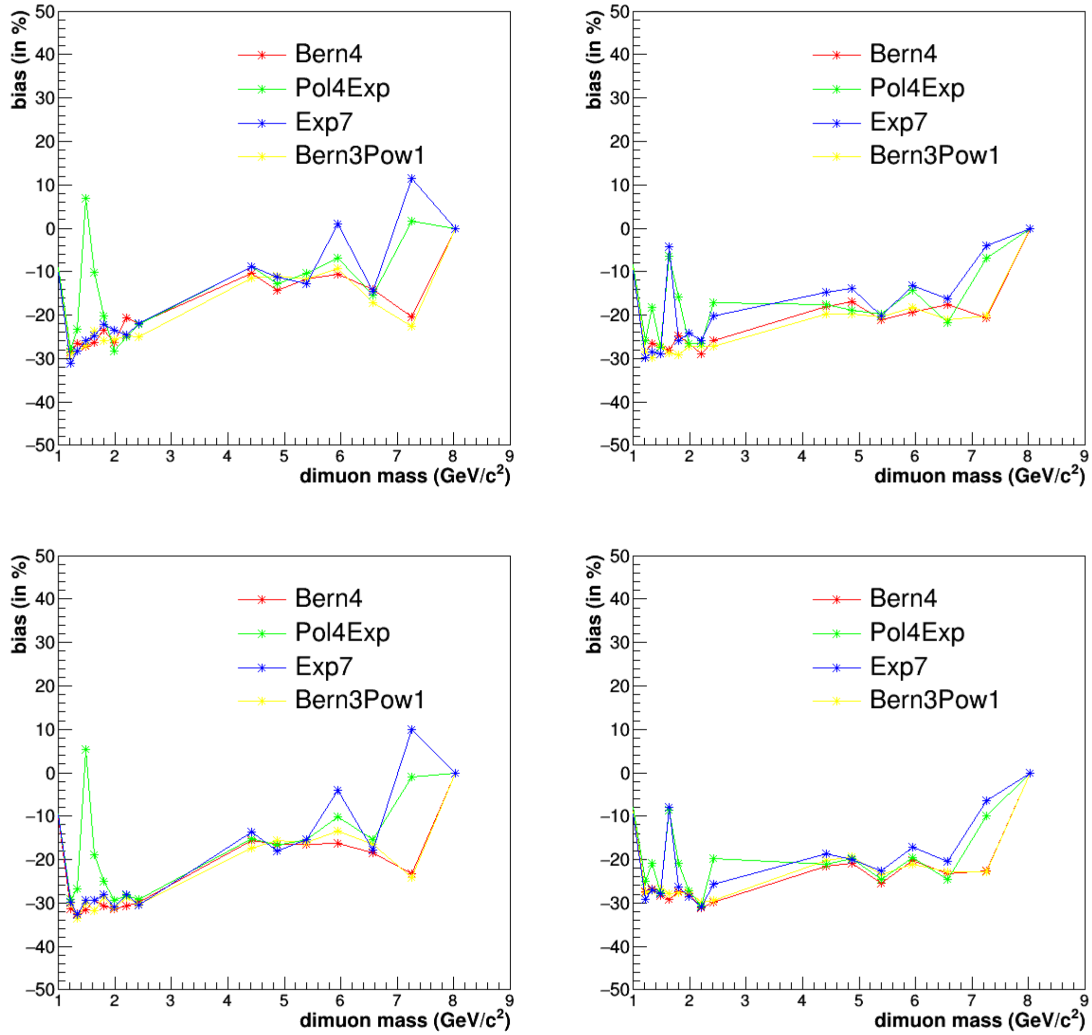

Figure 4·50: Bias distribution for the combined background function profile using toys drawn from 2017 (left) and 2018 (right) mass windows in the inclusive category. The top row shows the bias with zero signal present in the toys. The bottom row shows the bias when a small signal is injected into every toy.

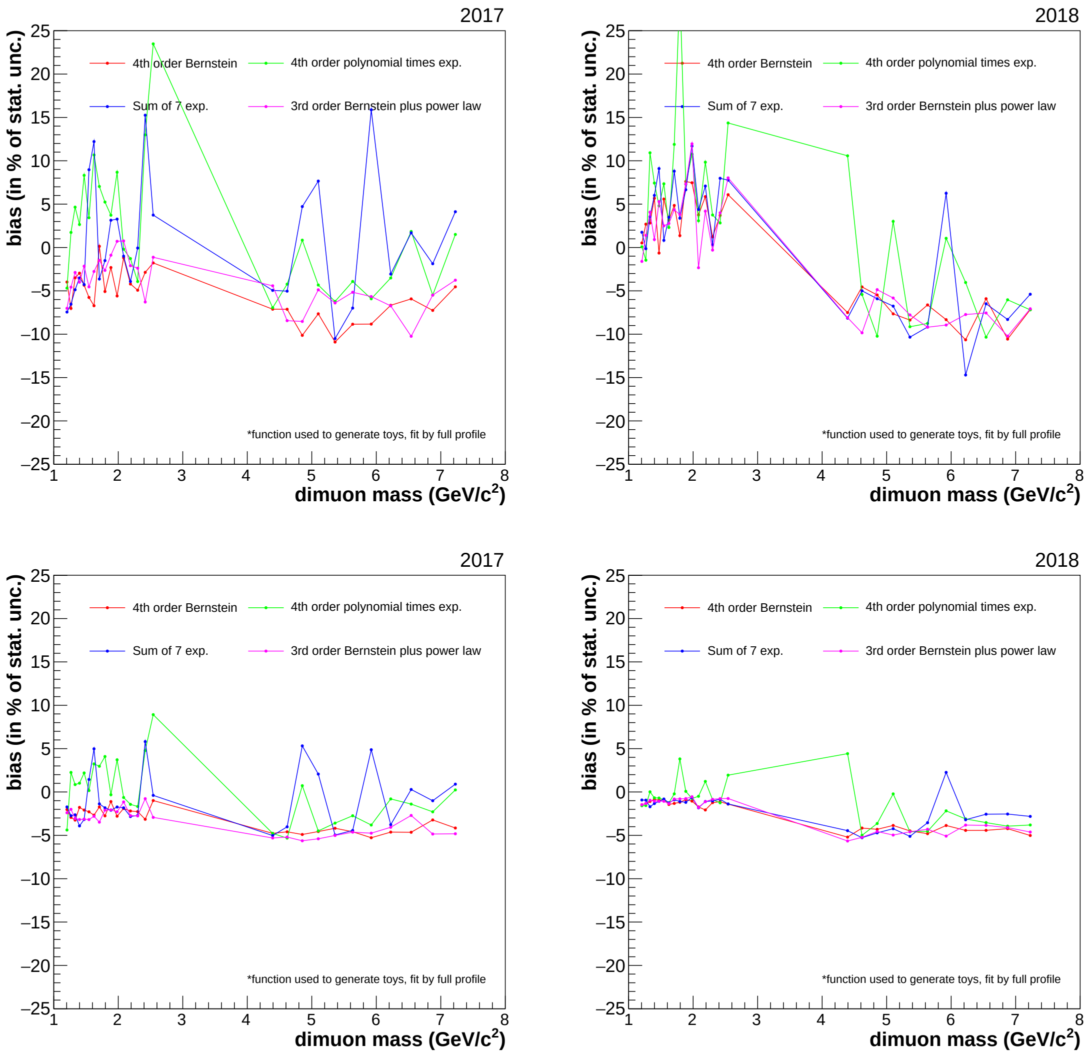

Figure 4·51: Bias distribution for the combined background function profile using toys drawn from 2017 (left) and 2018 (right) mass windows in the boosted category. The top row shows the bias with zero signal present in the toys. The bottom row shows the bias when a small signal is injected into every toy.

4.7.3 Peaking Background from D^0 Meson

When examining the F-test results and the scatter plots of the background only goodness-of-fit measurements, it is clear that there is some general poor behavior of our background models primarily between 1 and 2 GeV. Upon initial unbinding of this analysis, two significant excess were observed in this region; a 5.2σ excess at 1.58 GeV and a 4.6σ excess at 1.74 GeV, in the inclusive category. The windows that manifested these excesses we subjected to exhaustive studies, to verify that the excesses were real effects in the data, and not simply the result of a failed background fit or a poorly modeled feature in the dimuon mass spectrum. A visualization of the 1.54 GeV window is shown in Figure 4·52.

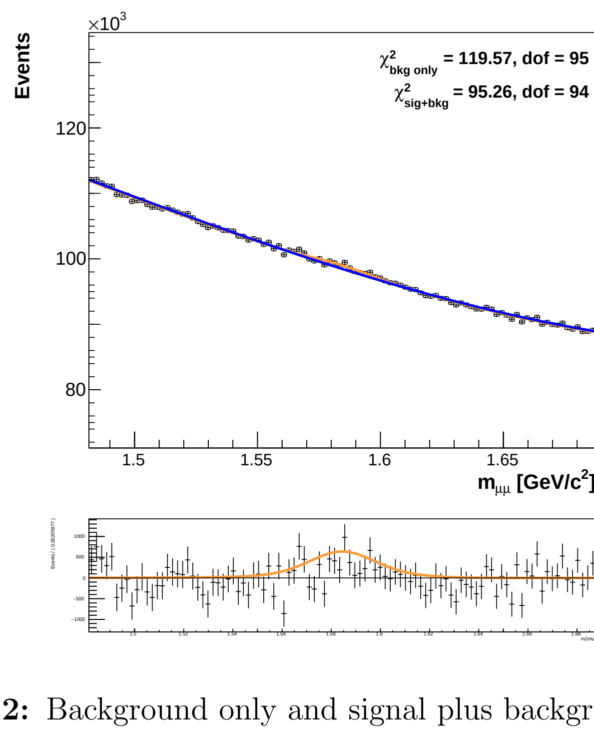

Figure 4·52: Background only and signal plus background fit for the window containing the 5.2σ excess in the inclusive category.

It was determined that both excesses could be attributed to real peaks in the data, resulting from the decay of SM D^0 mesons into charged kaons and pions. When kaons and pions travel through the detector, they often decay into a muon with a track very close to that of the charged meson, which may be mistakenly reconstructed as a single prompt muon track. If the D^0 decays into a pair of opposite charge kaons (K^+K^-) or an opposite charged kaon and pion ($K^\pm\pi^\mp$), there is a chance that both decay-in-flight particles will be reconstructed as muons that satisfy our analysis criteria and are matched to the same dimuon vertex. For this fake dimuon pair, the invariant mass calculation performed by the software will be incorrect, as the reconstructed 4-vectors assume the rest mass of both particles to be the muon mass ($\sim$105 MeV), rather than the significantly higher kaon mass ($\sim$493 MeV). This results in the reconstructed mass of the parent particle being shifted downwards according to the formula:

$$M_{\mu\mu} = \sqrt{M_{AB} + 4m_\mu^2 - (m_A + m_B)^2}, \tag{4.17}$$

where m_A and m_B are the masses of both of the final state particles to be mis-identified as muons, M_{AB} is the true invariant mass of the parent particle that decays into A and B, and $M_{\mu\mu}$ is the mass where this particle will appear when mapped onto the dimuon mass spectrum in our data. For the case where the D^0 decays into two kaons, both of which are mis-identified as muons ($M_{AB} = 1864$ MeV, $m_A = m_B = 493$ MeV), the shifted invariant mass is roughly 1.58 GeV, the same as the 5.2σ excess. If one of the final state mis-identified particles is a kaon and the other is a pion ($m_A = 493$ MeV, $m_B = 139$ MeV), the reconstructed mass is 1.74 GeV, coinciding with the other excess. Though the odds of two particles being mis-identified this way simultaneously is very small, D^0 mesons are copiously produced in LHC collisions and a significant number of such events may accumulate in our data. Given the precise agreement between the shifted D^0 mass calculations and the observed masses of these excesses in the data, we identify these peaks as SM backgrounds from D^0 decays.

Generally this analysis makes no attempt to measure a signal on the peaks of any other SM resonant process. However, given the comparatively small yield of these

D^0 peaks it is feasible to estimate the expected number of fake signal events and account for this contribution in our background model. The strategy to do this is to add a peaking signal-like shape into the background function at the precise masses of the excesses and allow the amplitude to float as part of the ML fit. This would accommodate the excesses, but also remove any hope of measuring a BSM signal at that mass, since the peaking background function could inflate itself to absorb any real signal that coincided with that mass. To allow ourselves to still set some constraints on BSM production in the vicinity of these peaks, we seek to identify a control region (CR) in the data that is nearly entirely void of events from our desired signal processes but still includes a contribution from the $D^0 \to KK/K\pi$ processes. By measuring the yield of the D^0 processes in the CR and determining the expected ratio of these events in the signal region (SR) versus the CR, we can constrain the amplitude of the peaking background shape in the SR to match the expected yield of D^0 events.

The best variable to define the CR is found to be $\sigma_{L_{\mathrm{xy}}}$. Figure 4·53 shows the distribution of $\sigma_{L_{\mathrm{xy}}}$ in $D^0 \to KK/K\pi$ MC samples where both mesons are reconstructed as muons, compared to the distribution in data and the samples of the relevant signal processes. Some of the samples included have the kaon and pion lifetime artificially decreased, to enhance the rate of reconstructing a fake muon. The fraction of DY and pseudoscalar events above the analysis cut of $\sigma_{L_{\mathrm{xy}}} < 3.5$ is nearly zero, while the D^0 processes have significantly longer tails. The boundary for the CR that most enhances the contribution from D^0 events relative to continuum background is found to be $3.5 < \sigma_{L_{\mathrm{xy}}} < 11$. Further studies on the optimizations of the CR are included in Appendix D.

The D^0 peaks are not well described by the nominal signal shape used for this analysis and each requires a custom functional form to capture its distribution. Fig-

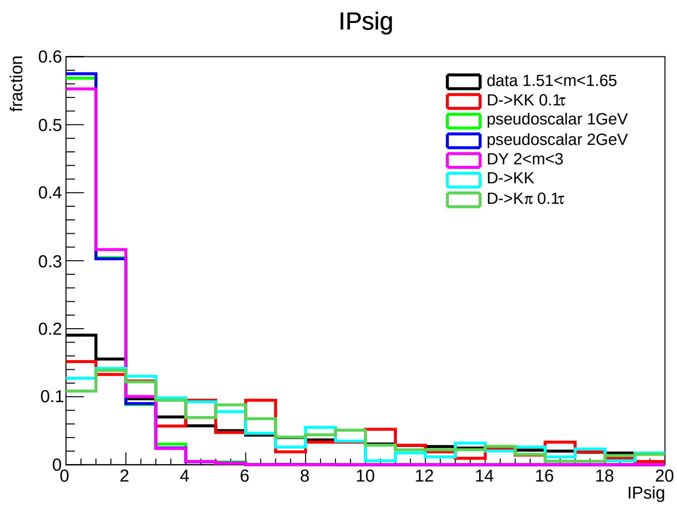

Figure 4·53: Distribution of $\sigma_{L_{\rm xy}}$ in data, DY signal MC, pseudoscalar signal MC, and different samples of D^0 mesons decaying into kaons/pions mis-identified as muons. Labels in the legend that include "0.1τ" indicate that the pion and kaon lifetime in that sample has been decreased, to improve the efficiency of the muon mis-identification.

ure 4·54 shows the mass spectrum of the $D^0 \to KK/K\pi$ MC samples, and their respective best-fit functions. The $D^0 \to KK$ peak is found to be well described by a Voigtian function, and the $D^0 \to K\pi$ peak is fit to a bifurcated Gaussian. Figure4·55 shows that the expected shape of these distributions does not vary much between the SR and CR, so the same function can be used for both regions when fitting the data.

To constrain the yield from the D^0 decays in the SR, we need to know the expected number of fake events in the SR relative to the CR. We derive this transfer factor from the $D^0 \to KK/K\pi$ MC samples, by taking the ratio of the events with $3.5 < \sigma_{L_{\rm xy}} < 11$ to events with $\sigma_{L_{\rm xy}} < 3.5$ in each sample. For the $D^0 \to KK$ process, this value is: $N(\mathrm{CR})/N(\mathrm{SR}) \approx 0.88$. The $D^0 \to K\pi$ process has a value of $N(\mathrm{CR})/N(\mathrm{SR}) \approx 1.09$. Because both of these values are derived from MC, we must account for the possibility that the there is some different between the distribution of $\sigma_{L_{\rm xy}}$ in data and simulation, and apply an uncertainty for the mismodeling in the transfer factor. Additionally,

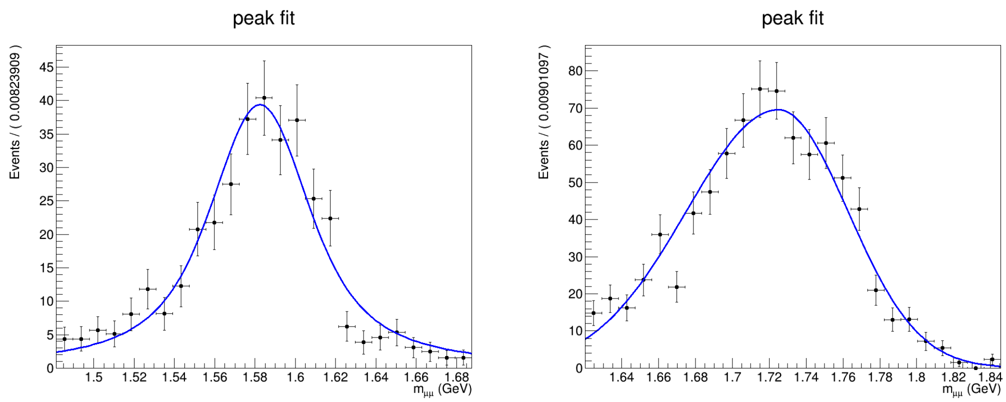

Figure 4·54: Mass spectrum of D^0 decays for the KK (left) and $K\pi$ (right) channels in MC. The KK peak is fit with a Voigtian function, and $K\pi$ with a bifurcated Gaussian.

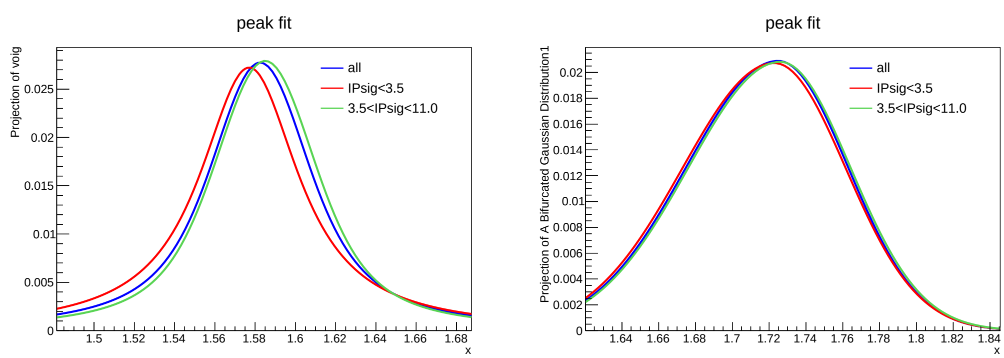

Figure 4·55: Fits of $D^0 \to KK$ (left) and $D^0 \to K\pi$ (right) modeling functions to the SR and CR distributions, as well as the entire sample. Little variation is observed in the shape between different $\sigma_{L_{\mathrm{xy}}}$ regions.

the measurement of the $K\pi$ transfer factor depended entirely on D^0 samples with shortened meson lifetimes in order to accrue sufficient statistics, which may also have influenced the distribution of $\sigma_{L_{\mathrm{xy}}}$. We assign uncertainties on the transfer factors due to both of these effects, which are discussed in Section 4.8.9.

Figure 4·56 shows the signal plus background fits in the window containing the 5.2σ excess, now with the addition of the new peaking background function. The figure shows the fits for both years, in both the signal and control regions, and illustrates that the vast majority of the peaking structure in this window can be attributed to the $D^0 \rightarrow KK$ process, indicated by the green line. Figure 4·57 shows the same fits for the window with the 4.6 sigma excess; again nearly all of the peaking contributions are fully described by the addition of the new background shape and control region. With the implementation of these new background modeling techniques, both of the significant excesses in the model-independent limit for the inclusive category are greatly reduced, and both are fully attributed to this previously unmodeled SM background. Appendix D provides comparisons of the model-independent limit before and after applying the peaking background models, as well as example fits for windows where the peaking backgrounds are included, but are not located at the central mass of that window.

The presence of the D^0 peaks is a feature that is only represented in the inclusive analysis category. The model-independent limit for the boosted category was unblinded at the same time and notably had no indication of excesses at 1.58 or 1.74 GeV. This is due to the fact the additional $p_{\mathrm{T}\mu}$ and $p_{\mathrm{T}\mu\mu}$ cuts applied to the boosted category are extremely inefficient for the $D^0 \rightarrow KK/K\pi$ events. These cuts accept only $\sim$2% of events in the $K\pi$ channel and $< 1\%$ of events in the KK channel. For consistency, the same SR+CR fits are applied to the effected windows in the boosted category, but no contribution from the D^0 peaks are observed as shown in Figure 4·58. Note that while both of these windows do include a small local excess, the green lines, which are proportional to the yield in the CR, are both nearly flat.

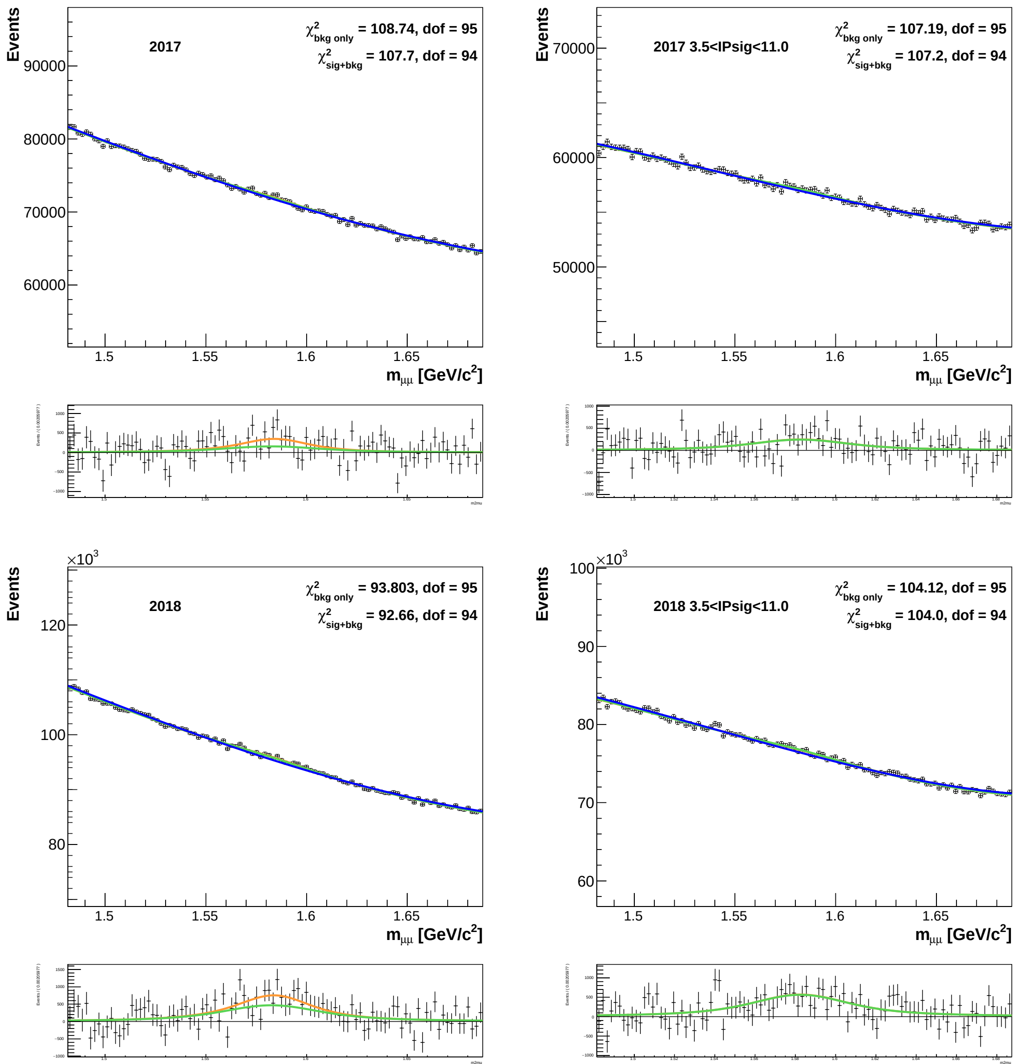

Figure 4·56: Simultaneous fits of of the SR (left) and CR (right) to 2017 (top) and 2018 (bottom) data in the 1.58 GeV mass window. The orange lines in the SR plots represent the best fit signal values, while the green lines are the expected contribution from $D^0 \rightarrow KK$, based on the yield measured in the CR.

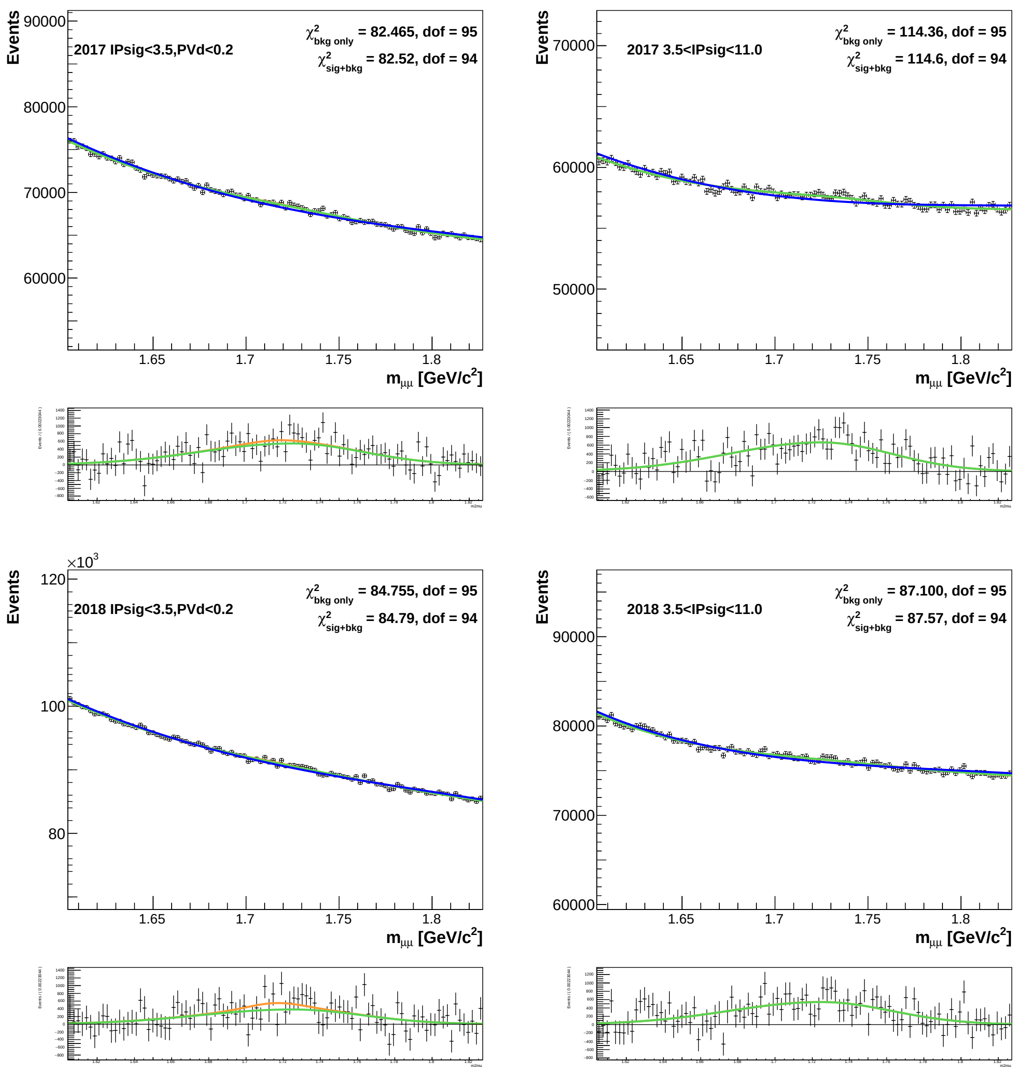

Figure 4·57: Simultaneous fits of of the SR (left) and CR (right) to 2017 (top) and 2018 (bottom) data in the 1.72 GeV mass window. The orange lines in the SR plots represent the best fit signal values, while the green lines are the expected contribution from $D^0 \rightarrow K\pi$, based on the yield measured in the CR.

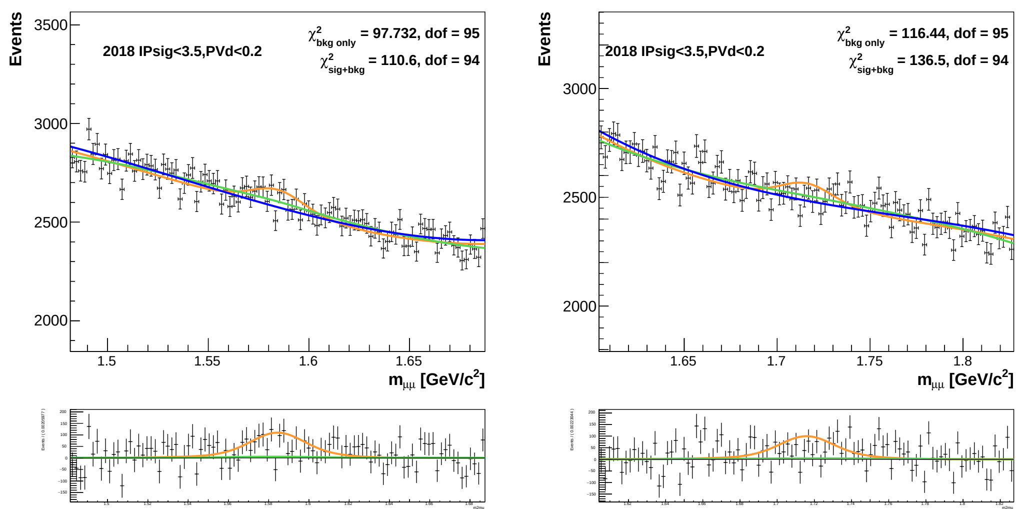

Figure 4·58: Fits of of the SR in 2018 data for the 1.58 (left) and 1.72 (right) GeV mass windows in 2018. The green line being flat in the lower pad indicates negligible measurement of $D^0 \to KK/K\pi$ events in the CR.

4.8 Systematic Uncertainties

Numerous systematic effects impact our confidence in the signal measurement during the fit. The leading sources of uncertainty are described in this section, as well as rationalizations for why certain effects are neglected. A summary of all major systematic effects is presented in Table 4.8.

4.8.1 Integrated Luminosity

To convert the measured post-fit signal contribution into a constraint on a production cross section, we must have a high degree of confidence in the total number of pp collisions represented in our data. The total integrated luminosity recorded in each year of data taking is studied by dedicated working groups within the CMS collaboration, who report uncertainty ranges on this quantity. These groups assign an uncertainty of 2.3% for 2017 data [85] and 2.5% for 2018 [86]. These values conflate the uncertainty due to correlated and uncorrelated systematics between the two years. Roughly half of the total uncertainty from each year is due to effects that are correlated, which results in an "effective" combined uncertainty of $\sim$2.0% on the full integrated luminosity.

4.8.2 Dimuon Scouting Trigger Efficiency

Uncertainty on dimuon yield due to incorrectly measuring the efficiency of the dimuon scouting trigger is a non-trivial effect and must be accounted for. The uncertainty is conservatively estimated by comparing the efficiency when measured with respect to data verses MC as described in Section 4.6. The percent value of this effect is $\left|1 - \frac{eff(\text{Data})}{eff(\text{MC})}\right| \times 100\%$. The magnitude of this value depends on the dimuon mass being considered, and ranges from 1% up to $\sim 16\%$.

4.8.3 Vertex Selection Efficiency

The uncertainty on the efficiency of the displacement vertex requirements are determined by comparing the the difference in performance of the cut when applied to resonances at different masses. For the requirement $L_{\mathrm{xy}} < 0.015$ cm, the nominal data/MC scale factor is ~ 0.93 when measured using the Υ peak and ~ 0.90 when measured with the J/Ψ peak. The percent difference between these values is roughly 3%, which we apply for the inclusive category above 4 GeV, where this cut is applied.

For the selection $\sigma_{L_{\mathrm{xy}}} < 3.5$, the efficiencies in data and MC are in near perfect agreement and the efficiency of the cut is itself nearly 1.00, so we choose not to assign an uncertainty.

4.8.4 BDT Selection Efficiency

The error in the measured efficiency of the BDT identifier is estimated similarly to the trigger efficiency, by taking the percent difference when it is measured with data and MC. This effect also varies by mass and is generally less than 10%. However, because the performance of the BDT is highly dependant on the distribution of the variable L_{xy} in training data, we additionally seek to model the impact on the efficiency if there is a small change in the average displacement in the data.

This study is indented to counteract an intrinsic weakness of using the TnP method with SM resonances to develop and qualify our selection, which is the fact that it includes only signal events in a narrow range of masses. L_{xy} varies as a function of mass in data and there may be some disagreement between the BDT performance when measured on a resonance versus one of the analysis mass hypotheses. Even weighting the application of the efficiency to match the expected dimuon kinematics in signal MC at a given mass does not entirely address this effect as, even within a given range of ΔR, there is disagreement in L_{xy} between samples of different masses.

This effect is illustrated in Figure 4·59, and is found to be a $\sim 10\%$ difference in the average value of L_{xy}.

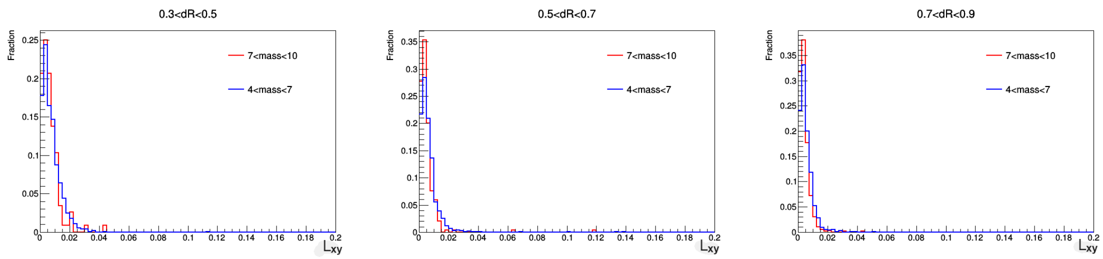

Figure 4·59: Distribution of L_{xy} in MC events with ΔR values of [0.3,0.5] (left), [0.5,0.7] (center), and [0.7,0.9] (right). The red line corresponds to a higher mass range, which represent events used for training and studying the BDT. The blue line is a range of lower masses, more like what will be used for this analysis.

To estimate the extent of this effect on the BDT efficiency, we manually shift the value of L_{xy} in all events in the TnP samples up by 10% and repeat the efficiency measurement. We again estimate the uncertainty by taking the percent disagreement between this efficiency measurement and the nominal efficiency measured in Section 4.6, finding it to be $< 3\%$ for events with $\Delta R > 0.5$ and $< 8\%$ for $\Delta R < 0.5$. This uncertainty is quadrature summed with the uncertainty from disagreement with MC, yielding a final BDT efficiency uncertainty ranging between $4 - 20\%$ depending on the mass being considered.

4.8.5 Muon Efficiency Application

Our measurement of the BDT performance via TnP depends entirely on the assumption that the quality of the two muons is completely uncorrelated, such that the sample of probe muons in the passing and failing samples are completely unbiased, i.e. a higher quality tag muon cannot "help" its corresponding probe to satisfy the BDT criteria. This assumption is not entirely true for the selection, given that one of theBDT contains vertex information, which is a strictly correlated aspect of the muons. Additionally, when the muons are very close together, the relative isolation cones can overlap, further correlating the quality. These variables are strong discriminants for the BDT, so it is not desirable to remove them from the analysis. Rather, we estimate and apply another uncertainty to allow for our mis-modeling in the trigger efficiency measurement strategy.

If our measurement of the probe muon efficiency was fully uncorrelated with the behavior of the tag, we could treat it as a per-muon efficiency and simply apply two factors of the efficiency to estimate the likelihood of the BDT passing a signal dimuon event. We call this value the "calculated efficiency," (ε_{μ}^{2}) representing our purely data driven measurement of the BDT performance. To gauge the extent to which correlated variables are impacting the efficiency, we use signal MC to contrast this value with the "true efficiency," ($\varepsilon_{\mu\mu}$). The true efficiency is the ratio of dimuon events in a given sample where both muons pass the BDT selection, divided by the total number of dimuon events that pass the analysis fiducial cuts and vertex selection. If the correlated variables have a significant impact on the BDT efficiency, we will expect the calculated efficiency to be an underestimation relative to the true value, as it is penalized twice for missing events where the correlated variables performed poorly. Table 4.7 shows values of the calculated and true efficiencies for pseudoscalar and DY samples at a range of masses, as well as the ratio of the two. Predictably, at lower masses, where the isolation cones are more overlapped, the correlation is greater and the discrepancy is increased. We assign an additional uncertainty on the efficiency of the BDT selection equal to the percent differences in these quantities. This effect is approximately 4% for the Υ trained ID and 8% for the J/Ψ trained ID.

4.8.6 Signal Resolution

As described in Section 4.7.1, the width of the signal hypothesis is fixed at 0.013 times the central mass value of the window being analyzed, determined via fits to the SM

Sample	Calculated Efficiency (ε_μ^2)	True Efficiency ($\varepsilon_{\mu\mu}$)	Ratio
pseudoscalar 1 GeV	80.2%	88.2%	0.91
pseudoscalar 2 GeV	86.0%	91.1%	0.94
pseudoscalar 5 GeV	84.1%	86.5%	0.97
pseudoscalar 8 GeV	87.3%	91.3%	0.96
DY $2.0 < m < 4.0$ GeV	85.1%	92.2%	0.92
DY $6.0 < m < 8.0$ GeV	87.8%	90.8%	0.96

Table 4.7: The calculated and true ID efficiencies of the dimuon events in a variety of signal MC samples.

resonances. We assign an uncertainty of 20% on this value, which encompasses all variation between the different resonances considered, as well as the slight systemic shift observed between analysis categories.

4.8.7 Choice of Signal Model

The decision to use a dCB plus Gaussian function to model the signal contribution is motivated entirely by the apparent goodness of fit to the SM peaks, and low bias when studied alongside the candidate background functions. However, it is possible that it is not the most accurate descriptor of the form of a resonance in our data, and may thus misconstrue the number of signal events. We estimate the extent of this effect by comparing the measured signal yield of our model when fitting the SM resonances, to the yield when measured with an alternate function, which in this case is the sum of two Gaussian distributions. The differences in these values is small, ranging from $1.3 - 2.7\%$ in for masses less than 4 GeV, and $1.4 - 1.7\%$ for masses above 4 GeV. We include these values as nuisance parameters in the ML fit, similarly to any other parameter which impacts the expected rate of dimuon events.

4.8.8 Choice of Background Function

The rigorous bias studies described in Section 4.7.2 and Appendix C indicate that bias introduced by our choice of background function is a negligible effect towards our ability to accurately resolve the presence of signal in the data. Regardless, though no quantitative uncertainty is stated, the inclusion of our four distinct background functions in the ML fit acts as an additional nuisance parameter. This will affect the shape of the likelihood function response when scanning across different signal strength hypothesis, as shown in Figure 4·60. Note that the 4th order Bernstein polynomial (red) dominates the profile for most signal hypotheses, but is overtaken by other functions at certain values. Conversely, sum of 7 exponentials has very little impact on the profile as a whole, because it is significantly penalized by the discrete profiling technique for its high number of free parameters.

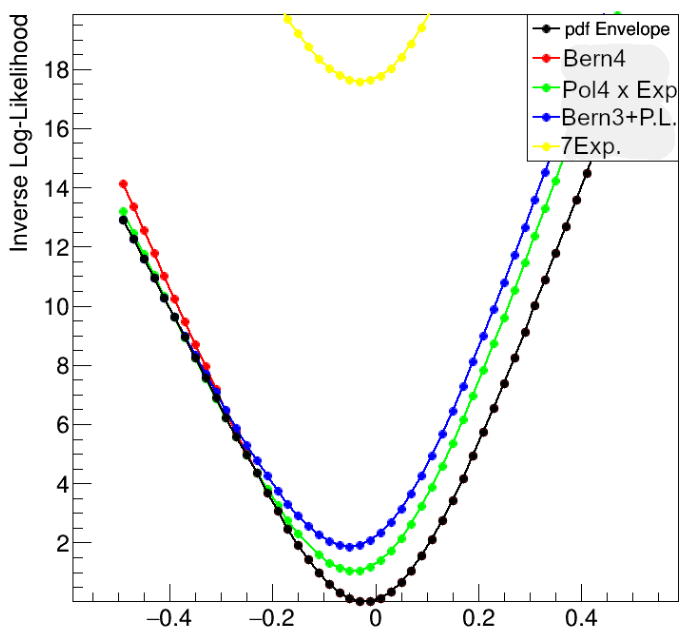

Figure 4·60: Likelihood scan across signal strength value for an example mass window with the complete background function profile (black). Colored lines represent scans where all but one of the background candidates are removed from the fit, as indicated in the legend.

4.8.9 D^0 Transfer Factor

In the windows that contain some contribution from the D^0 peaks, the amplitude of the peaking background is fixed to exactly the yield in the CRs times the corresponding transfer factors determined from MC. Due to the extremely low rate of pions and kaons to fake muons in the simulation it is difficult to accrue many events, and we must assign a statistical uncertainty of 11% to both the $D^0 \to KK$ and $D^0 \to K\pi$ transfer factors.

There is also a possibility that the reconstruction of $\sigma_{L_{\text{xy}}}$ in simulation is not the same as in data, meaning our CR definition in the data is not perfectly described by the MC. We quantify this difference in $\sigma_{L_{\text{xy}}}$ by comparing J/Ψ events in data and MC, with the same non-prompt background subtraction scheme used in Section 4.6.3. Figure 4·61 shows the distribution of $\sigma_{L_{\text{xy}}}$ in both of these samples, before and after the background subtraction. While this is not descriptive of the $\sigma_{L_{\text{xy}}}$ distribution in the D^0 peaks, we assume that the mismodeling of $\sigma_{L_{\text{xy}}}$ will be mirrored in the true $D^0 \to KK(K\pi)$ samples we use when determining the transfer factor. We determine a proxy transfer factor from both samples by taking the ratio of events above (CR) and below (SR) $\sigma_{L_{\text{xy}}} = 3.5$. The ratio in data and MC disagree by roughly 16%, which we treat as an additional uncertainty on the transfer factor.

The final effect we consider is specific to the $D^0 \to K\pi$ sample used to determine the transfer factor for the 1.79 GeV excess. The efficiency of a simulated pion to successfully fake a muon is even lower than for a kaon, making it extremely impractical to generate sufficient statistics to model this process. To enhance the rate of fake pions in the sample, we decrease the lifetime of the simulated mesons to 1/10th of their true lifetime. To estimate the extent to which this causes mismodeling of $\sigma_{L_{\text{xy}}}$, we apply the same lifetime shortening to the $D^0 \to KK$ process and compare the transfer factor to the full-lifetime sample. The normal $D^0 \to KK$ sample has a transfer factor of

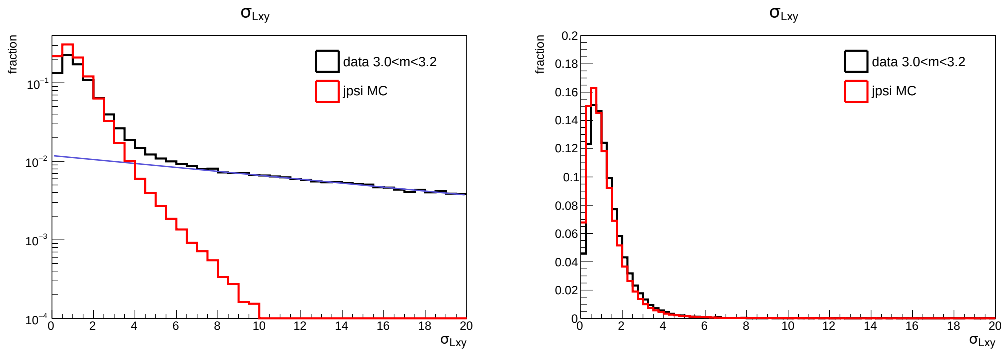

Figure 4·61: Comparison of $\sigma_{L_{\mathrm{xy}}}$ distributions in data and MC before (left) and after (right) subtracting the displaced background contribution from data. Background in modeled by an exponential fit to the region $8 < \sigma_{L_{\mathrm{xy}}} < 20$.

0.88, while the 1/10th lifetime sample has a transfer factor of 0.97, giving a difference of roughly 10%. Because it is not certain that the kaon and pion reconstructions are effected in the same way by this operation, we cautiously inflate this effect to 15% when applying it to the $K\pi$ transfer factor.

The total uncertainty on the KK transfer factor is the quadrature sum of the statistical uncertainty and the $\sigma_{L_{\mathrm{xy}}}$ mismodeling: $\sqrt{0.11^2 + 0.16^2} \approx 20\%$. Similarly, the total uncertainty on the $K\pi$ transfer factor is sum of those effects, plus the lifetime mismodeling: $\sqrt{0.11^2 + 0.16^2 + 0.15^2} \approx 25\%$.

4.8.10 Systematic Effects Summary

Effect	$m_{\mu\mu} < 2.6$ GeV	$m_{\mu\mu} > 4.2$ GeV
Integrated luminosity	2.3–2.5%	
Trigger efficiency	1–16%	
Vertex selection	—	3%*
Muon ID efficiency	4–9%	12–20%
Efficiency application	8%	4%
Mass resolution	20%	
Signal shape	1.3–2.7%	1.4–1.7%
D^0 meson normalization	20–25%	—

Table 4.8: Summary of the experimental systematic uncertainties for a signal model in the model independent search for a dimuon resonance. Uncertainties are similar between both analysis categories, with the exception of the vertex selection uncertainties, which is not applied in the boosted category.

4.9 Results

This section will briefly describe studies of the ML fit after unblinding the data, as well as the model independent limits on the BSM production cross-section, and constraints on our targeted benchmark model parameters. Conventionally, the unblinded results should only be observed after the selection and modeling methods have been frozen, such that analysis parameters cannot be fine-tuned to enhance sensitivity without proper motivation. Given the novelty of this analysis and some unanticipated features in the data, both analysis categories were subjected to small, targeted adjustments after initial unblinding to accommodate pathological behaviors found in the limits. The inclusive analysis was augmented to model the unexpected presence of $D0$ resonances below 2 GeV (see Section 4.7.3 and Appendix D), and the boosted category was restructured to use a larger mass window for the sideband fit to correct unnatural noisy behavior in the limit resulting from low statistics (see Appendix B). Neither of these post-unblinding adjustments had a significant effect on the expected sensitivity, and in both cases served only to normalize the analysis, reducing the significance of what were previously the largest excesses.

4.9.1 Nuisance Parameter Pulls

Figure 4·62 shows the pull of the highest ranking nuisance parameters included in the ML fit, listed according to their impact on the observed limit at two representative mass points in the inclusive category. For these example fits, the forth-order Bernstein polynomial is the best performing background function for both years, so its free parameters have the strongest direct impact on the signal measurement. The next most dominant effect is the variation in the width of the signal shape, which is allowed to float within the assigned 20% uncertainty. When considered together, the uncertainties on the efficiencies and total integrated luminosity also have a small effect on the fit. Typically, the parameters included in the 2017 fit rank lower than their counterparts in 2018, as the sample contains roughly half as much data. All parameters are uncorrelated between the years except for the signal strength, signal resolution, and a portion of the luminosity uncertainty.

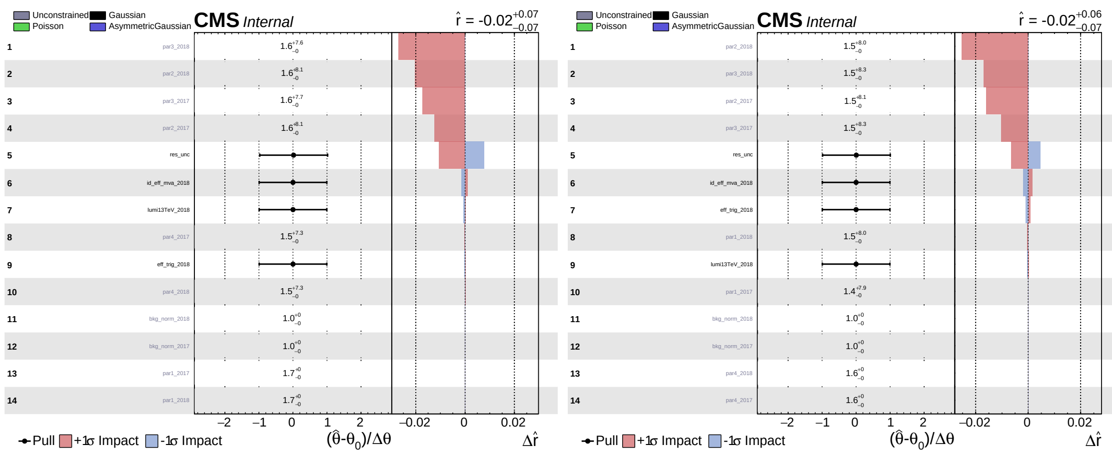

Figure 4·62: Impact ranking of nuisance parameters from example windows with central masses of 2.2 GeV (left) and 6.0 GeV (right). Both plots represent simultaneous fits to the 2017 and 2018 data in the inclusive category.

Figure 4·63 shows an alternate visualization the effect of the nuisance parameters within these windows, via the shape of the likelihood profile when scanned across different signal values. The black line represents the true signal measurement with all uncertainties included, and the lowest point represents the signal value that maximizes the likelihood function. We then repeat the fit and successively freeze any variation in the leading uncertainties; first signal resolution (blue), then all efficiency effects (violet), then all remaining systematics (red). Visibly, the greatest effect comes from removing the resolution uncertainty, with some additional adjustment from the efficiency uncertainties, and little change attributed to the other uncertainties. For the scan corresponding to the 6.0 GeV window, the freezing of the signal resolution has caused a small shift in the best fit value of the signal strength.

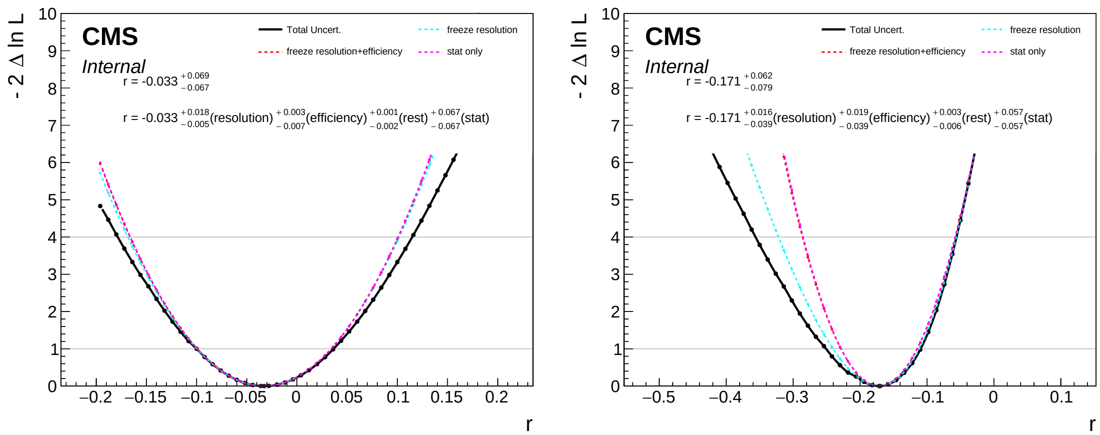

Figure 4·63: Inverse log likelihood scanned around the best fit value of the signal strength for windows with central masses of 2.2 GeV (left) and 6.0 GeV (right). Both plots represent simultaneous fits to the 2017 and 2018 data in the inclusive category. Different colors correspond to the likelihood profile with different combinations of nuisance parameters frozen.

4.9.2 Model-Independent Limits

The model-independent limits on BSM production cross-section, times branching fraction to muons, times the fiducial acceptance of our experimental volume ($p_{\mathrm{T}} > 4$ GeV and $|\eta| < 1.9$ inclusive; $p_{\mathrm{T}} > 5$ GeV, $p_{\mathrm{T}_{\mu\mu}} > 20$ (35) GeV, and $|\eta| < 1.9$ boosted) are shown in Figures 4·64 and 4·65 for the inclusive and boosted categories respectively at a confidence-level (CL) of 95%. Both limits are presented as an analysis of 2017 and 2018 data separately, as well as a combined analysis where the signal fits are fully correlated, which is the main result. The largest excess is seen in the boosted category at a dimuon mass of 2.41 GeV, with a local significance of 2.6σ.

Both analysis categories exhibit an unusual behavior, where the expected limit will, on occasion, sharply increase for a number of mass points, before returning to the nominal value. This effect occurs in windows where the ML fit favors a background function other than the forth-order Bernstein polynomial. This discrete change in the background function has a slight impact on the sensitivity of the analysis in those

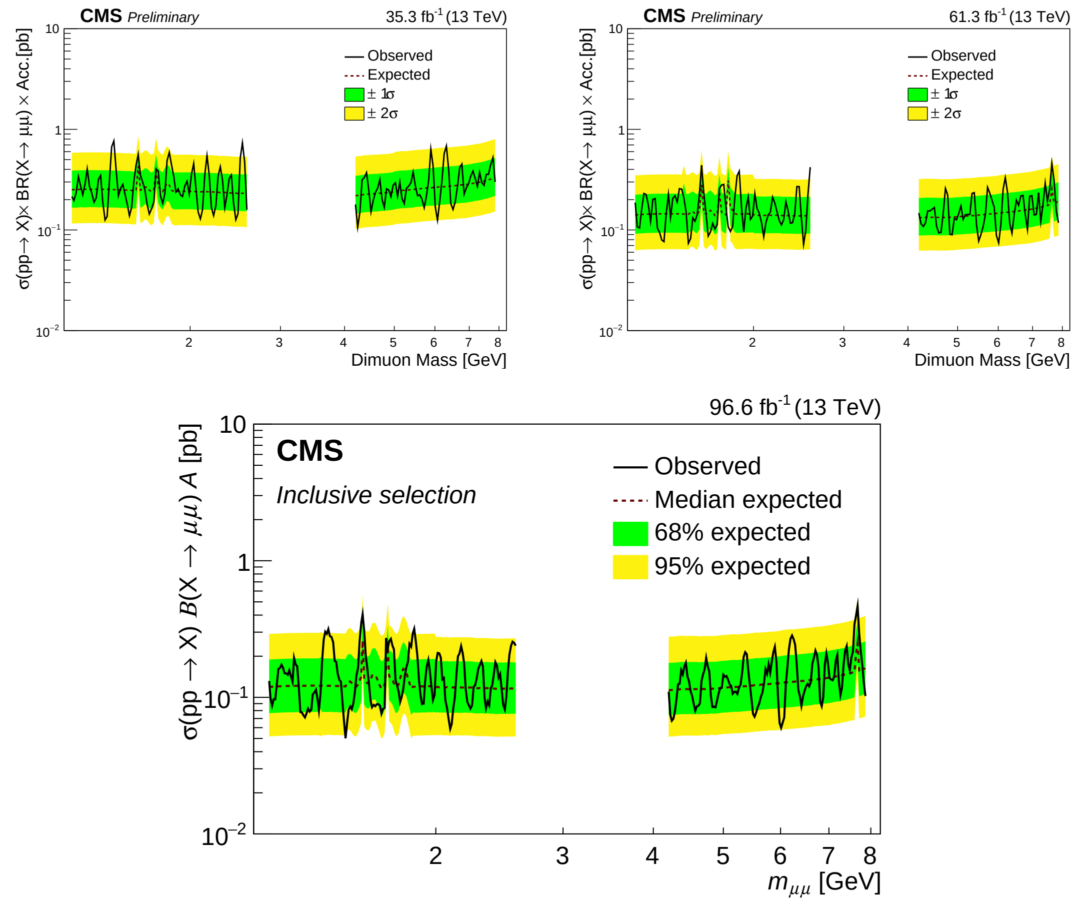

Figure 4·64: Model independent limit on BSM resonant dimuon production in the inclusive category at 95% CL. Limit is shown for an analysis with just 2017 (top left) and 2018 (top right) data, and both years combined (bottom).

windows. Because $B_4(x)$ generally has the best fit, and is least penalized by the discrete profiling method for its small number of free parameters, these jumps will likely only occur in windows that contain some strange feature in the data that is fit exceptionally well by another function. The behavior is most pronounced in the windows that include the fake $D0$ resonances, as the background shape is further augmented by the inclusion of the peaking background shape.

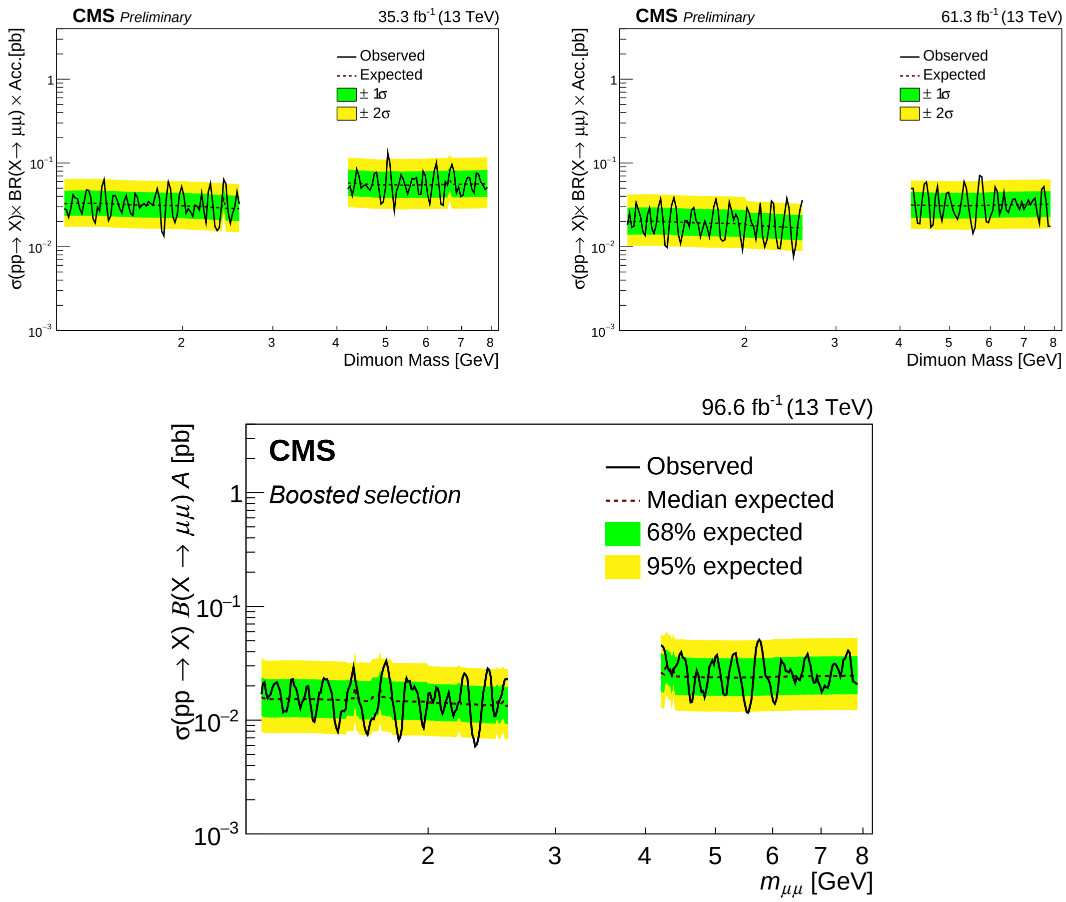

Figure 4·65: Model independent limit on BSM resonant dimuon production in the boosted category at 95% CL. Limit is shown for an analysis with just 2017 (top left) and 2018 (top right) data, and both years combined (bottom).

4.9.3 Theory Parameter Calculations

The previously shown limits on excess dimuon production are the most directly useful presentation of these results for the particle physics community at large, because they are strictly model-independent. A theorist developing a novel BSM model may find that they predict a new particle (X) which decays into a pair muons, with an invariant mass between 1 and 8 GeV, which they wish to compare with our limit. They would begin their interpretation by assuming that the cross-sections shown in Figures 4·64 and 4·65 correspond specifically and exclusively to the production of their new particle, giving the relationship:

$$\sigma_{\text{obs.}} = \sigma_{pp\to X}\left(\vec{\alpha}\right)\ \mathcal{B}_{X\to\mu\mu}\ A_X, \tag{4.18}$$

where $\sigma_{\text{obs.}}$ is the value of our model-independent limit, $\sigma_{pp\to X}\left(\vec{\alpha}\right)$ is the production cross-section of the new particle for some set of parameters that define its coupling to the SM fields, $\vec{\alpha}$. $\mathcal{B}_{X\to\mu\mu}$ is the branching fraction of the new particle to decay into muons, and A_X is the fraction of dimuons produced by this particle that will be accepted by our analysis fiducial cuts. By studying the couplings and final state kinematics of this particle, when produced in pp collisions, the theorist can make predictions for the branching fraction and acceptance of their new particle in our analysis, enabling them to state an upper bound on the production cross section of their particle:

$$\sigma_{pp\to X}\left(\alpha\right) = \frac{\sigma_{\text{obs.}}}{\mathcal{B}_{X\to\mu\mu}\ A_X}. \tag{4.19}$$

Because this exclusively describes 13 TeV pp collisions at the LHC, this is not very useful for characterizing this model and understanding how it might impact other experimental results. However, if we assume the rate of this process scales linearly with some single parameter of the model, α_i, then we can convert the upper limit on the cross-section to an upper limit on that parameter, provided we have some reference value of the production cross-section at a known value of this parameter. Because the production rate scales directly with α_i, we gain the relationship:

$$\frac{\sigma_{pp\to X}\left(\alpha_i\right)}{\alpha_i} = \frac{\sigma_{pp\to X}\left(\alpha_{i,0}\right)}{\alpha_{i,0}} \qquad \longrightarrow \qquad \sigma_{pp\to X}\left(\alpha_i\right) = \frac{\sigma_{pp\to X}\left(\alpha_{i,0}\right)\alpha_i}{\alpha_{i,0}}, \tag{4.20}$$

where $\alpha_{i,0}$ is some chosen value of the parameter, and $\sigma_{pp\to X}\left(\alpha_{i,0}\right)$ is the calculated value of the cross-section for $\alpha_i = \alpha_{i,0}$. Substituting this in factors out the aspects of the measurement that are dependent on the mechanism of producing the particle:

$$\frac{\sigma_{pp\to X}\left(\alpha_{i,0}\right)\alpha_i}{\alpha_{i,0}} = \frac{\sigma_{\text{obs.}}}{\mathcal{B}_{X\to\mu\mu}\ A_X}. \tag{4.21}$$

Solving for α_i:

$$\alpha_i = \frac{\sigma_{\text{obs.}}\alpha_{i,0}}{\sigma_{pp\to X}\left(\alpha_{i,0}\right)\ \mathcal{B}_{X\to\mu\mu}\ A_X}. \tag{4.22}$$

This equation expresses α_i entirely in terms of the observed limit, and model specific quantities that can be determined via calculation and simulation studies. We will follow this prescription to constrain parameters from the previously discussed benchmark models: a massive dark photon and pseudoscalar in a 2HDM+S scenario.

Dark Photon Model Parameters

For the case of a dark photon (Z_D), the rate of excess dimuons produced through the DY process scales directly with the square of the kinetic-mixing parameter (ε^2), as described in Section 4.2.1. For this particle, Equation 4.22 becomes:

$$\varepsilon^2 = \frac{\sigma_{\text{obs.}}\ \varepsilon_0^2}{\sigma_{pp\to Z_D}(\varepsilon_0)\ \mathcal{B}_{Z_D\to\mu\mu}\ A_{Z_D}}. \tag{4.23}$$

We determine the model specific parameters $\sigma_{pp\to Z_D}(\varepsilon_0)$, $\mathcal{B}_{Z_D\to\mu\mu}$, and A_{Z_D} with a mixture of different software tools. The predicted cross-section and branching fraction to muons are calculated simultaneously using MADGRAPH5_aMC@NLO. This calculation is performed at next-to-leading order (NLO) and uses the HAHM configuration HAHM_variableMW_UFO with a reference coupling of $\varepsilon_0 = 0.02$ and the pdf set NNPDF31_nnlo_as_0118. An inclusive calculation of the dark photon DY production cross-section will consist of the LO $2 \to 2$ process, as well as $2 \to X$ diagrams with additional jets from ISR, therefore we allow MadGraph to produce up to two additional partons in the final state. The matrix element calculation for $q\bar{q} \to \mu^+\mu^- +$ N jets diverges in the limit where the jets are extremely soft (low-momentum) or collinear with other objects. The typical remedy for this is to use a parameter in the generator that quantifies the "softness" and "collinearity" of the partons in the event (k_T), and preemptively kill the generation of an event if it falls above some tunable threshold (xqcut). A benchmark value for this parameter that gives accurate calculations for processes with partonic momentum exchange (Q^2) greater than $\sim$10 GeV is xqcut = 10 GeV. However, the momentum spectrum of ISR jets emitted from events with low momentum exchange ($Q^2 < \quad \sim 10$ GeV) trends increasing soft as Q^2 decreases. This means that for the masses considered in our analysis, much of the physical phase space that should be included in the calculation will be killed by the xqcut requirement, leading to an underestimation of the dark photon production cross-section.

To address this, we use a numerical tool called DYTurbo [87], which is capable of accurately calculating the DY process cross section at next-to-next-to-leading order (NNLO). While it is not specific to a dark photon, it will model how DY cross section values should trend at lower masses, which can be can be used to correct the MadGraph prediction. To do this, we rescale cross-section values calculated with DYTurbo until they agree with MadGraph at higher masses, where the xqcut threshold still results in accurate calculations. At lower masses, as the MadGraph values become less reliable, the NNLO calculation from DYTurbo remains accurate and we tune the value of xqcut in the MadGraph calculations until the cross section distribution matches the shape of the DYTurbo result. Table 4.9 summarizes the calculated cross sections at each mass and shows the tuned values of xqcut that were implemented to match the trend of DYTurbo. Figure 4·66 shows the interpolation of these values, as well as the fine agreement between the rescaled DYTurbo values and the tuned MadGraph. To contrast, it also shows the MadGraph cross section when calculated at LO precision, which is significantly lower. The QCD factorization and renormalization scale used for the DYTurbo calculation is $\mu = 0.5\sqrt{m_{\mu\mu}^2 + p_{\mu\mu}^2}$. The theoretical uncertainty on the cross-section is estimated by taking the percent difference in the cross sections when varying the value of this parameter up and down by a factor of 2, which results in an upwards uncertainty of 3.4% and a downwards uncertainty of 9.7%.

m_{Z_D} (GeV)	Cross section $\times\mathcal{B}$ (pb)	xqcut (GeV)
1.1	9.0×10^4	1.7
2.0	4.5×10^4	2.0
3.0	1.31×10^4	5.0
4.0	5181	5.0
5.0	3126	5.0
6.0	2198	5.0
7.0	1446	6.0
8.0	1007	7.5
9.0	738	9.0
10.0	559	10.0
12.5	352	10.0
20.0	136	10.0

Table 4.9: Cross section of $pp \to Z_D \to \mu\mu$ process from MadGraph, with optimized xqcut identified using DYTurbo.

The remaining theoretical quantity to be determined is the acceptance of two muons produced in a dark photon DY process to be included in our analysis fiducial cuts ($p_{\mathrm{T}} > 4$ GeV, $|\eta| < 1.9$). Because the final state kinematics of a dark photon are the same as regular DY events, we use DYTurbo as our main tool for this measurement. We again use the QCD scale form of $\mu = 0.5\sqrt{m_{\mu\mu}^2 + p_{\mu\mu}^2}$, and perform the calculation at NNLO precision. We infer the acceptance to be the ratio of the DY cross section calculated with the outgoing phase space restricted to our analysis

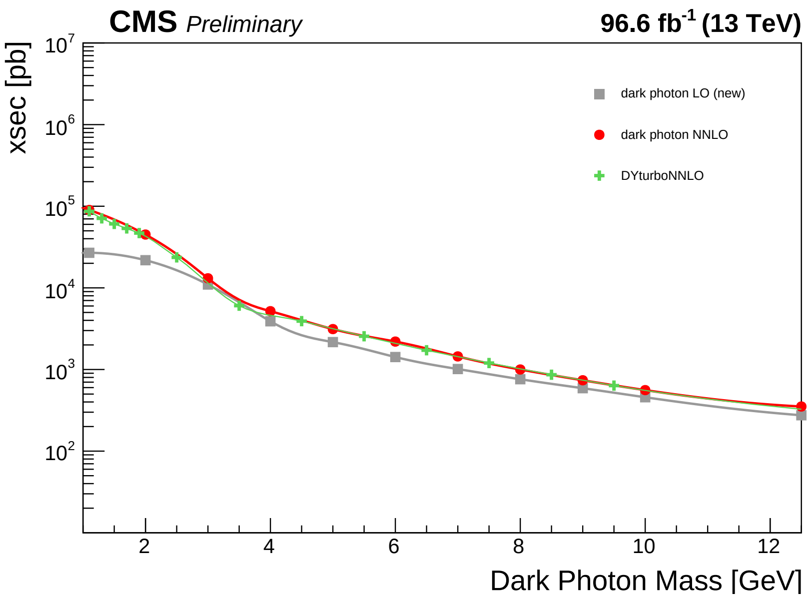

Figure 4·66: Tuned xqcut dark photon to muon cross sections from MadGraph, with rescaled DYTurbo calculations, and LO only comparison.

cuts divided by the total inclusive cross section. Figure 4·67 shows the result of this calculation, represented by the gray squares. The other points represent checks on similar measurements done with alternate configurations of DYTurbo, described in the caption, or different generators altogether. The acceptance from MadGraph is calculated with and without the tuning of xqcut from the cross-section measurement, and is in very close agreement when the tuning is applied. We estimate the the uncertainty on the acceptance to be the average percent difference between the DYTurbo values in the nominal configuration, and the values from MadGraph with the tuned xqcut, which is roughly 30%.

Pseudoscalar Model Parameters

For the pseudoscalar in the Type IV 2HDM+S scenario, with a fixed value of $\tan(\beta) = 0.5$, the rate scaling parameter we are interested in constraining is the square of the pseudoscalar fraction of the light pseudoscalar state $(\sin(\theta_{\mathrm{H}})^2)$, described in Sec-

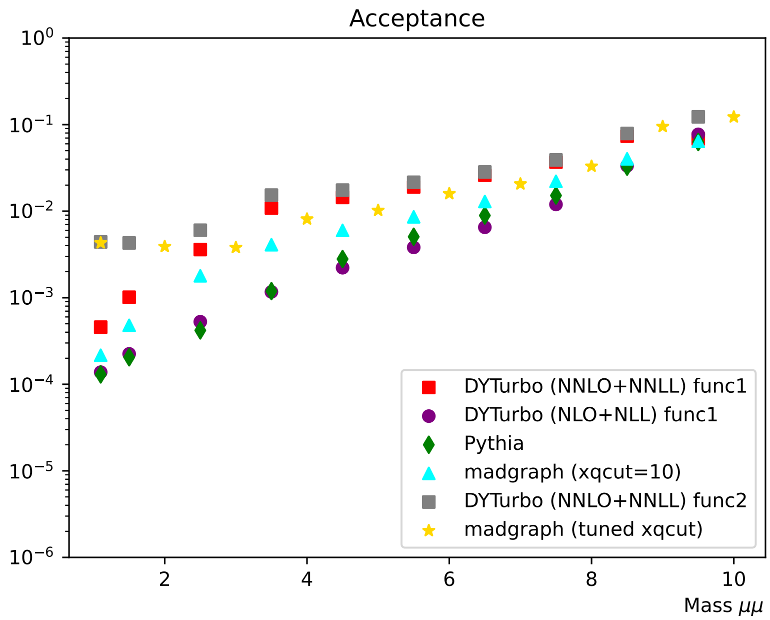

Figure 4·67: Calculated fiducial acceptance of dark photon signals from different generators and configurations, including Pythia, MadGraph, and DYTurbo with different QCD scales. The MadGraph results are shown with (yellow) and without (cyan) the tuning of xqcut applied to the cross-section calculation. For the different DYTurbo measurements, "func1" represents the unfavored QCD scale functional form $\mu = 0.5m_{\mu\mu}$, while the "func2" represents the nominally used form $\mu = 0.5\sqrt{m_{\mu\mu}^2 + p_{\mu\mu}^2}$.

tion 4.2.2. Similar to the dark photon, we have the relationship:

$$\sin\left(\theta_{\mathrm{H}}\right)^2 = \frac{\sigma_{\mathrm{obs.}}\ \sin\left(\theta_{\mathrm{H},0}\right)^2}{\sigma_{pp\to a}(\sin(\theta_{\mathrm{H},0})^2)\ \mathcal{B}_{a\to\mu\mu}\ A_a}. \tag{4.24}$$

For this measurement, $\sigma_{\mathrm{obs.}}$ corresponds to the model-independent limit in the boosted category, as its selections were tailored specifically to this signal mode. The reference cross section calculation is performed using HIGLU, a specialized program for measuring the production cross section of Higgs particles via ggF. This software does not include settings for the specific 2HDM theory we are hoping to constrain, but it can calculate the cross section for a generic pseudoscalar Higgs state, which we can modify to more accurately describe our scenario. Because this state is a "pure" pseudoscalar, in our theory it would represent the case of maximal mixing between the a particle and the SM fields, $\sin(\theta_{\mathrm{H}}) = 1$, which we will treat as our reference value for the coupling. Next, our chosen value of $\tan(\beta) = 0.5$ will modify the couplings of the a particle to each of SM fermion fields according to the values shown in Table 4.2. For a Type IV 2HDM, this means the couplings to up-type quarks and charged leptons is doubled, and the coupling to down-type quarks is halved.

The calculation is performed with the NNPDF30_NNLO_AS_011 pdf set, and nominal QCD scales of $\mu = 0.5 m_{\mu\mu}$ for refactorization and renormalization. Figure 4·68 shows the production cross section for this proxy a particle, as well as an uncertainty band determined by doubling the QCD scale and treating the percent difference as a symmetrical uncertainty about the nominal value. The branching fractions to muons are calculated separately, following the prescription from the appendices of Ref. [[64]]. Table 4.10 summarizes the production cross sections and branching fraction calculations[2].

The fiducial acceptance for the final state muons is determined by generating samples of pseudoscalar events produced by ggF, and measuring the ratio of events that satisfies the analysis cuts for the boosted category ($p_{\mathrm{T}} > 5$ GeV, $p_{\mathrm{T}_{\mu\mu}} > 20/35$ GeV, $|\eta| < 1.9$) versus the total number in the sample. The samples are produced using PYTHIA 8, at a range of masses between 1 and 8 GeV. The couplings to SM fermions are adjusted similarly to the cross section calculation, but it is found to have no effect on the acceptance values. As with the dark photon acceptance, we estimate the uncertainty by repeating the measurement using MadGraph and taking the mean disagreement across all masses, which is again roughly 30%.

[2] Branching fractions are sourced from a previously conducted calculation campaign for this model, available at https://github.com/cecilecaillol/CMS_HAA_SummaryPlots

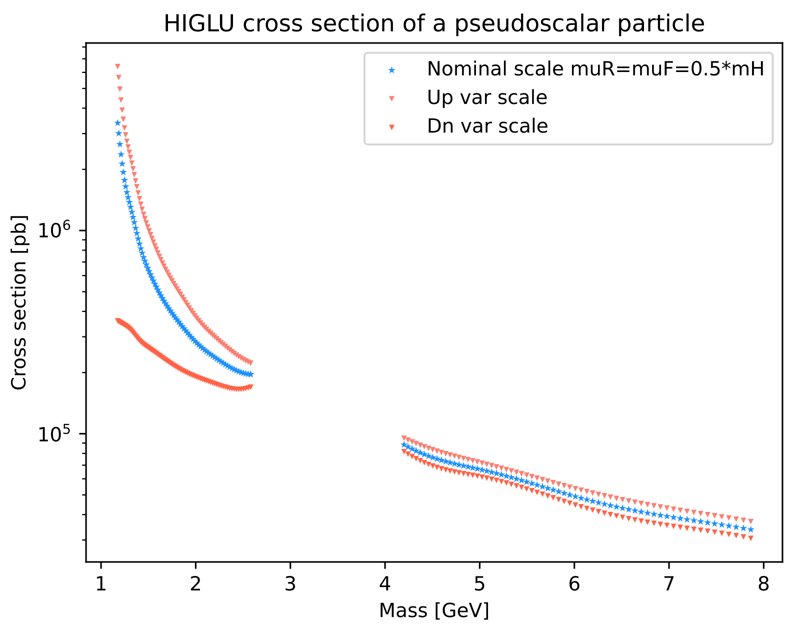

Figure 4·68: HIGLU calculated production cross section for a pure pseudoscalar state in a Type IV 2HDM+S scenario with a fixed value of $\tan(\beta)$ =0.5. Orange lines represent the uncertainty band, determined by varying the QCD scale.

Summary

Figure 4·70 shows the product of the BSM production cross-section, times branching fraction to muons, times fiducial acceptance for both models. This values represents the denominator of the right-hand side of Equations 4.23 and 4.24, so a larger value of this parameter represents a smaller value on the parameter upper limit, meaning more phase space is excluded. The shaded bands represent one standard deviation of the combined cross section and acceptance uncertainties described previously.

m_a (GeV)	Cross section (pb)	$\mathcal{B}$
1.1	77910740	0.024185
1.2	2631990	0.025447
1.3	1372070	0.273753
1.4	881140	0.280388
1.5	642097	0.028781
1.8	365239	0.028135
2.0	283080	0.026048
2.2	237671	0.022763
2.5	199051	0.013887
4.0	101029	0.003951
5.0	66799	0.002415
6.0	49249	0.001905
7.0	39060	0.001802
8.0	328129	0.001778
20	8062	0.001244

Table 4.10: Production cross section of our pseudoscalar particle a from the HIGLU generator and corresponding branching fraction for a decay into muons for select mass points.

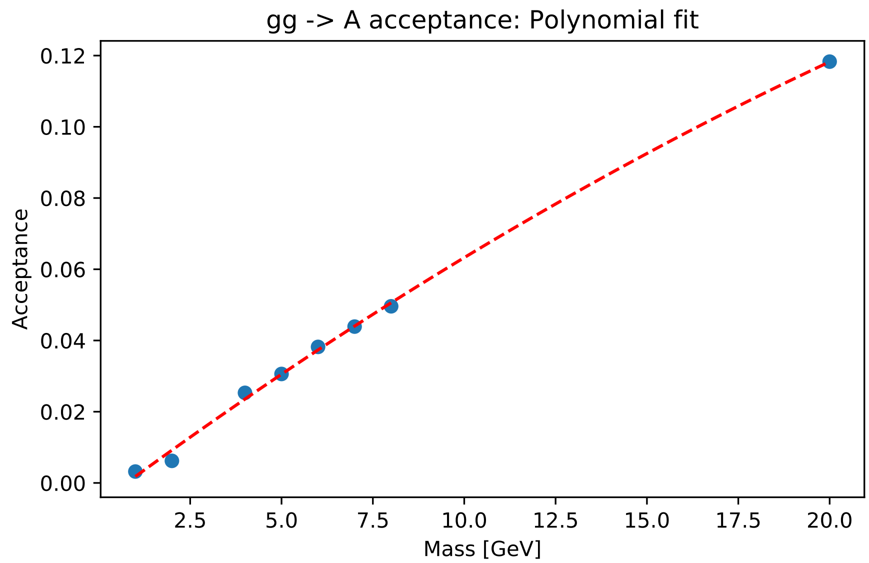

Figure 4·69: Fraction of simulated ggF pseudoscalar events that satisfy boosted category analysis cuts, fit to second-order polynomial.

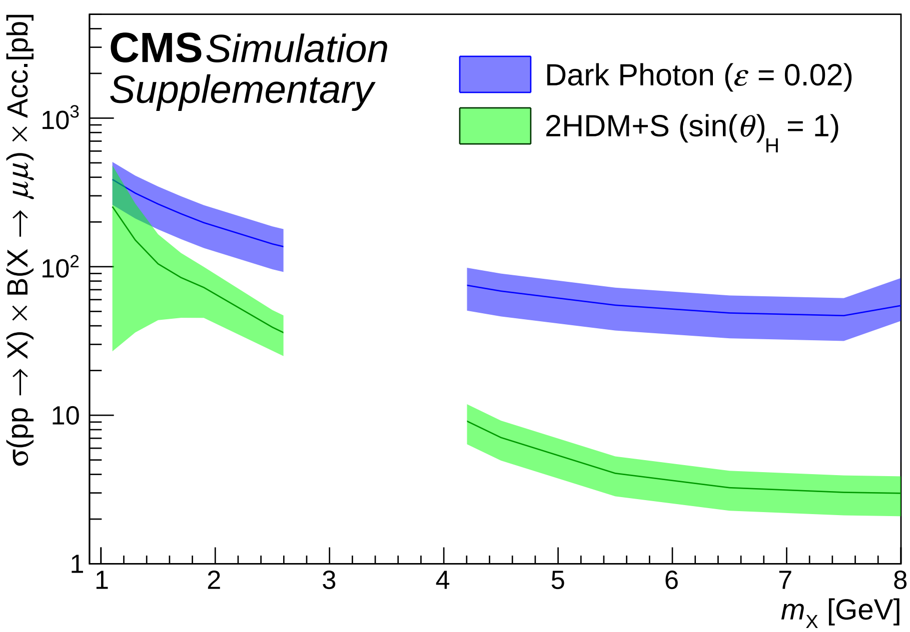

Figure 4·70: Product of all model specific parameters needed for BSM constraints. Shaded band represents 1σ theoretical uncertainty estimations.

4.9.4 BSM Parameter Limits

With all model parameters defined for both BSM scenarios, we can produce upper bounds on the rate scaling parameters of interest. For the ε^2 limit, $\sigma_{\text{obs.}}$ corresponds to the inclusive model-independent limit, and for $\sin(\theta_{\text{H}})$ it is the boosted model-independent limit. Figure 4·71 shows both parameter limits, with a shaded band representing the variance allowed by our theory uncertainty estimations. To enable direct comparison with results from other experiments, the model-independent limits are re-evaluated at a CL of 90% when used for these results.

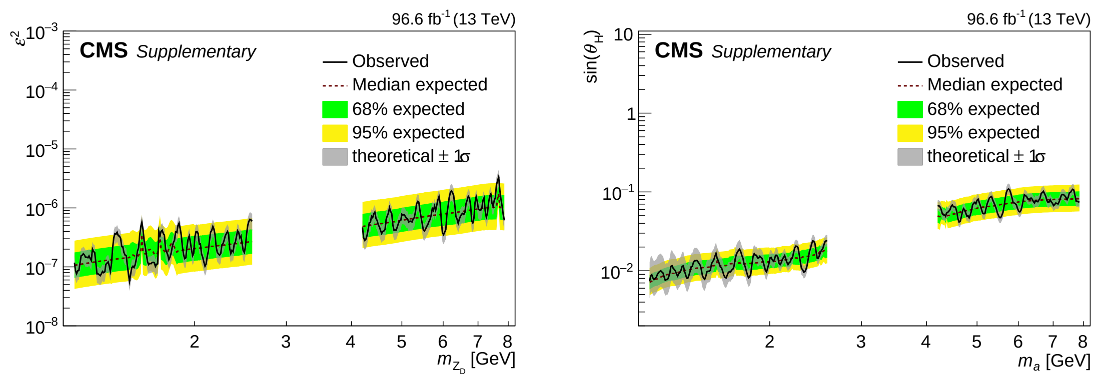

Figure 4·71: Upper limits on ε^2 (left) and $\sin(\theta_{\text{H}})$ (right) at 90% CL. The gray band surrounding the observed limit represents the effect of a $\pm 1\sigma$ pull on the theory value uncertainties.

Searches for similar signatures have been performed by the LHCb [88] and BaBar [89] experiments. Figure 4·72 (upper) shows the ε^2 constraints from this analysis, compared with recent results from both experiments, reported at 90% CL [90, 91], and reinterpreted with the DarkCast software package, as found in Ref.[92]. We find that our analysis generally provides tighter constraints on ε^2 than the LHCb results, and is of comparable sensitivity to BaBar at most mass values, with a slight advantage at very low masses. We attribute this to the enhancement of the dark photon cross section in pp collisions with low Q^2. Both experiments have superior mass resolution to CMS, allowing them to set limits much closer to the SM peaks, without significant effects from the resonance tails. Figure 4·72 (lower) shows a similar comparison of the $\sin(\theta_{\mathrm{H}})$ constraints to results from BaBar and a more recent LHCb analysis targeting pseudoscalar production [93]. Our result is competitive with LHCb at masses below 4 GeV, and dramatically better than both experiments above. BaBar is significantly weaker at constraining pseudoscalar production, because it records e^+e^- collision data, which cannot facilitate the dominant ggF production mode. These results are currently published, and available at Ref. [94].

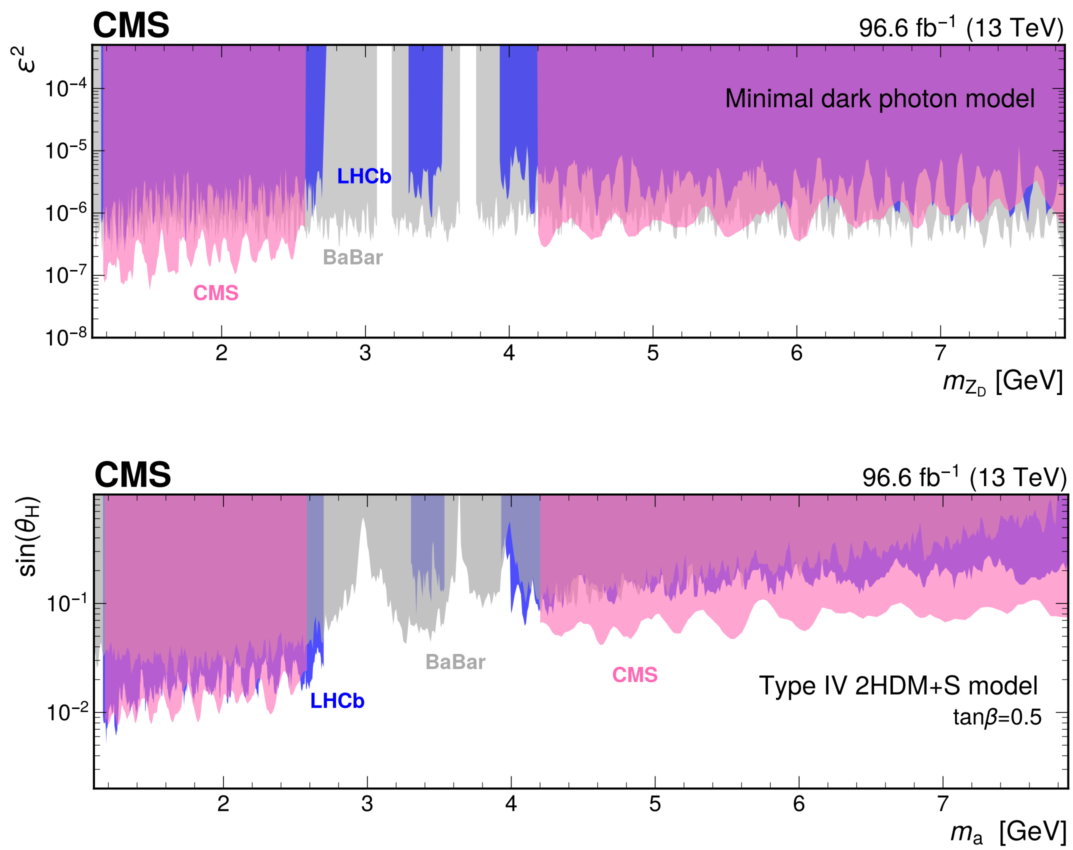

Figure 4·72: Observed upper limits at 90% CL on the square of the kinetic mixing coefficient ε (top) and the mixing angle θ_H for a Type IV 2HDM+S scenario (bottom). Filled in regions represent values excluded by this analysis (pink) as well as recent results from LHCb [93, 95] (blue) and BaBar [91] (gray).

Chapter 5

Conclusions

In the years since the first observation of the Higgs boson in 2012, particle physicists all over the world have produced an incredible volume of research to test the limits of the standard model (SM). Thousands of papers have been published using data collected from LHC collisions, both in the form of precision measurements of SM effects and novel searches for new particles and phenomena. In spite of this effort, no sure deviations from the SM or hints of new particles have yet been observed, and many of the promising models which previously motivated many LHC searches, such as super-symmetry, are increasingly disfavored. The current impetus for our field is to seek out any anomalous effects that might be hiding in the available data. A strong measurement of any phenomenon that cannot be explained by the SM could galvanize the particle physics community and spark a new wave of model-making and tailored analyses to explain this new mystery.

This thesis has described the methods and results of a search for such an anomaly, in the form of resonant excess production in the dimuon mass spectrum between 1 and 8 GeV in LHC Run-II data collected by CMS. This search was enhanced by the use of the CMS scouting data stream, which allowed for the recording of many events with low dimuon masses, most of which would be rejected by the standard triggers. With this wealth of statistics available in the scouting data, this analysis set the first limits on resonant dimuon production with CMS data below 10 GeV. We report two sets of model independent limits on excess dimuon production; an inclusive result which represents the kinematics of Drell-Yan-like production, and a boosted results with selections that better describe signal modes with enhanced probability of a large recoil, such as gluon-gluon fusion. The results in both categories were consistent with SM dimuon production, with no observed excesses of local significance greater than 2.6σ. The results are also interpreted as upper bounds on BSM parameters for a dark photon model and a pseudoscalar in a 2HDM+S scenario. Our constraints on these parameters are competitive with contemporary experiments, and are worlds best at many mass hypotheses.

Though this search has not found evidence of BSM physics, it has excluded a large section of the coupling parameter space for the two models we consider, and our model independent limits are available for reinterpretation by members of the theory community. Additionally, this analysis has resulted in an improved understanding of the dimuon mass spectrum in the CMS scouting data, such as the discovery and subsequent modeling of resonant backgrounds from decays of $D0$ mesons into charged kaons and pions. These lessons will benefit future CMS analyses exploring similar signatures. At the time of writing this thesis, LHC Run-III is underway and the data scouting stream has been expanded to include additional particles as well as an increase to the maximum allowed rate. By studying this huge volume of data in new and creative ways, CMS physicists may be able to find the first ever evidence of physics beyond the standard model.

Appendix A

AMC13 for HCAL Readout and TTC

Omitted from the discussion of the detector in Section 3, is the method of buffering the contents of an event while it awaits a judgement from the global trigger. Further, the event fragments as they are delivered by the front ends are encoded with metadata that associates them with a specific bunch crossing and LHC orbit index, which requires precise timing control to keep synchronized. For most of HCAL, both of these functions are managed by a single electronics component, the AMC13 [96]. Developed by CMS physicists and engineers at Boston University, the AMC13 is a μTCA card that supervises a crate full of front end buffer modules, called μHTRs for the HCAL use case. Figure A·1 shows a photograph of an AMC13 card from multiple angles. The AMC13 receives and distributes trigger, timing, and control

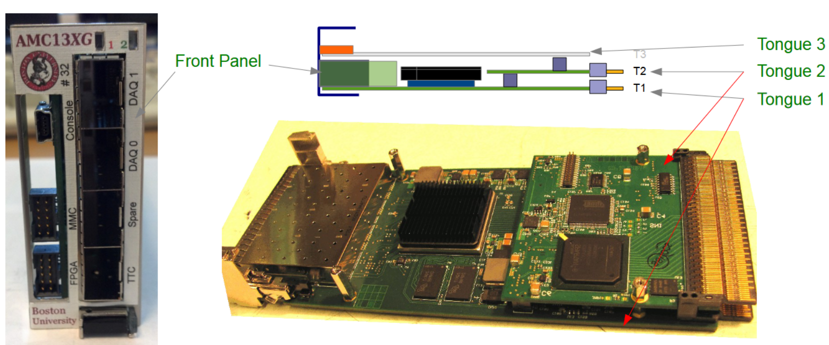

Figure A·1: Photograph of AMC13 front plate (left) and side view (right). Taken from Ref. [96].

(TTC) information, which includes the LHC clock as well as L1 accept signals from the global trigger. When an L1A is received, the AMC13 collates the corresponding event fragments stored in the μHTR buffers and transmits the event to the central DAQ system via up to three 5.0 Gb/s optical readout links, so it can be reconstructed and evaluated by the HLT. The AMC13 is also used to throttle the trigger rate when there is a risk of buffers overflowing. Its direct link to the global trigger allows it to veto L1A signals if it is in a state where an event cannot be read out, either due to a back-pressure on the link to central DAQ, or violation of the L1 trigger rules, which require some minimum number of bunch crossings between successive events that can be read out.

The DAQ and TTC functions are mostly segmented across the two main boards (tongues) that make up the device, each with their own FPGA chip running custom firmware. A block diagram illustrating the most important components of each board is shown in Figure A·2. The larger of the boards is the "T1," which mostly handles the event building process and data readout to central DAQ. It houses a Xilinx Kintex 7 (or Vertex 7) FPGA, as well as the four optical transceivers that extend to the front panel. Three of these are dedicated to data readout, and the remaining one accepts and distributes clock, trigger, and timing information. Assembled events are also saved in an circular SDRAM buffer on the T1, which can be accessed over ethernet, to study of potentially corrupted data or the effect on saved events when errors occur. Up to 1000 events can we stored in this buffer at a time.

The other major board is the "T2," which houses a Xilinx Spartan 6 FPGA and mostly manages the TTC functionality of the device. It communicates with the μHTR cards in its crate via the μTCA backplane fabric connections. The T2 also includes a Flash memory chip that allows firmware to be preloaded onto the device, which can automatically program both FPGAs upon powering up. AMC13s TTC functionality has a wide range of applications, even for detectors that do not use of its data acquisition features. One example is the silicon strip detector, which makes use of the AMC13s TTC functions for controlling its front end electronics, but has a separate system to buffer and readout event information. The AMC13 was

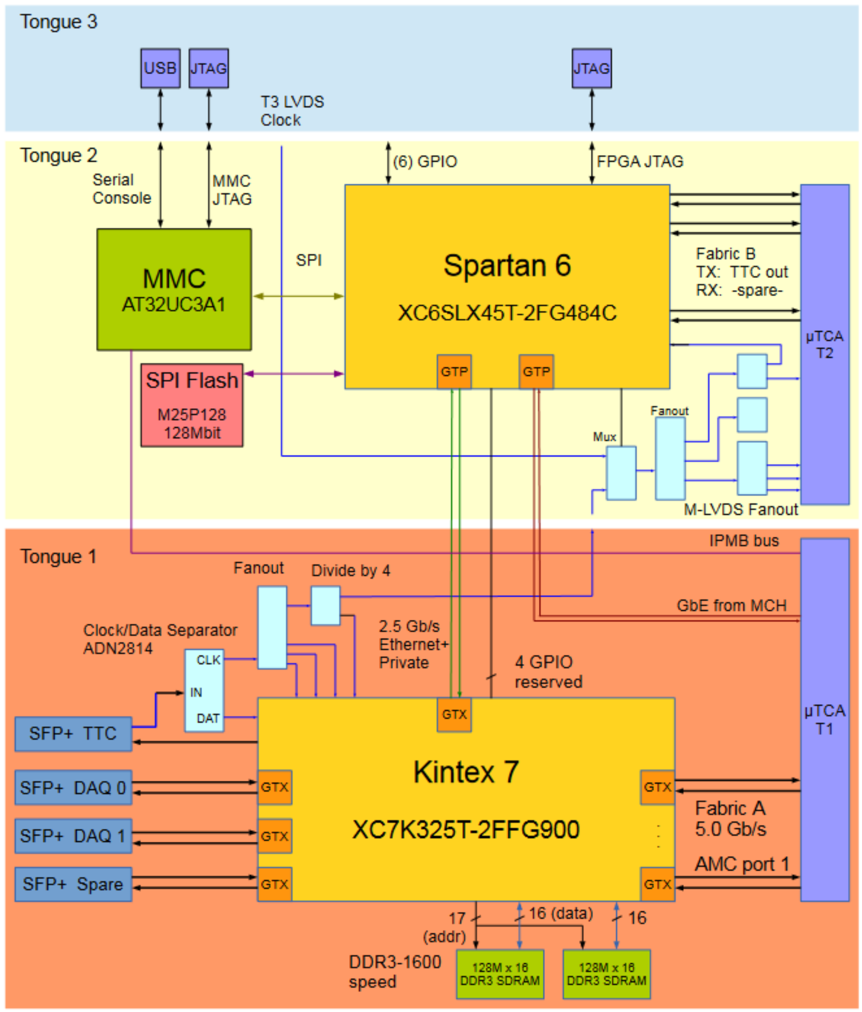

Figure A·2: Block diagram of the AMC13. Taken from Ref. [96].

originally implemented as part of the "Phase-I" CMS upgrade campaign, with the goal of standardizing the back end electronics for the different HCAL subsystems prior to Run-II. However, in the years leading up to Run-III there was a large amount of demand for more AMC13s to made available and sold. Many were requested for use on the detector itself, for the GEM backend system and to accommodate the additional HCAL channels after the depth segmentation increase. Many were also requested to provide TTC information for test-stand set ups of other detector systems. To meet this demand, 49 new AMC13s were commissioned in late 2020 all of which were subjected to rigorous testing by BU graduate students, to ensure that they were fit to distribute. These test included multiple runs with simulated data, where a μTCA crate was was filled with AMC13s programmed to emulate μHTR cards, each of which produced their own fake event fragments.

AMC13 will be mostly phased out for its data readout application in CMS Phase-II, given that the L1 Trigger rate is expected to be significantly higher than the current μTCA electronics can handle ($\sim$750 kHz). However, it may retain its role as a timing distributor and front end controller in some sub-detectors.

Appendix B

Boosted Category Window Adjustments

At the time of unblinding, the model independent limit in the boosted category used a window size of ± 5 times the mass resolution, identical to the inclusive category. During review of the analysis, it was noted that the limit seemed to exhibit "spiky" behavior, where the observed limit would oscillate intensely between troughs and peaks. Even more strange, in many regions the amplitude of local excesses seemed to be exactly influenced by the size of its neighboring deficits, as can be seen in Figure B·1. This behavior was judged to be an artifact of the analysis techniques and

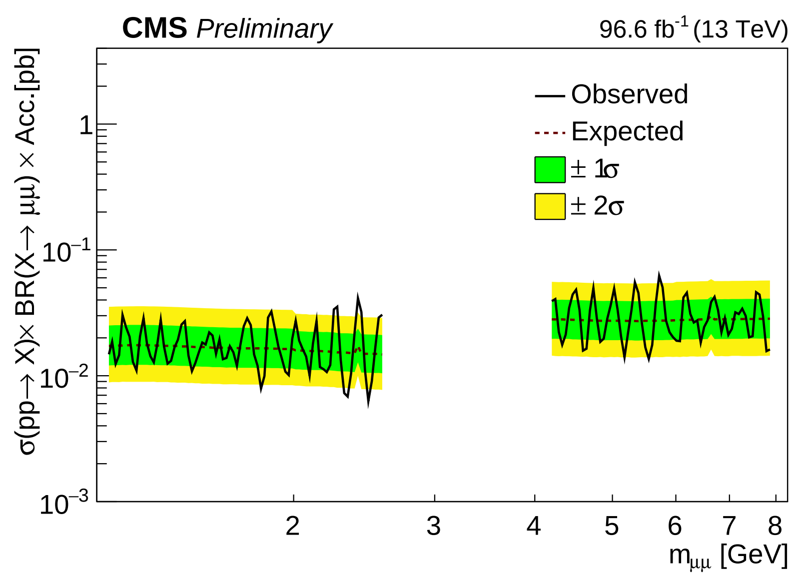

Figure B·1: Model independent limit in boosted category, with a window size of ± 5 mass resolutions.

was eventually attributed to the size of the sliding window. Because of the tightened muon and dimuon p_{T} cuts in the boosted category, there are less available events in the sidebands of each window, causing larger fluctuations which the background model must fit to. By increasing the window size more sideband data can contribute to the background fit, smoothing over spurious fluctuations and stabilizing the background function fit. The optimal window size for this category is determined by generating toys of the analysis input distribution, thrown from a spline fit to the data. We judge the behavior of the limit when it is run with progressively larger window sizes, and stop when it roughly converges to the limit obtained by using the original spline as the background model, which we consider to be the most accurate representation of the limit. Figure B·2 shows that the spiking behavior is mostly suppressed when using a window size of ±8 times the mass resolution.

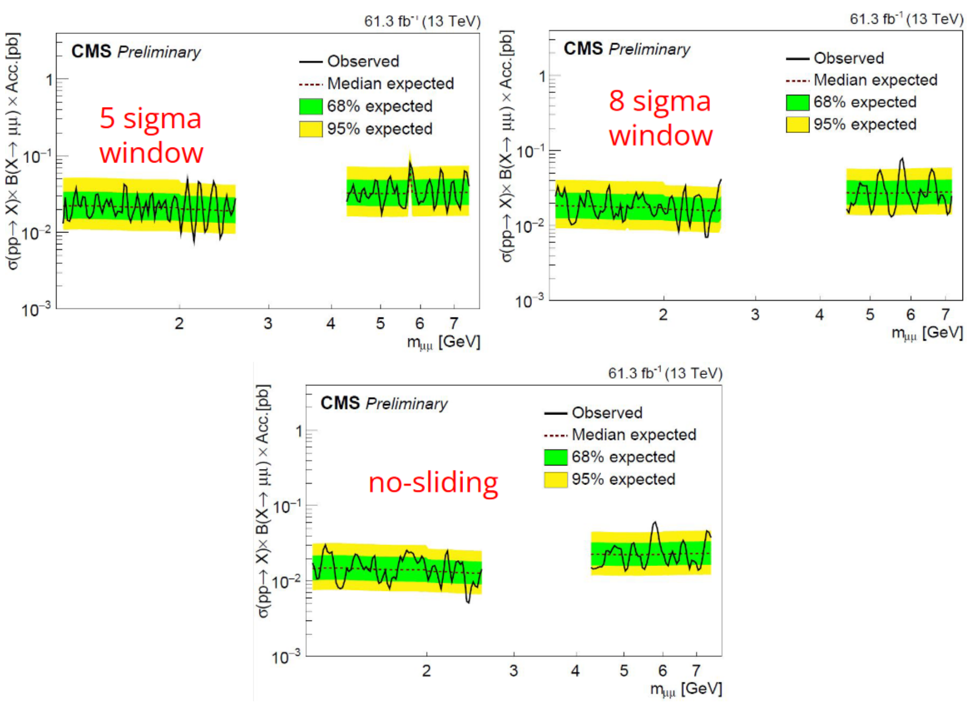

Figure B·2: Model independent limits, taken from a toy of the boosted category mass distribution fit with a window that is ±5 (top left) and ±8 (top right) times the mass resolution. The bottom shows the limit with no sliding window, and a single background function mapping the whole range.

Figure B·3 shows the model independent limit in data with both window sizes overlaid. Given the more natural character of this limit, the analysis strategy was amended, and use the larger window for the boosted category.

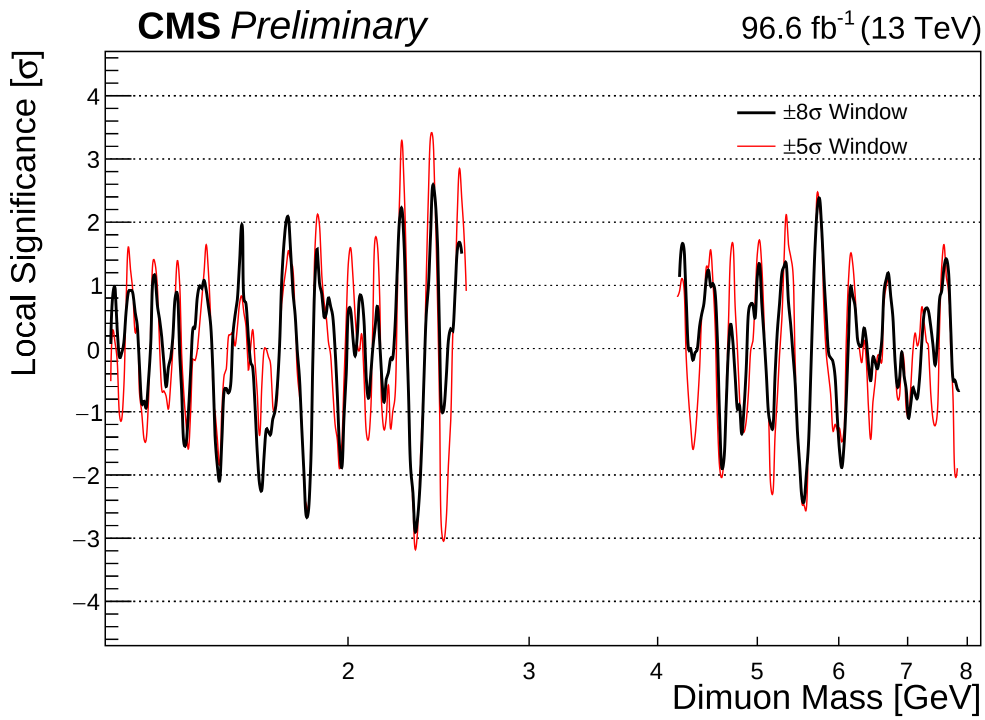

Figure B·3: Comparison of boosted category limit significance when using mass windows with a size of ± 5 times the mass resolution versus ± 8 times the mass resolution.

Appendix C

Extended Bias Studies

When developing the background function profile to be applied during the analysis, each of the included background functions were studied to gauge their individual bias when fit to a variety of toy data sets. The following figures show the typical comparison studies that would be performed between two functions. For the two functions studied in each plot, the dashed line represents the measured bias of the function when it is fit to a set of toys drawn from the background only fit of the same function (self-bias). The solid lines are the bias when fit the function in question is fit to toys drawn from the opposite functions background only fit. The forth order Bernstein polynomial has the best overall fit for most windows, and is treated as a baseline to compare the other functions against. These studies are performed with and without injected signal contributions. Figure C·1 shows the bias measurement between the forth order Bernstein polynomial and a forth order polynomial times an exponential function. Figure C·2 shows bias measurement between the forth order Bernstein polynomial and a third order Bernstein polynomial plus a power-law. Finally, Figure C·3 shows bias measurement between the forth order Bernstein polynomial and the sum of seven exponential functions. The latter of these appears to exhibit extremely high bias, particularly at low mass hypotheses, but this is likely due to strong disagreement in the shapes of the background only fits between the two functions. The self-bias (dashed blue) is within 30% for all measurements.

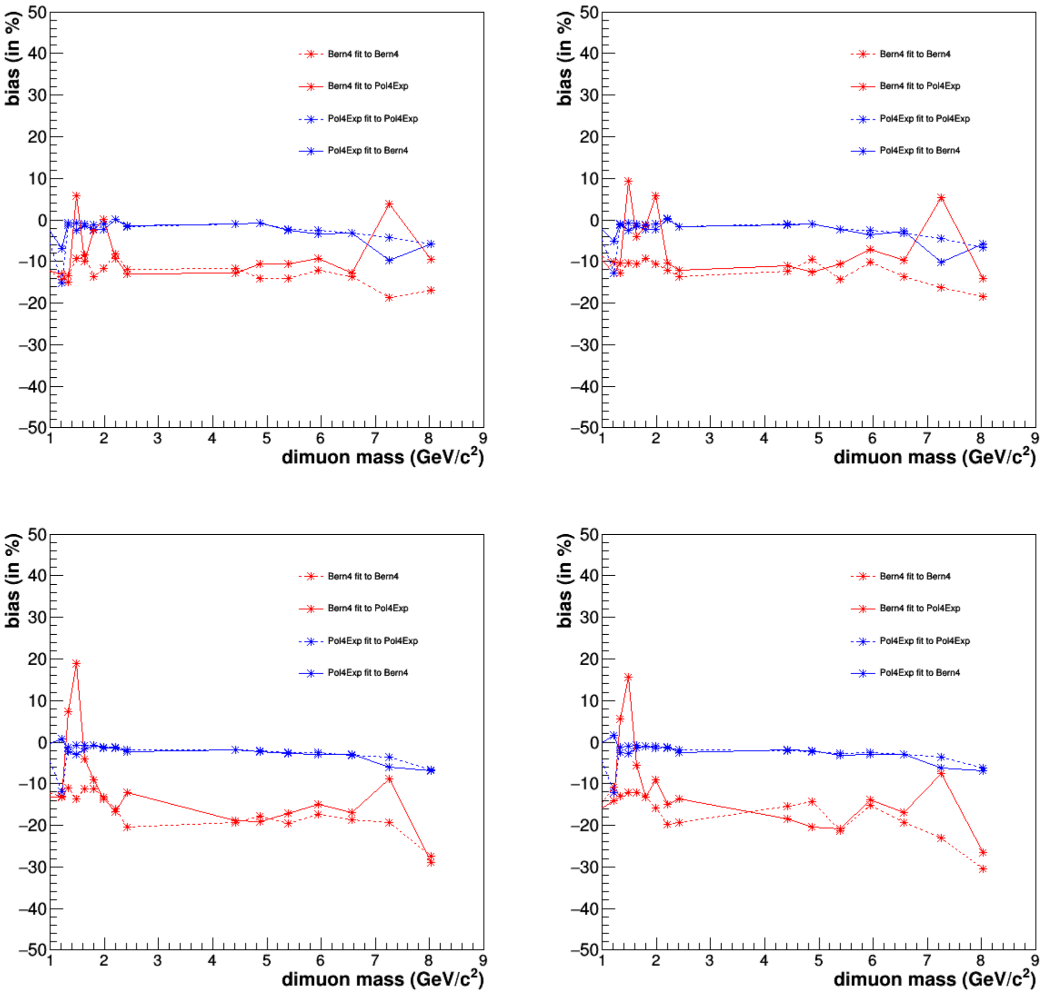

Figure C·1: Bias comparison between fourth order Bernstein polynomial and a forth order polynomial times exponential, when fit to 2017 (top) and 2018 (bottom) data. The left column is the measurement when run with no signal injected into the toys. The right includes a small injection of signal, on the order of the anticipated signal for a potential discovery in this analysis.

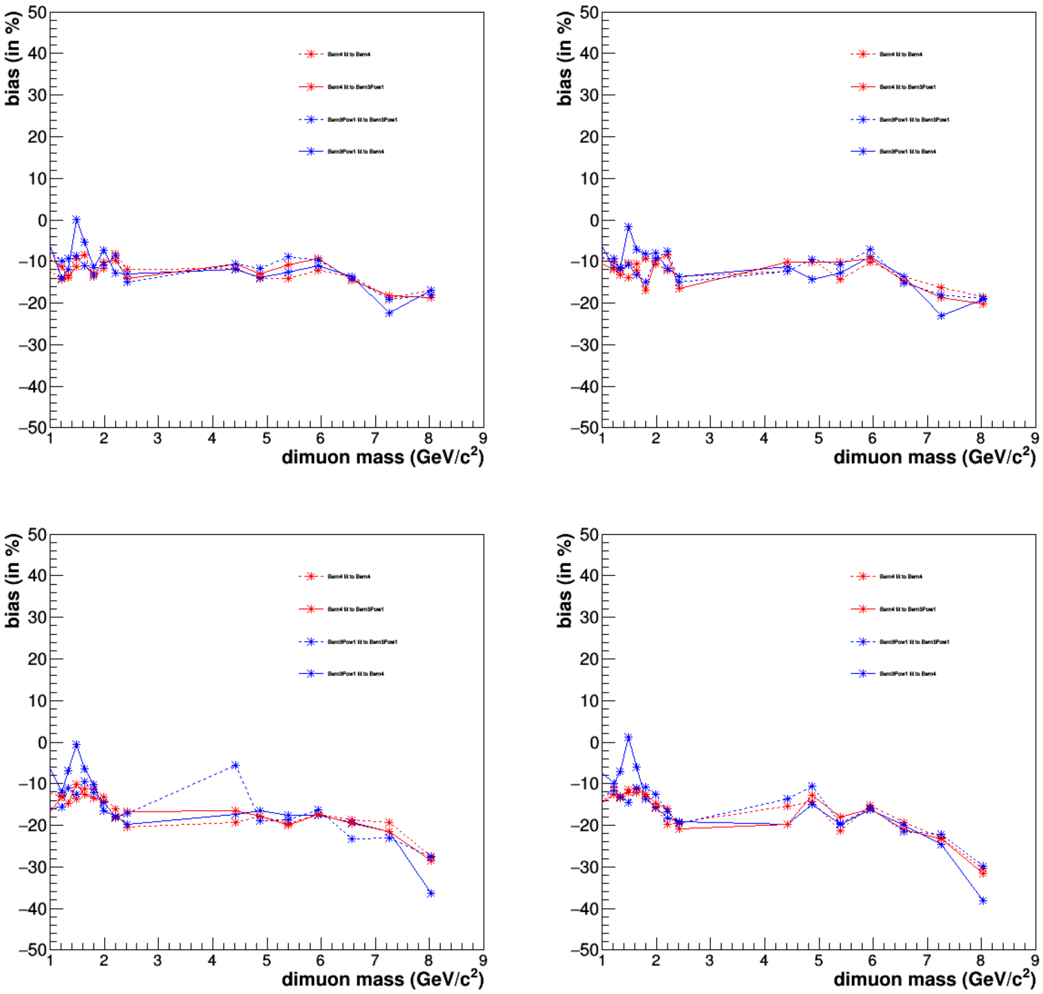

Figure C·2: Bias comparison between fourth order Bernstein polynomial and a third order Bernstein polynomial plus a power-law term, when fit to 2017 (top) and 2018 (bottom) data. The left column is the measurement when run with no signal injected into the toys. The right includes a small injection of signal, on the order of the anticipated signal for a potential discovery in this analysis.

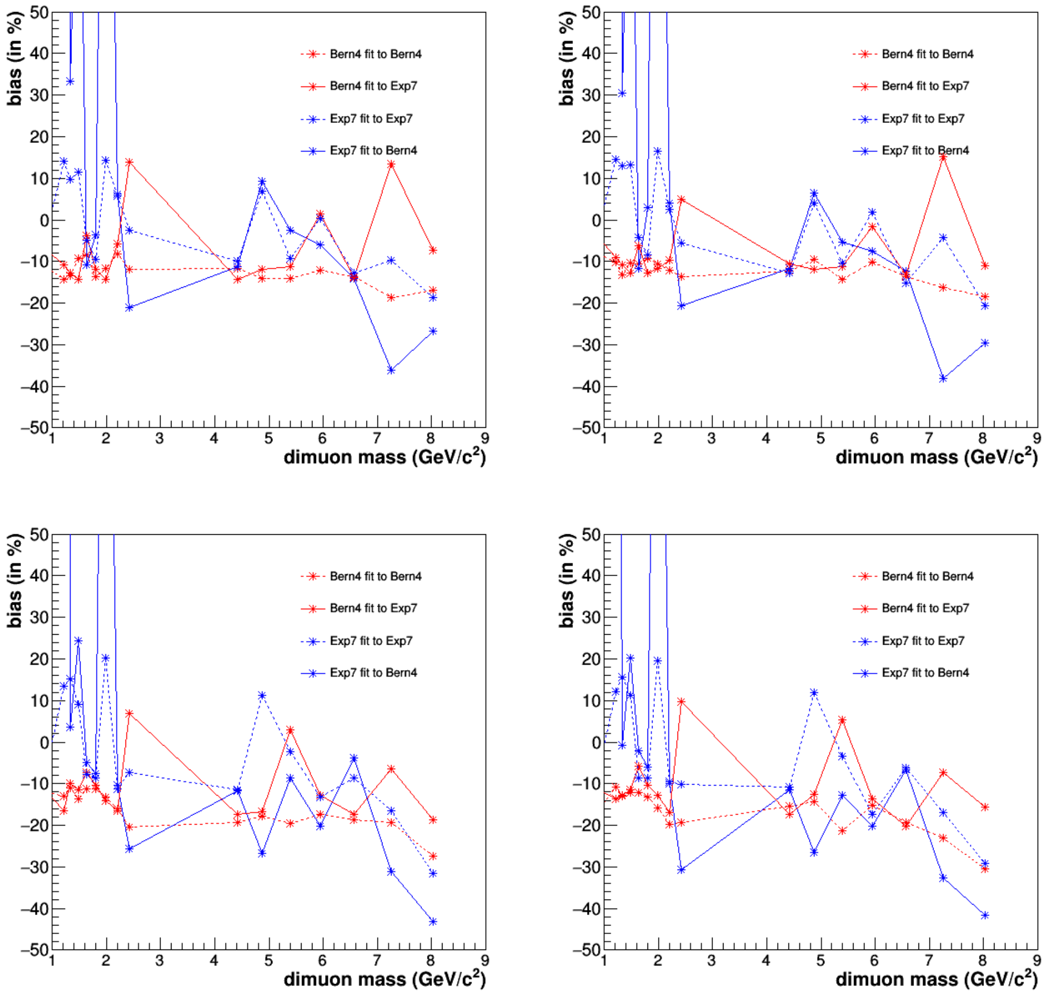

Figure C·3: Bias comparison between fourth order Bernstein polynomial and a sum of seven exponential terms, when fit to 2017 (top) and 2018 (bottom) data. The left column is the measurement when run with no signal injected into the toys. The right includes a small injection of signal, on the order of the anticipated signal for a potential discovery in this analysis.

Appendix D

Unexpected Peaking SM Backgrounds

Prior to unblinding, significant effort was spent to understand a cluster of windows in the inclusive category where background-only fits resulted in low fit probabilities. The effect is most pronounced between 1 and 2 GeV in fits, as can be seen in Figure D·1. Studying the shape of the dimuon mass distribution in these areas did not reveal any large, easily visible features that might be complicating the fit, so initial approval to unblind was given.

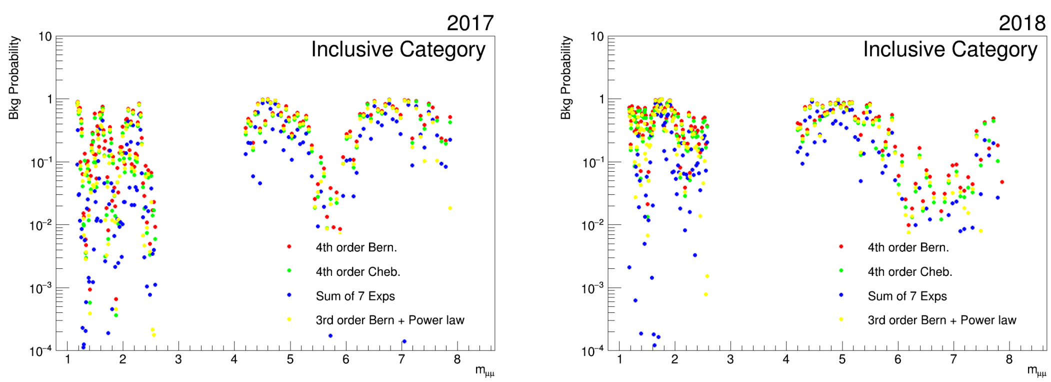

Figure D·1: Distribution of background only fit probabilities for all background functions. Each dot represents a fit to an inclusive category windows in 2017 (left) or 2018 (right) data.

Figure D·2 shows the model-independent limit in the inclusive category at initial unblinding, which includes significant excesses in the same areas as the clustered low-probability fits. The local p-value of all excesses in this limit are shown in Figure D·3.

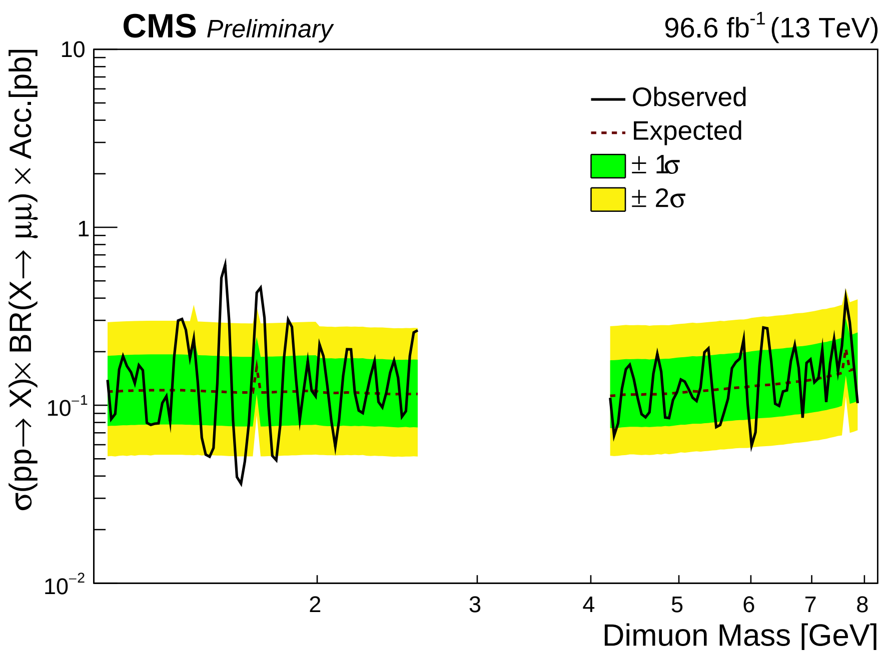

Figure D·2: Model-independent limit in the inclusive category prior to identifying and modeling contributions from D^0 meson decays.

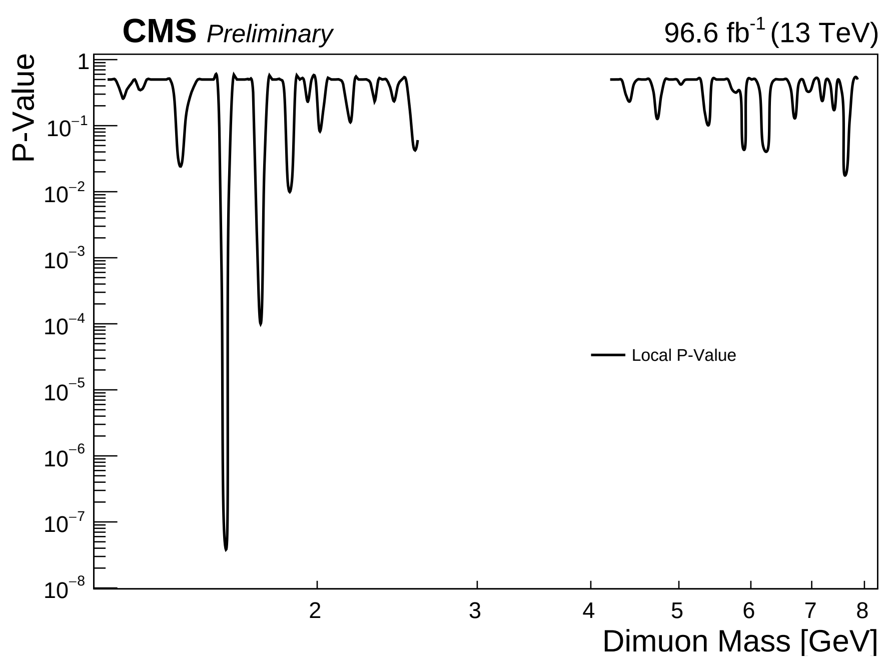

Figure D·3: Local p-value of excesses the inclusive category prior to identifying and modeling contributions from D^0 meson decays.

List of Journal Abbreviations

Astron. Astrophys. .... Astronomy and Astrophysics

Astrophys. J. .... The Astrophysical Journal

Comm. Kharkov Math. Soc. .... Communications of the Kharkov Mathematical Society

Commun. Theor. Phys. .... Communications in Theoretical Physics

Comput. Phys. Commun. .... Computer Physics Communications

Eur. Phys. J .... The European Physical Journal

IEEE Trans. Appl. Supercond. .. IEEE Transactions on Applied Superconductivity

IEEE Trans. Nucl. Sci. .... IEEE Transactions on Nuclear Science

JCAP .... Journal of Cosmology and Astroparticle Physics

JCSS .... Journal of Computer and System Sciences

JHEP .... Journal of High Energy Physics

JINST .... Journal of Instrumentation

J. Phys. Conf. Ser. .... Journal of Physics: Conference Series

J. R. Stat. Soc. .... Journal of the Royal Statistical Society

Nucl. Instrum. Meth. A .... Nuclear Instruments and Methods A

Nucl. Phys. B .... Nuclear Physics B

Phys. Lett. B .... Physics Letters B

Phys. Rept. .... Physics Reports

Phys. Rev. D .... Physical Review D

Phys. Rev. Lett. .... Physical Review Letters

Bibliography

[1] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).

[2] A. Pich, in 2004 European School of High-Energy Physics (Feb. 2005), pp. 1–48, arXiv:`hep-ph/0502010`.

[3] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[4] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964).

[5] The LEP Collaborations, Phys. Lett. B **307**, 187 (1993).

[6] R. Ofierzynski, J. Phys. Conf. Ser. **110**, 042019 (2008).

[7] J. C. Kapteyn, Astrophys. J. **55**, 302 (1922).

[8] V. Rubin and J. Ford W. Kent, Astrophys. J. **159**, 379 (1970).

[9] P. A. R. Ade, N. Aghanim, M. Arnaud, and M. Ashdown, Astron. Astrophys. **594**, A13 (2016).

[10] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, Astrophys. J. **648**, L109 (2006).

[11] *40th Anniversary of the First Proton-Proton Collisions in the CERN ISR*, CERN Yellow Reports: Conference Proceedings (Geneva, Jan. 2011).

[12] S. van der Meer, (1978).

[13] D. Drijard et al., Phys. Lett. B **85**, 452 (1979).

[14] F. Büsser et al., Phys. Lett. B **56**, 482 (1975).

[15] C. W. Fabjan and N. McCubbin, Phys. Rept. **403-404**, 165 (2004).

[16] C. Rubbia, P. McIntyre, and D. Cline, in Workshop on Producing High Luminosity Proton-Antiproton Storage Rings (1976), pp. 683–687.

[17] G. Arnison, A. Astbury, B. Aubert, C. Bacci, G. Bauer, A. Bézaguet, R. Böck, and T. Bowcock, Phys. Lett. B **122**, 103 (1983).

[18] M. Banner, R. Battiston, P. Bloch, F. Bonaudi, K. Borer, M. Borghini, and J.-C. Chollet, Phys. Lett. B **122**, 476 (1983).

[19] G. Arnison, A. Astbury, B. Aubert, C. Bacci, G. Bauer, and A. Bézaguet, Phys. Lett. B **126**, 398 (1983).

[20] P. Bagnaia, M. Banner, R. Battiston, P. Bloch, and F. Bonaudi, Phys. Lett. B **129**, 130 (1983).

[21] *Status of the Fermilab Main Injector*, IEEE Particle Accelerator Conference (June 2001), pp. 27–29.

[22] S. Abachi, B. Abbott, M. Abolins, and B. S. Acharya (D0 Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).

[23] F. Abe, H. Akimoto, A. Akopian, and M. G. Albrow (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).

[24] V. Shiltsev, "Tevatron: The world's first fully superconducting collider and the discovery of the top and bottom quarks", in *Challenges and Goals for Accelerators in the XXI Century* (2016) Chap. 14, pp. 211–222.

[25] T. Aaltonen, V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, and G. D. Alexeev (CDF Collaboration and D0 Collaboration), Phys. Rev. Lett. **109**, 071804 (2012).

[26] G. Aad et al. (ATLAS), Phys. Lett. B **716**, 1 (2012), arXiv:`1207.7214 [hep-ex]`.

[27] S. Chatrchyan et al. (CMS), Phys. Lett. B **716**, 30 (2012), arXiv:`1207.7235 [hep-ex]`.

[28] S. Chatrchyan et al. (CMS), JINST **3**, S08004 (2008).

[29] T. Sakuma and T. McCauley, J. Phys. Conf. Ser. **513**, 022032 (2014).

[30] The CMS Collaboration (CMS), *The CMS magnet project: Technical Design Report*, tech. rep. (Geneva, 1997).

[31] H. A. Lorentz, "La Théorie Électromagnétique de Maxwell et Son Application Aux Corps Mouvants", in *Collected Papers: Volume II* (Dordrecht, 1936), pp. 164–343.

[32] Z. Charifoulline, IEEE Trans. Appl. Supercond. **16**, 1188 (2006).

[33] TheCMS Collaboration, JINST **5**, T03021 (2010).

[34] W. Adam et al., JINST **16**, P02027 (2021).

[35] The CMS Collaboration, JINST **3**, S08004 (2008).

[36] G. L. Bayatian et al. (CMS), (2006).

[37] P. Azzurri, J. Phys. Conf. Ser. **41**, 127 (2006).

[38] L. Borrello, A. Messineo, E. Focardi, and A. Macchiolo, *Sensor Design for the CMS Silicon Strip Tracker*, tech. rep. (CERN, Geneva, 2003).

[39] A. Annenkov, M. Korzhik, and P. Lecoq, Nucl. Instrum. Meth. A **490**, 30 (2002).

[40] The CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, tech. rep. (Geneva, 1997).

[41] A. Benaglia, JINST **9**, C02008 (2014).

[42] A. Benaglia, E. Auffray, P. Lecoq, H. Wenzel, and A. Para, IEEE Trans. Nucl. Sci. **63**, 574 (2016).

[43] J. Mans et al., *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*, tech. rep. (2012).

[44] The CMS Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, tech. rep. (CERN, Geneva, 2023), arXiv:2309.05466.

[45] G. Abbiendi et al., JINST **14**, C12010 (2019).

[46] G. L. Bruno, "The RPC detectors and the muon system for the CMS experiment at the LHC", PhD thesis (Pavia U., 2001).

[47] A. Colaleo, A. Safonov, A. Sharma, and M. Tytgat, *CMS Technical Design Report for the Muon Endcap GEM Upgrade*, tech. rep. (2015).

[48] E. Fontanesi (CMS), *New trigger strategies for CMS during Run 3*, tech. rep. (CERN, Geneva, 2022).

[49] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, B. Arnold, and H. Bergauer, JINST **15**, P10017 (2020).

[50] A. Bocci and on behalf of the CMS Collaboration, J. Phys. Conf. Ser. **2438**, 012016 (2023).

[51] S. Mukherjee (CMS), in 5th Large Hadron Collider Physics Conference (2017), arXiv:1708.06925 [hep-ex].

[52] The CMS Collaboration, *Enriching the physics program of the CMS experiment via data scouting and data parking*, tech. rep. (CERN, Geneva, 2024), arXiv:2403.16134.

[53] J. J. Aubert et al., Phys. Rev. Lett. **33**, 1404 (1974).

[54] J. E. Augustin et al. (SLAC-SP-017 collaboration), Phys. Rev. Lett. **33**, 1406 (1974).

[55] S. W. Herb et al., Phys. Rev. Lett. **39**, 252 (1977).

[56] B. Holdom, Phys. Lett. B **166**, 196 (1986).

[57] F. Compagnin, S. Profumo, and N. Fornengo, JCAP **2023**, 061 (2023).

[58] B. Brahmachari and A. Raychaudhuri, Nucl. Phys. B **887**, 441 (2014).

[59] M. Fabbrichesi, E. Gabrielli, and G. Lanfranchi, *The Physics of the Dark Photon* (Springer Cham, 2020).

[60] G. Branco, F. P.M., L. Lavoura, M. Rebelo, M. Sher, and J. Silva, Phys. Rept., 10.1016/j.physrep.2012.02.002 (2012).

[61] A. M. Sirunyan et al. (CMS Collaboration), Phys. Rev. Lett. **129**, 032001 (2022).

[62] L. Wang, J. M. Yang, and Y. Zhang, Commun. Theor. Phys. **74**, 097202 (2022).

[63] D. Curtin et al., Phys. Rev. D **90**, 075004 (2014), arXiv:1312.4992.

[64] U. Haisch, J. F. Kamenik, A. Malinauskas, and M. Spira, JHEP **03**, 178 (2018), eprint: 1802.02156.

[65] A. M. Sirunyan et al. (CMS), Phys. Rev. Lett. **124**, 131802 (2020), arXiv:1912.04776 [hep-ex].

[66] G. Aad et al. (ATLAS), Phys. Lett. B **796**, 68 (2019), arXiv:1903.06248 [hep-ex].

[67] CMS Collaboration, *Dimuon scouting at CMS*, Detector Performance Summary CMS-DP-2023-070 (2023).

[68] CMS Collaboration, *HLT Dimuon Invariant Mass Distributions In 2017 and 2018*, Detector Performance Summary CMS-DP-2018-055 (2018).

[69] T. Sjöstrand et al., Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].

[70] J. Alwall et al., JHEP **07**, 079 (2014), arXiv:1405.0301.

[71] A. M. Sirunyan et al. (CMS), Eur. Phys. J. C **80**, 4 (2020), arXiv:1903.12179 [hep-ex].

[72] S. Agostinelli et al. (GEANT4), Nucl. Instrum. Meth. A **506**, 250 (2003).

[73] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Nucl. Instrum. Meth. A **543**, 577 (2005), arXiv:physics/0408124.

[74] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Phys. Rept. **810**, 1 (2019).

[75] Y. Freund and R. E. Schapire, JCSS **55**, 119 (1997).

[76] The CMS collaboration (CMS), JINST **7**, P10002 (2012), arXiv:1206.4071.

[77] H. Albrecht, U. Binder, G. Harder, I. Lembke-Koppitz, and A. Philipp, Phys. Lett. B **162**, 395 (1985).

[78] The CMS collaboration, JINST **13**, P06015 (2018).

[79] G. Breit and E. Wigner, Phys. Rev. **49**, 519 (1936).

[80] J. E. Gaiser, "Charmonium spectroscopy from radiative decays of the J/Ψ and Ψ'", SLAC Report SLAC-R-255, PhD thesis (Stanford University, 1982).

[81] M. J. Oreglia, "A study of the reactions $\Psi' \to \gamma\gamma\Psi$", SLAC Report SLAC-R-236, PhD thesis (Stanford University, 1980).

[82] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, JINST **10**, P04015 (2015), arXiv:1408.6865 [physics.data-an].

[83] R. A. Fisher, J. R. Stat. Soc. **85**, 87 (1922).

[84] S. Bernstein, Comm. Kharkov Math. Soc. **13**, 1 (1912).

[85] The CMS Collaboration, *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ *TeV*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001 (2018).

[86] The CMS Collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-LUM-18-002 (2019).

[87] S. Camarda et al., Eur. Phys. J. C **80**, 251 (2020), arXiv:1910.07049 [hep-ph].

[88] The LHCb Collaboration, JINST **3**, S08005 (2008).

[89] B. Aubert et al., Nucl. Instrum. Meth. A **479**, Detectors for Asymmetric B-factories, 1 (2002).

[90] R. Aaij et al. (LHCb), Phys. Rev. Lett. **120**, 061801 (2018), arXiv:1710.02867 [hep-ex].

[91] J. P. Lees et al. (BaBar), Phys. Rev. Lett. **113**, 201801 (2014), arXiv:1406.2980 [hep-ex].

[92] P. Ilten, Y. Soreq, M. Williams, and W. Xue, JHEP **06**, 004 (2018), arXiv:1801.04847 [hep-ph].

[93] R. Aaij et al. (LHCb), JHEP **10**, 156 (2020), arXiv:2007.03923 [hep-ex].

[94] The CMS collaboration, JHEP **70**, 10.1007/JHEP12(2023)070 (2023).

[95] R. Aaij et al. (LHCb), Phys. Rev. Lett. **124**, 041801 (2020), arXiv:1910.06926 [hep-ex].

[96] E. Hazen, A. Heister, C. Hill, J. Rohlf, S. X. Wu, and D. Zou, JINST **8**, C12036 (2013).

Curriculum Vitae

Christopher Cosby

Contact Information

Email: christophercosby0@gmail.com
cosbyc@bu.edu

Phone Number: (425) 591-7989

Local Address: 833 Beacon St. Apt. C
Boston, MA, 02215

Permanent Address: 14626 168th Ave NE
Woodinville, WA, 98072

Education

Boston University, Boston, MA

- Ph.D. in Physics **May 2024**

  Thesis Title: *A search for GeV-scale dark mediators decaying to a pair of muons using data from the CMS detector*

University of Washington, Seattle, WA

- B.Sc. in Physics **June 2018**

Research Achievements

CMS Experiment, *CERN, Geneva, Switzerland* **Sep. 2018 –**

Hardware, Software, and Operations

- Maintainer and developer for AMC13, a high-speed data assembly and clock distribution electronics card. Responsibilities included software development for the slow control software and tools, maintaining the git repository to build and publish the software, as well as hardware work to test and debug at BU the fleet of modules utilized by CMS.
- Was the point of contact for HCAL operations of the AMC13. Frequently interfaced with HCAL leads to diagnose AMC13 related issues, or make updates to the HCAL AMC13 control software.
- Developed and implemented NIM+, an FPGA based replacement for standard NIM modules, as a tool for triggering HGCAL hexaboards during testbeam, developed in conjunction with FNAL engineers.

Physics Analysis

- Performed an analysis with online-reconstructed muons to search for dark photons at low masses. The limits from this analysis are currently the worlds best limits on prompt dimuon resonance production for masses between 1 and 8 GeV.
- Acted as the muon object contact for the exotica physics group, reviewing early stage analyses that made use of muons and green-lighting their methodology prior to pre-approval.

DarkQuest Experiment, *Fermilab, Batavia, Illinois USA* **Mar. 2023 –**

- Developing readout electronics for the EMCal detector; the primary feature of the DarkQuest upgrade to the SpinQuest spectrometer.
- Calibrating and characterizing of EMCal module performance prior to test-beam at FNAL in late 2023.
- Designed and implemented a SiPM based cosmic muon trigger for energy response calibrations.

Select Publications

The CMS Collaboration, **Search for prompt production of a GeV scale resonance decaying to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV**, J. High Energ. Phys. **12** (2023) 070, arxiv:2309.16003.

The CMS Collaboration, **Measurement of differential cross sections for the production of a Z boson in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV**, arxiv:2205.02872. Submitted to PRD.

The CMS Collaboration, **Dimuon scouting at CMS**, CMS Technical Report, CMS-DP-2023-070.

Teaching Experience

Boston University, *Teaching Fellow*

- Physics 581 "Advanced Lab", Fall 2018
- Physics 581 "Advanced Lab", Fall 2019
- Physics 681 "Electronics Lab", Spring 2019
- Physics 252 "Electromagnetics", Spring 2020

University of Washington, *Teaching Assistant*

- Physics 122 “Electromagnetism”, Winter 2013
- Physics 121 “Mechanics”, Spring 2014
- Physics 123 “Waves, Light, and Heat”, Fall 2014
- Physics 122 “Electromagnetism”, Spring 2015

Outreach Activities

Boston University Undergraduate Mentorship Program, 2021-2022

- Volunteered to pair with a BU undergraduate majoring in physics, to meet with them throughout the year to offer advice about how to advance their career into into graduate school.

University of Washington SPS High School Outreach Initiative, 2016-2017

- Travelled with other undergrads from UW SPS to local high-schools, to speak with students about physics as a field of study and potential career path. Included physics ”demos” presented to groups, as well as some one-on-one conversations with interested students. A focus was placed on visiting diverse and underserved high-schools in the greater Seattle area.

Pacific Science Center, Science Communicator, 2014-2017

- Performed interactive shows and experiences to families and school groups at Pacific Science Center, a non-profit science museum located in Seattle. The goal of this work was to foster scientific interest and literacy in guests of all ages, by breaking down complex topics and presenting them in entertaining, easily digestible ways. This included learning to work one-on-one with guests who had varying levels of learning disabilities, to ensure that learning about science could be fun and accessible for everyone.