Study of the Vector Boson Scattering $W^{\pm}W^{\pm}$ Production at $\sqrt{s} = 13.6$ TeV

Monika Ghimiray
Supervised by
Michał Szleper
National Centre for Nuclear Research (NCBJ)
PhD seminar 2025

18 December 2025

Outline

- Goals of the project
- Introduction
- CMS Experiment
- Details of the analysis
- Results
- EFT (Effective Field Theory) interpretation.

Goals of the Project

Study of the Vector boson scattering (VBS) process for $W^{\pm}W^{\pm}$ with Run3 data from CMS at CERN.

- Develop analysis code for Run 3 studies.
- Improve the nonprompt background estimation.
- Observe EW $W^{\pm}$ $W^{\pm}$ poduction at $\sqrt{s} = 13.6\ TeV$
- Perform differential cross-section measurement for EW $W^{\pm}$ $W^{\pm}$ process.

Perform an Effective Field Theory (EFT) interpretation.

- Especially focused in understanding the quartic-gauge couplings with dim-8 operators.
- Study the Interference and quadratic terms separately.
- Application of the full clipping method for CMS analysis, only using the interference term (first-time).
- Exploring a wide range of kinematic variables to enhance sensitivity to interference terms in the EFT expansion.

Hunt for New physics: Direct and indirect searches

Direct Searches

- Supersymmetry
- Long-lived particles
- Dark matter
- Heavy resonances

Indirect Searches

- Check for standard model deviations: measurement of known quantities using high-energy physics experiments. Ex:
  - precise measurement of the properties of the Higgs boson
  - study of the high-energy behavior of W and Z bosons (VBS)

Electroweak Sector of Standard Model physics

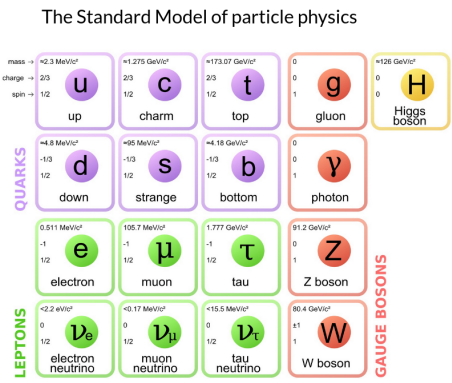
The Standard Model is a theory that describes:

- three of the four fundamental forces — electromagnetic, weak nuclear, and strong nuclear
- all known elementary particles, except for graviton.

The electroweak gauge part of Standard Model

- 2 massive vector bosons: $W^{\pm}$, Z
- 1 massless vector boson: γ
- 2 triple gauge couplings: WWγ, WWZ
- 4 quartic gauge couplings: WWWW, WWZZ, WWZγ, WW$\gamma\gamma$

All the above couplings, as well as W and Z masses, are completely determined by theory via the sole requirement of no tree-level divergences in any allowed processes.



Vector Boson Scattering processes

- Indirect search for new physics.
- VBS is a process of the type: $VV \to VV$ with vector-bosons V = ($W^{\pm}$), Z).
- It is an event where two quarks from the high energy collision of protons, each radiates an electroweak vector-boson, which then scatter and decay.

VBS signature in the detector:

- 2 forward "tagging" jets
- 2 leptons in the central region.

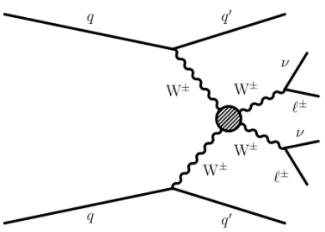
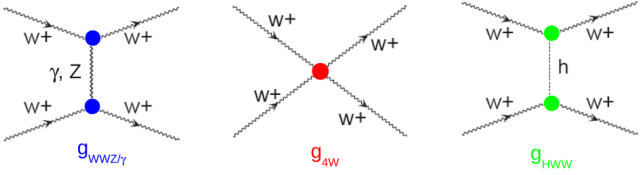
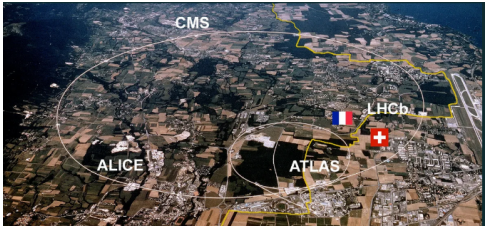


Figure: Same-sign WW process



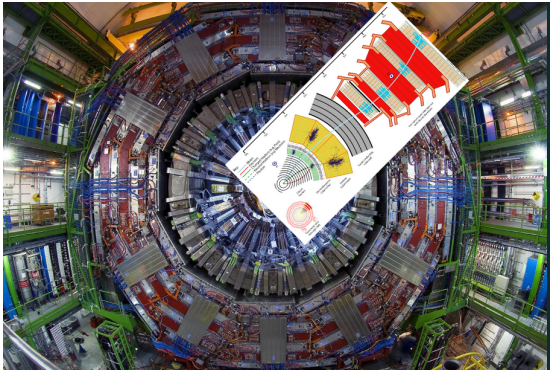
Anomalous (non-SM) couplings ⇔ **New (non-SM) particles**

CMS (Compact Muon Solenoid) Experiment



- One of the largest detectors at the Large Hadron Collider at CERN.
- Approximately 4,000-tonne detector, located about 100 m underground.

- **Main subsystems:**
  - **Silicon tracker** → reconstructs charged-particle tracks and interaction vertices.



- **Electromagnetic calorimeter** → measures the energy of electrons and photons.
- **Hadronic calorimeter** → measures hadron energy
- **Superconducting solenoid magnet** → bend tracks and measure momentum.
- **Muon detector** → identifies muons

Kinematic Selections $W^{\pm}W^{\pm}$

We define one **signal region** to enhance the electroweak $W^{\pm}W^{\pm}$ signal, and **control regions** to validate and constrain the main backgrounds.

Signal Region (Signal-enriched)

- invariant mass of the two leading jets $>$ 500 GeV (jets from high energy processes)
- $|\Delta\eta_{jj}| > 2.5$ (Sufficient separation in rapidity)
- Zeppenfeld variable < 1.0 (Centrally located lepton)
- $p_T^{miss} > 30$ GeV (Neutrino must be present)
- Anti b-tagging Applied (Remove leptons from $t\bar{t}$)

Control Regions (Background-enriched)

- Anti b-tagging Applied/Inverted
- Relax selected VBS requirements (e.g. m_{jj} and/or $|\Delta\eta_{jj}|$) to increase background contribution.

Three Control Regions Used:

- QCD enriched b-tag inverted region: Region dominated by non-prompt leptons and QCD WZ backgrounds.
- QCD enriched b-tagged region: Region dominated by QCD and non-prompt lepton background.
- Non-prompt region: Dominated by non-prompt lepton background.

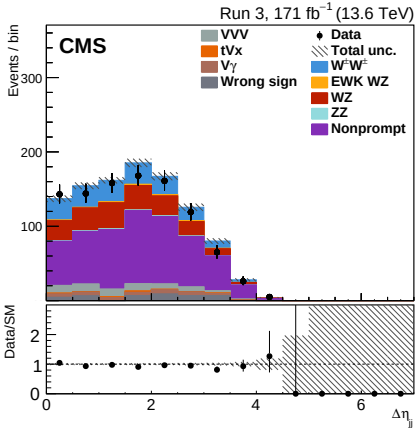
Backgrounds in $W^{\pm}W^{\pm}$ processes

- Non-prompt leptons (Improvements introduced)
- EW induced and QCD induced WZ processes

  One of the lepton from Z decay can be misidentified as from the decay of W boson.
- Wrong sign leptons

  Charge misidentification of one of the leptons makes an OS pair appear as a SS pair (detector reconstruction).
- QCD induced ssWW processes

  Similar final states

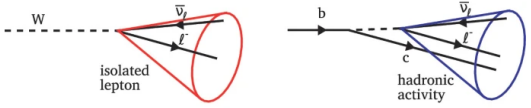


Prompt vs non-prompt leptons

- Prompt:
  - Leptons from W/Z bosons.
  - Properly isolated leptons.
- Non-prompt:
  - leptons from Photon conversions or Hadron decays or any misidentified objects.
  - Not properly isolated.

Data Driven Method

- Jet fragmentation is extremely difficult to model.
- The probability of a jet being detected as a lepton is not modeled well in MC.
- In data-driven method we estimate a background using the data itself.



Non prompt background estimation

- Largest background source for ssWW, not well modeled in MC → need data driven methods
- Two non-prompt background estimation methods

Lepton Selection

- Two categories of leptons selected:
  - Loose: Loosely selected lepton, fakeable objects
  - Tight: used to define signal. The difference wrt. loose is mainly isolation.
- Lepton pairs: TT, TF/FT, FF
  - Tight: passes all selections,
  - Fail: passes loose selection but fails tight selection

Method I: Fakerate method

- The original method used in Run2.
- This method requires all Monte Carlo samples.

Method II: Matrix method

- A bit more complex math → this work.
- Less use of MC.

Non prompt background through matrix method

- Reduces dependency in MCs

  ϵ = fakerate: Efficiency for a fake lepton to pass tight selection

  δ = promptrate: Efficiency for a prompt lepton to fail the tight selection

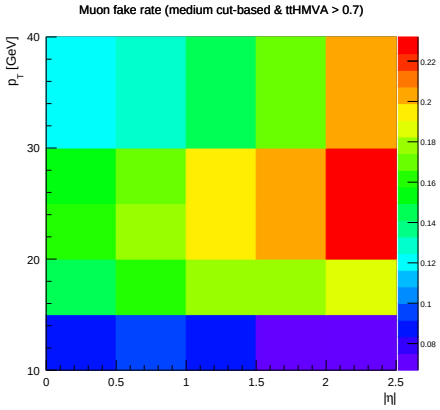
$$\begin{pmatrix} \mathrm{TT} \\ \mathrm{TF} \\ \mathrm{FT} \\ \mathrm{FF} \end{pmatrix} = \begin{pmatrix} (1-\delta_1)(1-\delta_2) & (1-\delta_1)\varepsilon_2 & \varepsilon_1(1-\delta_2) & \varepsilon_1\varepsilon_2 \\ \delta_2(1-\delta_1) & (1-\delta_1)(1-\varepsilon_2) & \varepsilon_1\delta_2 & \varepsilon_1(1-\varepsilon_2) \\ (1-\delta_2)\delta_1 & \delta_1\varepsilon_2 & (1-\varepsilon_1)(1-\delta_2) & (1-\varepsilon_1)\varepsilon_2 \\ \delta_1\delta_2 & \delta_1(1-\varepsilon_2) & (1-\varepsilon_1)\delta_2 & (1-\varepsilon_1)(1-\varepsilon_2) \end{pmatrix} \begin{pmatrix} \mathrm{PP} \\ \mathrm{PN} \\ \mathrm{NP} \\ \mathrm{NN} \end{pmatrix}$$

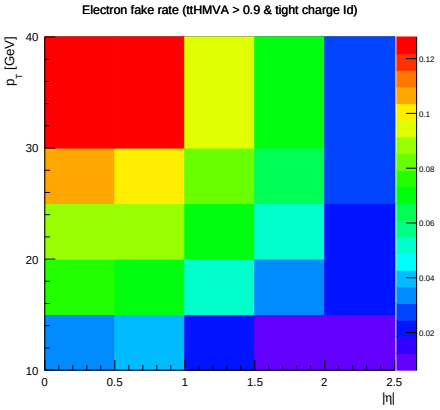
$$PP = \sum_{TT} \frac{(1-\varepsilon_1)(1-\varepsilon_2)}{(1-\varepsilon_1-\delta_1)(1-\varepsilon_2-\delta_2)} - \sum_{TF+FT} \frac{(1-\varepsilon_T)\varepsilon_F}{(1-\varepsilon_T-\delta_T)(1-\varepsilon_F-\delta_F)} + \sum_{FF} \frac{\varepsilon_1\varepsilon_2}{(1-\varepsilon_1-\delta_1)(1-\varepsilon_2-\delta_2)}$$

- The summation runs over the TT, TF+FT and FF events
- PP: prompt-prompt leptons
- After calculating PP, the signal can be estimated as:

  N_{signal}= (1-δ)^{2}PP

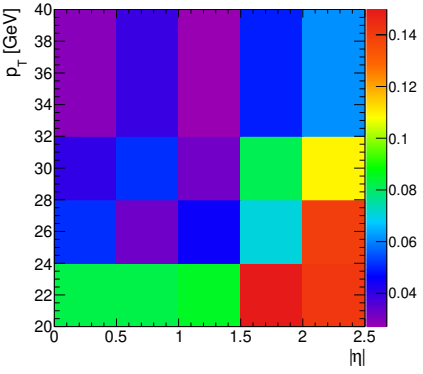
Nonprompt Fake Rates



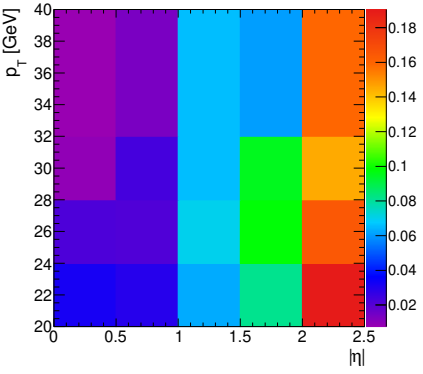


- Efficiency for a fake lepton to pass tight selections.
- Measured in a Non-prompt enriched sample.

Prompt Rates



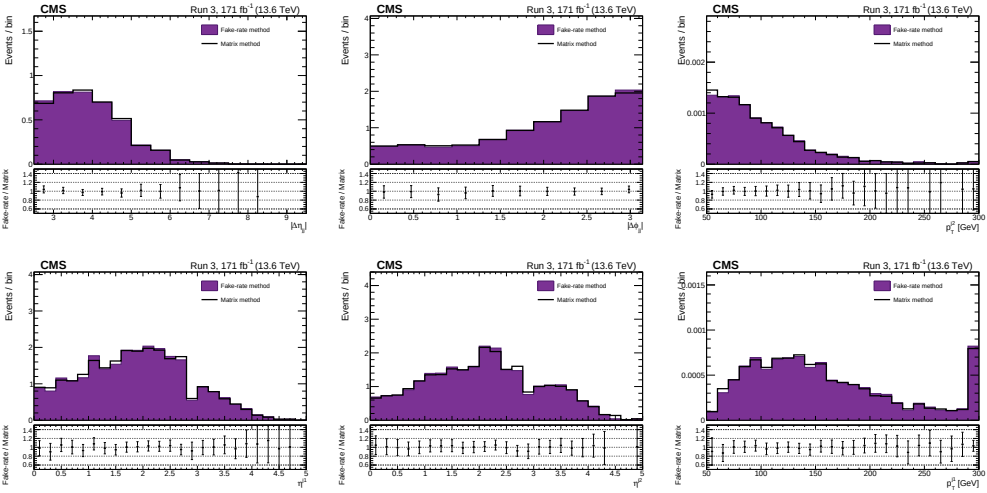
Muons



Electrons

- Efficiency for a prompt lepton to fail tight selections.
- Selections: Similar to VBS signal events.
- Prompt rate is checked on signal MC samples.

Comparing the Nonprompt distributions

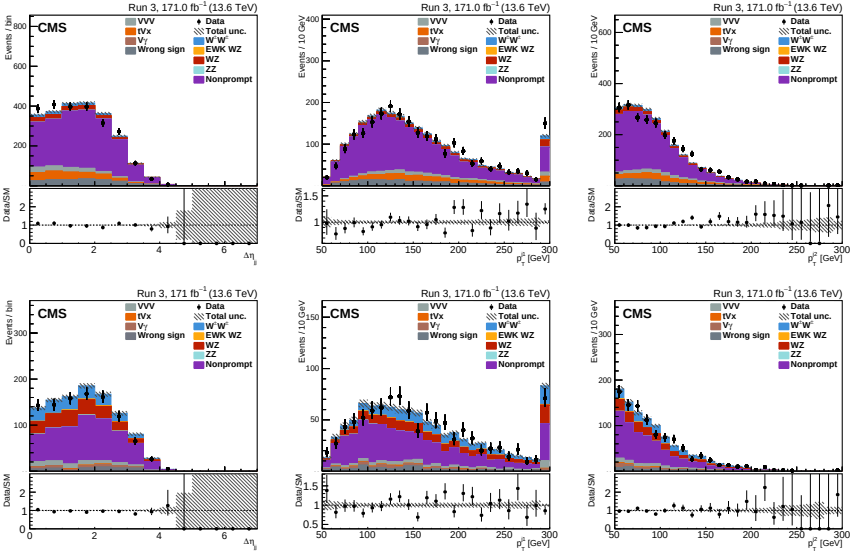


- Consistent nonprompt estimation between the two methods.

Corrections to Simulated Samples

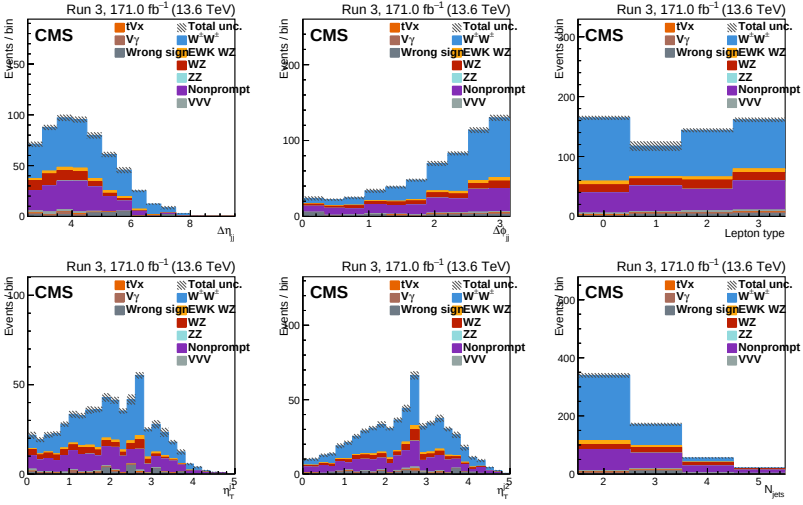
- Pileup
  - Event-by-event pileup weights are applied to MC to match the pileup profile observed in data as a function of the number of interactions/vertices.
- Using updated Jet Energy Corrections:
  - Removes any hot or cold region of the detector and corrects for the jet energy measurements.
- B-tagged corrections
  - The chance to correctly tag a *b*-jet (or mistag a light jet) differs between data and MC.
  - To correct the b-tagging efficiencies in MC to match data.
- Lepton momentum scale & resolution
  - to correct for the lepton scale and smearings.
- Lepton efficiency scale factors
  - Efficiencies of lepton IDs, for electrons and muons.
- Trigger efficiency correction
  - using p_T^{miss} triggers approach.

QCD-enriched CR Distributions b-Tagged and non b-Tagged



Observation: Good Data vs MC agreement.

$W^{\pm}W^{\pm}$ SR Distributions



Observations

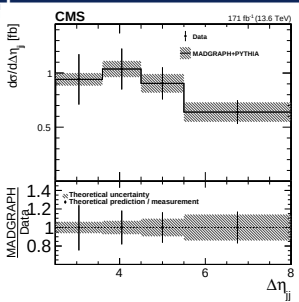
- Signal yield is $> 50\%$.
- Dominant background: non-prompt leptons.

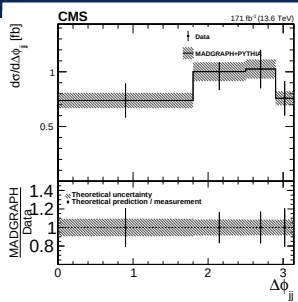
Systematic uncertainties

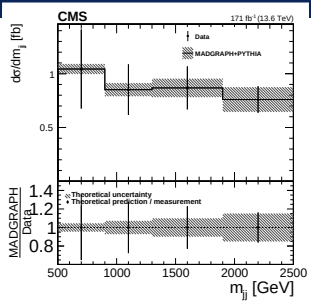
- Systematic uncertainties in μ, the measured inclusive cross sections divided by the predicted cross sections.
- Lepton experimental uncertainties encompass the effects of calibration of lepton momentum scale and resolution, as well as lepton trigger, reconstruction, identification, and isolation efficiencies.
- Jet experimental uncertainties encompass the effects of the jet energy scale and resolution.

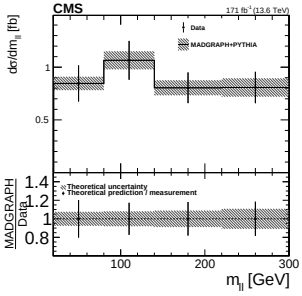
Source of uncertainty	EW WW	EW WZ
Integrated luminosity	0.015	0.013
Lepton experimental	0.012	0.019
Jet experimental	0.010	0.036
b-tagging	0.016	0.002
Nonprompt background	0.004	0.015
Limited sample size	0.033	0.031
Theory	0.020	0.025
Statistical	0.085	0.155
Total	0.097	0.166

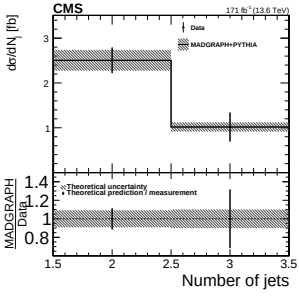
Expected differential cross sections











- The measured absolute $W^{\pm}W^{\pm}$ cross section measurements in different bins
- Shaded bands on data: measurement uncertainties.
- Error bars: combined statistical, PDF, and scale uncertainties.

Summary

- Study of VBS process with CMS Run 3 data has been done.
- Currently analysis is complete using 2022-2024 data sample.
- An alternative method to estimate nonprompt background is applied which gives a consistent picture.
- Reported inclusive differential WW cross section measurements with uncertainties around 10 %.
- Analysis still statistically dominated.
- The analysis note and paper draft is available for CMS members.

What's next?

- Perform unblinding.

Future Plan: EFT interpretation

SMEFT framework

- Standard Model Effective Field Theory (SMEFT): model–independent way to parameterise deviations from the SM.
- For anomalous quartic gauge couplings (QGCs) we focus on dimension-8 operators.

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \frac{c_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \frac{c_i^{(8)}}{\Lambda^4}\mathcal{O}_i^{(8)} + \cdots, \quad \textit{where} \quad f_i^{(8)} = \frac{C_i^{(8)}}{\Lambda^4}, \quad \ldots \quad \textbf{Wilson Coefficient}$$

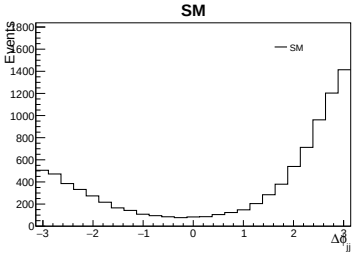
Cross section structure

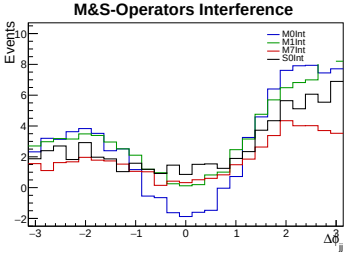
$$\sigma \propto |\mathcal{A}_{\text{full}}|^2 = |\mathcal{A}_{\text{SM}}|^2 + 2\Re(\mathcal{A}_{\text{SM}}\mathcal{A}^*_{\text{dim-8}}) + |\mathcal{A}_{\text{dim-8}}|^2$$

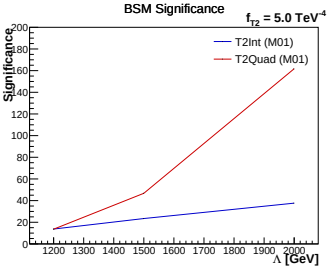
$$\text{with} \quad \mathcal{A}^*_{\text{dim-8}} \sim \frac{c^2}{\Lambda^4}, \quad |\mathcal{A}_{\text{dim-8}}|^2 \sim \frac{c^4}{\Lambda^8}$$

- Interference term: linear in the Wilson coefficients.
- Quadratic term: scales as $1/\Lambda^8$, same as linear dim-12 contributions, so without dim-10 and dim-12 physics interpretation may be severely compromised.

What do we have!!







Comparing SM and BSM contributions:

- Target variables with enhanced SM–BSM interference sensitivity.

BSM Significance

- Interference is generally small $\rightarrow$ challenging. Only $T2$ (and maybe $T1$) show interference comparable to quadratic with high significance.
- Some observables better than $M_{\ell\ell}$ and M_T.
- Different variables are best for different operators.
- For the final analysis we may need to apply Machine Learning.

Summary

- Did pheno studies on SMEFT.
- Identified several variables which provide a better sensitivity.

Tasks planned for the following years

- Include 2022-2026 Data (Full Run3).
- Perform studies on anomalous quartic-gauge couplings.
  - Using dim-8 operators for aQGC studies: S0, T0-T2, M0, M1, M7.
  - First CMS VBS EFT study focusing on interference terms alone.
  - Extend to WZ operators not covered in Run 2: T5-T7 and M2-M5..
  - MC simulation already in progress.
- Search for resonant VBS WW and WZ production
- Polarization measurements

Thank You

Backups

Lepton Selections

- Loose selection to define fakeable object
- Tight/strict selection to define full selected leptons
- Working points and requirements for leptons:

Muons

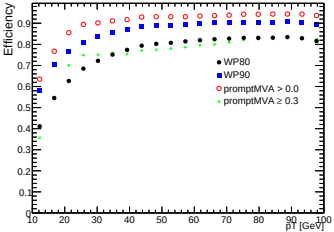
Requirement	Veto	Relax	Strict
Identification	*Loose*	*Medium*	*Medium*
$\lvert d_{xy}\rvert <$ (cm)	–	0.05	0.05
$\lvert d_z\rvert <$ (cm)	–	0.10	0.10
Isolation/p_T	–	0.4	0.4
Jet rel. iso $<$	–	0.5	0.5
promptMVA output	–	–	> 0.7

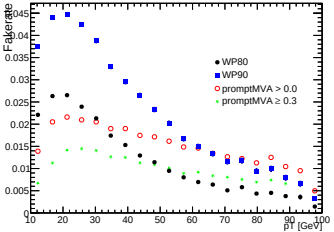
Electrons

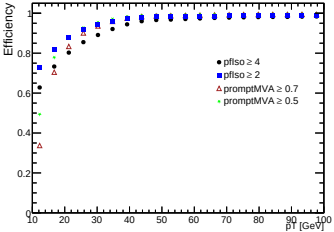
Requirement	Veto	Relax	Strict
ID & isolation	*Veto*	*Medium*	*Medium*
$\lvert d_{xy}\rvert <$ (cm)	–	0.05	0.05
$\lvert d_z\rvert <$ (cm)	–	0.10	0.10
Conversion veto	–	Applied	Applied
Triple charge sel.	–	–	Applied
promptMVA output	–	–	> 0.9

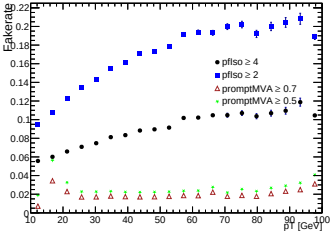
- Electrons: For 2022-23, the promptMVA output is required to be greater than 0.3.
- Discussed earlier during the EGamma meeting here

Lepton Efficiencies









- The cut value differs between the NANOAOD version 12/15

  Electrons:
- For 2024/v15:

  promptMVA $\geq$ 0.8 gives 90% eff.

  promptMVA $\geq$ 0.9 gives 80% eff.
- For 2023/v12:

  promptMVA $>$ 0.0 gives 90% eff.

  promptMVA $\geq$ 0.3 gives 80% eff.

  Muons:

  promptMVA $\geq$ 0.5 gives 90% eff.

  promptMVA $\geq$ 0.7 gives 80% eff.

Fiducial Region

Observable	WW
Lepton origin	Direct decay of a W boson
Lepton definition	Dressed-leptons
First lepton p_{T}	$> 25\,\mathrm{GeV}$
Second lepton p_{T}	$> 20\,\mathrm{GeV}$
$\lvert\eta\rvert$ of leptons	$\lvert\eta\rvert < 2.5$
m_{ll}	$> 20\,\mathrm{GeV}$
Jet p_{T}	$> 50\,\mathrm{GeV}$
$\lvert\eta\rvert$ of jets	$\lvert\eta\rvert^{\mathrm{j}} < 4.7$
m_{jj}	$> 500\,\mathrm{GeV}$
$\Delta\eta_{jj}$	> 2.5
Jet-lepton removal	$\Delta R(j, \ell) > 0.4$

- Selections at the generator level.
- Dressed-leptons: Pre-FSR lepton from the hard process (generator level).

Efficiencies

Filter	Eff. (%)
pass JSON	99.14
Passed trigger	59.04
Only two loose leptons	11.23
Two fake leptons	62.88
Same-sign leptons	1.83
No selected hadronic taus	99.81
ptl1 $>$ 25 && ptl2 $>$ 20	11.55
mll $>$ 20 GeV	98.12
Z veto	100.00
no good b-jets	67.07
At least two VBS jets	9.09
VBS selection	60.00
thePuppiMET_pt $>$ 30.0	100.00

Samples Generated

Monte Carlo Samples: MadGraph 2.9.7 (Simulation tool)

Separate samples for: SM, quadratic term, interference term, and a full sample for verification.

Process: $pp \to jj\ \mu^{\pm}\mu^{\pm}\ \nu\nu$ (Generator level only).

- Only the signal samples are generated and no backgrounds are included.
- Samples are scaled by an additional factor 4 to account for all the lepton (e, μ) combinations.
- Easy to rescale to any desired value:

  - interference term scales linearly
  - quadratic term scales quadratically.
- Previously, getting results for different values of Wilson coefficients (f_i) was time-consuming.

Nominal (generation) values of Wilson coefficients (f_i)

Interference	$\mathbf{f_i}$ [TeV^{-4}]
T2Int	2.0
T1Int	1.0
T0Int	2.0
M0Int	10
S0Int	15
M1Int	15
M7Int	15

Quadratic	$\mathbf{f_i}$ [TeV^{-4}]
T2Quad	2.0
T1Quad	1.0
T0Quad	2.0
M0Quad	10
S0Quad	15
M1Quad	15
M7Quad	15