First look at the Run 3 data for $t\bar{t}HH$ analysis

Hanyang University High Energy Physics Laboratory

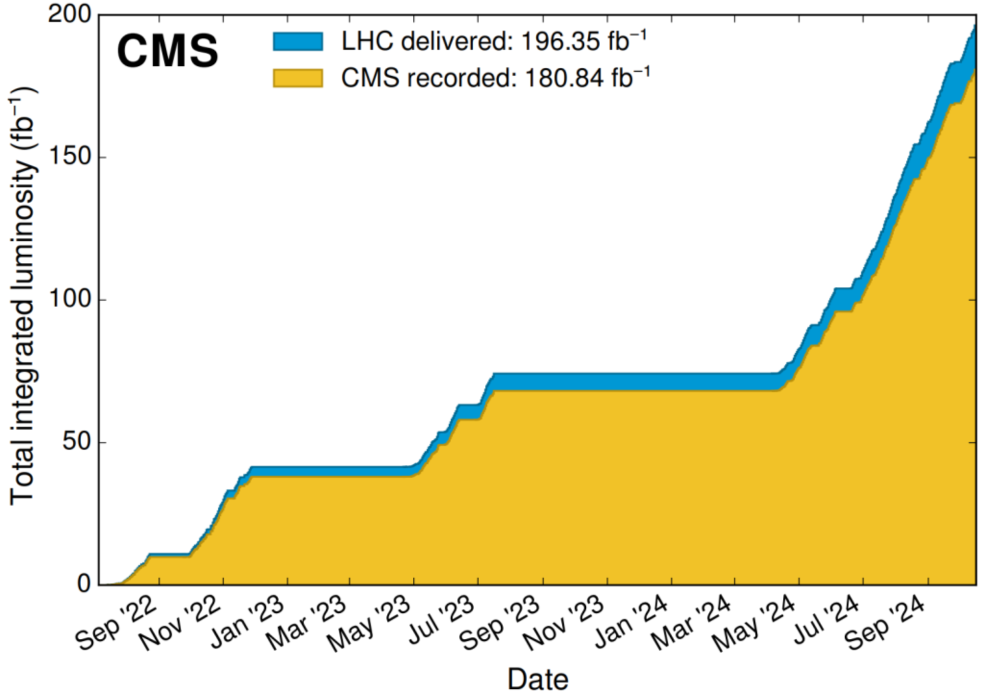
Sungbeom Cho, Taejeong Kim, Tianyi Yang, Yeonsu Ryou

Introduction

- This study aims at searching for di-Higgs events and measuring the triple Higgs coupling in the $t\bar{t}HH$ channel using CMS Run3 data.
- This study focuses on the **di-lepton** and **multi-lepton** **final states** to suppress Drell-Yan and multi-jet backgrounds.

- Currently, we are taking a first look at Run3 data and comparing data with background MC samples in the whole di-lepton and multi-lepton control region.
- Based on this, we are to determine the analysis strategy.

Run 3 dataset and luminosity



- From September 2022 until now, the Run 3 integrated luminosity has already become larger than Run 2.
- The Golden Json luminosity of 2022 and 2023 Runs is listed in the table below.
- The 2024 luminosity is under updating.

Year	2022					2023	
Era	C	D	E	F	G	C	D
first run number	355794	357487	359022	360332	362350	367080	369803
last run number	357486	359021	360331	362180	362760	369802	372415
Golden Json Luminosity /fb	5.0104	2.97	5.807	17.7819	3.0828	17.794	9.451
Total	34.6521					27.245	

Data

- Checking begins with the di-muon region
  - **Muon0** and **Muon1** primary datasets contain **all** the single muon and double muon triggered events.
  - **EGamma0/1** contain all the single and double electron-triggered events.

- Muon0 and Muon1 record different events in the same run and luminosity blocks.
- **No double counting** of events between Muon0 and Muon1.

- Using Run2023D era for the checking
  - Luminosity: $\mathbf{9.451\ fb^{-1}}$

https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVRun3Analysis

MC samples

- Drell-Yan: DYto2L-4Jets_MLL-50_TuneCP5_13p6TeV_madgraphMLM-pythia8
- $t\bar{t}$:
    - DL: TTto2L2Nu_TuneCP5_13p6TeV_powheg-pythia8
    - SL: TTtoLNu2Q_TuneCP5_13p6TeV_powheg-pythia8
- Di-boson:
    - WZ: WZto2L2Q_TuneCP5_13p6TeV_powheg-pythia8
    - WW: WWto2L2Nu_TuneCP5_13p6TeV_powheg-pythia8
    - ZZ: ZZto2L2Nu_TuneCP5_13p6TeV_powheg-pythia
      ZZto2L2Q_TuneCP5_13p6TeV_powheg-pythia8
- W+jets: WtoLNu-4Jets_TuneCP5_13p6TeV_madgraphMLM-pythia8

MC normalization and cut scheme

- ***genWeight*** applied for generator level nominal weights.
- Total evens is normalized to the luminosity before preselection:

$$w_{lumi} = \frac{\text{lumi} \times \text{cross–section}}{N(\text{MC events})}, \text{ where } N(\text{MC events}) = \sum \text{genWeight}$$

- The genWeight distribution for each MC dataset is checked as well.
- Cross-section:

	Cross-section /pb	Additional branching ratio
Drell-Yan 2L	6221.3	
$t\bar{t}$ DL	96.9	
$t\bar{t}$ SL	404.0	
$W \to l\nu$	63199.9	
WW	173.4	$W \to l\nu$: 21.75%
WZ	54.3	$W \to qq$: 67.41%, $Z \to ll$: 6.729%
ZZ	16.7	$Z \to qq$: 69.91%, $Z \to \nu\nu$: 20%

- Cross-section from AN2022_103
- Additional branching ratio from PDG

https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD#Weights

MC Luminosity scaling Run2023D

channel	σ /pb	BR	$\sigma \times$ BR /pb	Events count	$\sum w$	$\sum w^2$	Stats eff events	Expected events	count lumi weight	GenWeight lumi weight	Stats eff lumi / fb^{-1}	Count eff lumi / fb^{-1}
Drell-Yan 2L	6221.3	-	**6221.3**	6.90E+07	9.11E+11	1.20E+16	6.90E+07	5.88E+07	8.52E-01	**6.46E-05**	**11.10**	11.10
$t\bar{t}$ DL	921	10.50%	**96.71**	2.46E+07	1.98E+09	1.62E+11	2.43E+07	9.14E+05	3.71E-02	**4.61E-04**	**250.78**	254.89
$t\bar{t}$ SL	921	45.70%	**420.9**	8.21E+07	2.74E+10	9.32E+12	8.07E+07	3.98E+06	4.85E-02	**1.45E-04**	**191.82**	194.96
$W \rightarrow l\nu$		21.34%	**63199.9**	9.43E+07	1.19E+13	1.50E+18	9.43E+07	5.97E+08	6.33E+00	**5.02E-05**	**1.49**	1.49
WW	116.8	4.55%	**5.32**	6.39E+06	7.53E+07	8.88E+08	6.39E+06	5.03E+04	7.87E-03	**6.67E-04**	**1200.87**	1201.35
WZ	54.3	4.54%	**2.46**	4.27E+06	3.23E+07	2.45E+08	4.26E+06	2.33E+04	5.46E-03	**7.21E-04**	**1728.80**	1732.40
$ZZ \rightarrow 2l2q$	16.7	9.41%	**1.57**	1.49E+07	1.01E+08	6.96E+08	1.47E+07	1.48E+04	9.96E-04	**1.47E-04**	**9371.60**	9487.55
$ZZ \rightarrow 2l2\nu$	16.7	2.69%	**0.45**	1.49E+07	1.54E+07	1.59E+07	1.49E+07	4.25E+03	2.85E-04	**2.76E-04**	**33069.23**	33217.11

Direct count

Compute from weight

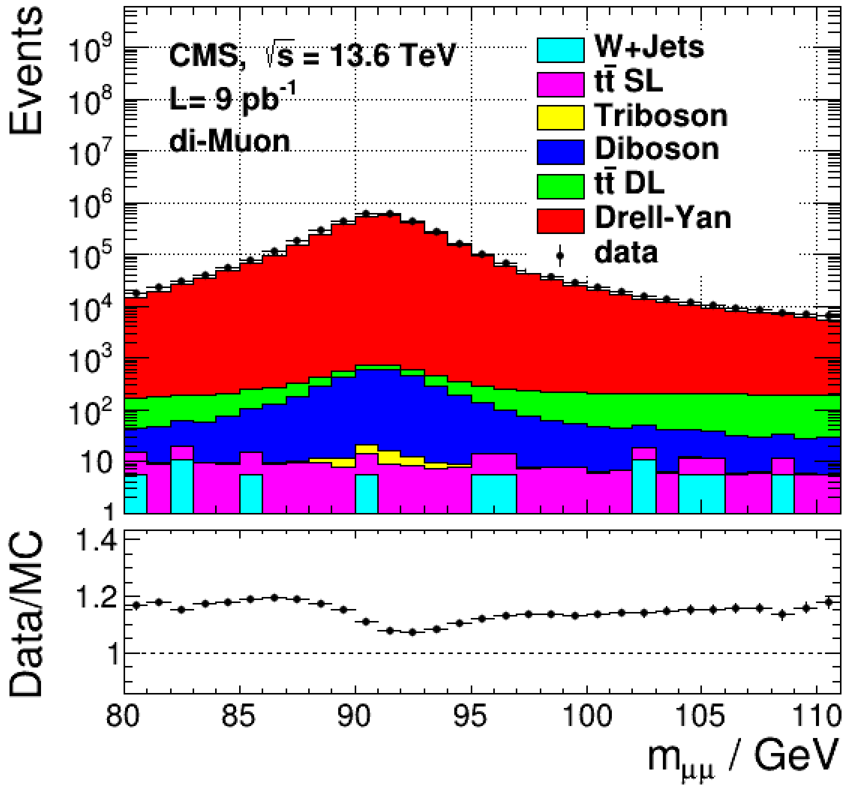
w_{lumi} Applied to events

- Red values are taken from AN2022_103 (Run 3 $t\bar{t}$ cross-section measurement note).
- $t\bar{t}$ cross-section 921 is @NNLO, in the plots in these slides use XS*BR DL 96.6 pb, SL 404.0 pb
- WW cross-section 116.8 is @NNLO, in the plots in these slides use Pythia8 computed XS 173.4 pb.
- Decay branching ratio in third column all taken from PDG2023:
  - $t\bar{t}$: SL 45.7%, DL 10.5%
  - $W \rightarrow qq$: 67.41%, $Z \rightarrow ll$: 6.729%, $Z \rightarrow qq$: 69.91%, $Z \rightarrow \nu\nu$: 20%
- Statistical effective events = $(\sum w)^2/\sum(w^2)$, $\boldsymbol{w}$ is **genWeight** branch.
- Expected events = cross-section * branching ratio / luminosity (9.451 fb^{-1})
- Effective luminosity = number of events / (cross-section * branching ratio)

XSDB values

- The cross-section values of MC samples from XSDB are not accurate enough, providing the under-estimation for MC samples.
- This shows the importance of using the cross-section @NNLO.

	XSDB value/pb	Branching ratio
Drell-Yan 2L	5467	
$t\bar{t}$ DL	762.1	10.5%
$t\bar{t}$ SL	762.1	45.7%
$W \rightarrow lv$	55390	
WW	80.23	
WZ	29.1	
ZZ	12.75	
WWW	0.2328	
WWZ	0.1851	
WZZ	0.06206	
ZZZ	0.01591	



V15 Nano?

- This check uses V12 Nano following the Run 3 Pdmv twiki recommendation.
- V15 Nano mentioned in the CMS week.
- The PPD group report shows NanoAODv15 MC and reReco data ready next year:

CMS PPD

Nov. 2024

Roadmap of 22, 23 and 24 DATA and MC Production

2024	Sep	Oct	Nov	Dec	2025	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2022/2023 pp in NanoAODv15 ; 14_0_X (AOD/SIM) and 15_0_X (MiniAOD+NanoAODv15)														
22/23 Full ReRECO														
22/23 PAG MC														
22/23 SF											-->			

- 2022 and 2023:
  - MC Campaigns: **RunIII2022Summer24** and **RunIII2023Summer24**
    - Ideal detector (no BPix, no EE+)
    - ECAL PF Calib ongoing, PFHC to follow
    - Planned to be launched by December;
  - Data Reprocessing:
    - Pending ECAL PF, PFHC (same as above)
  - Preparations: [gitlab]

Muon and event preselection

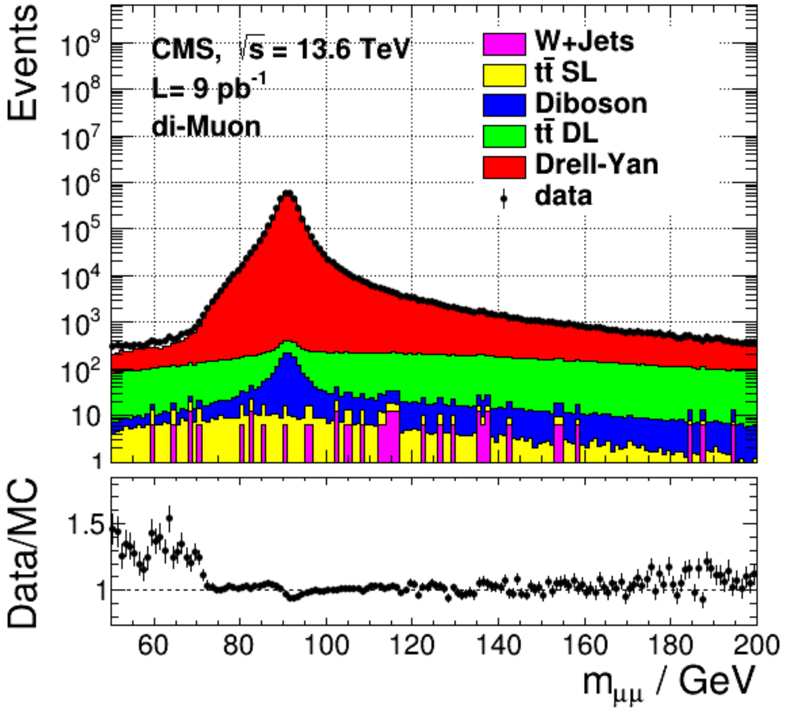
- Good Muon requirement:
  - Tight ID
  - Tight Iso ($I_{0.4}$ < 0.15)
  - Kinematic cut: $p_T > 35\mathrm{GeV}$, $|\eta| < 2.4$
- Events required:
  - **Di-Muon**: Select only events with exactly 2 good Muons.
  - Leading and subleading muon picked in p_T descending order.

  - **Multi-Muon** is also studied later, where requiring more than 2 good Muons.

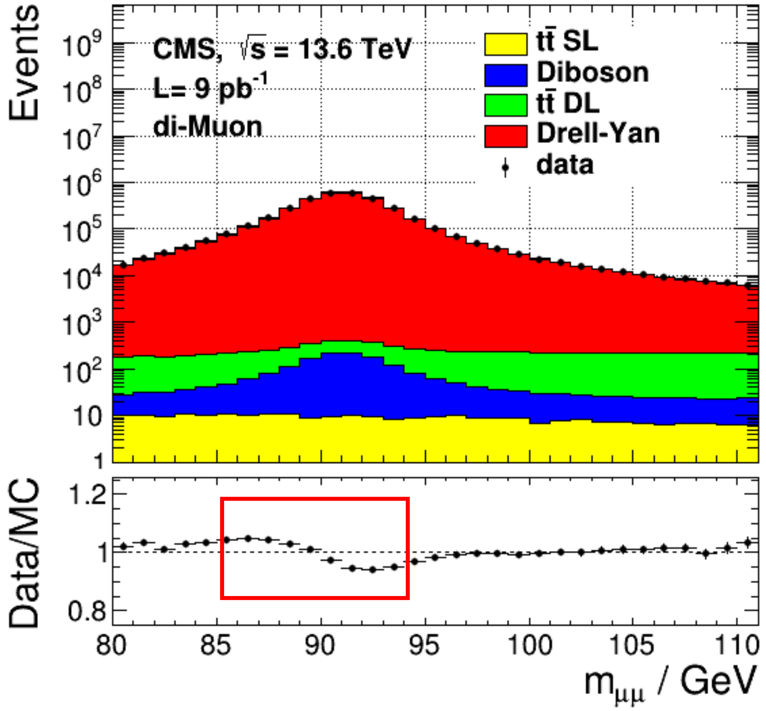
https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonSelection

Di-muon channel $m_{\mu\mu}$ data-MC comparison

$m_{\mu\mu}$ of data & MC

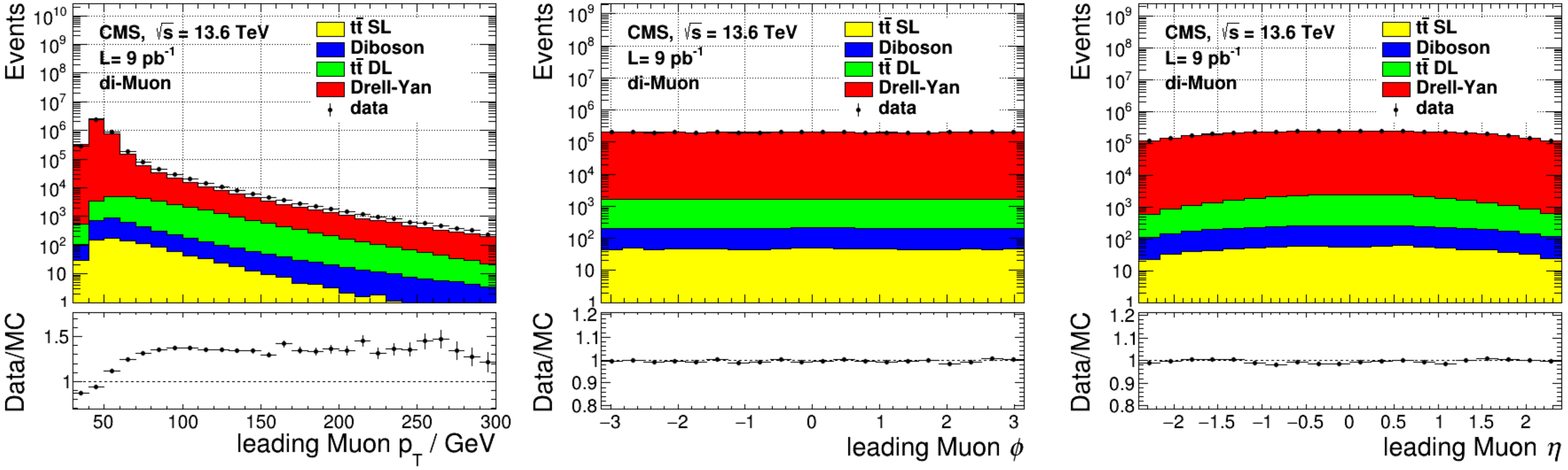


Zoom-in



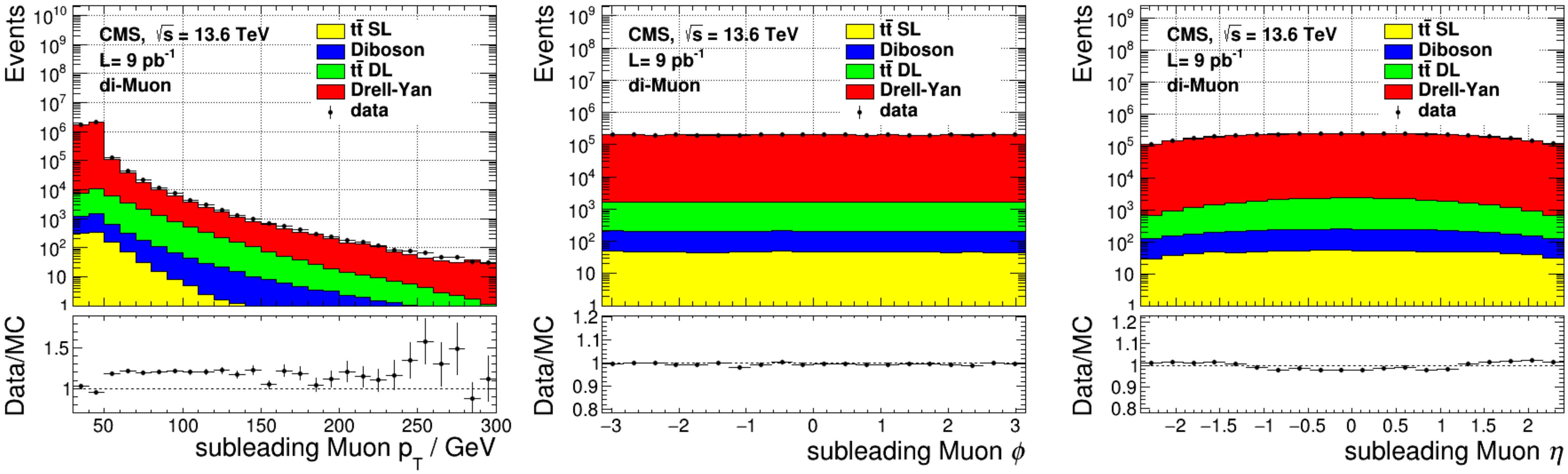
- W+jet and $t\bar{t}$ SL contribution very low, ignore in later plots.
- Missing QCD contribution in the low mass region.
- Zoom-in plot shows clearly the $m_{\mu\mu}$ shift between data and MC.

Leading muon kinematic variables



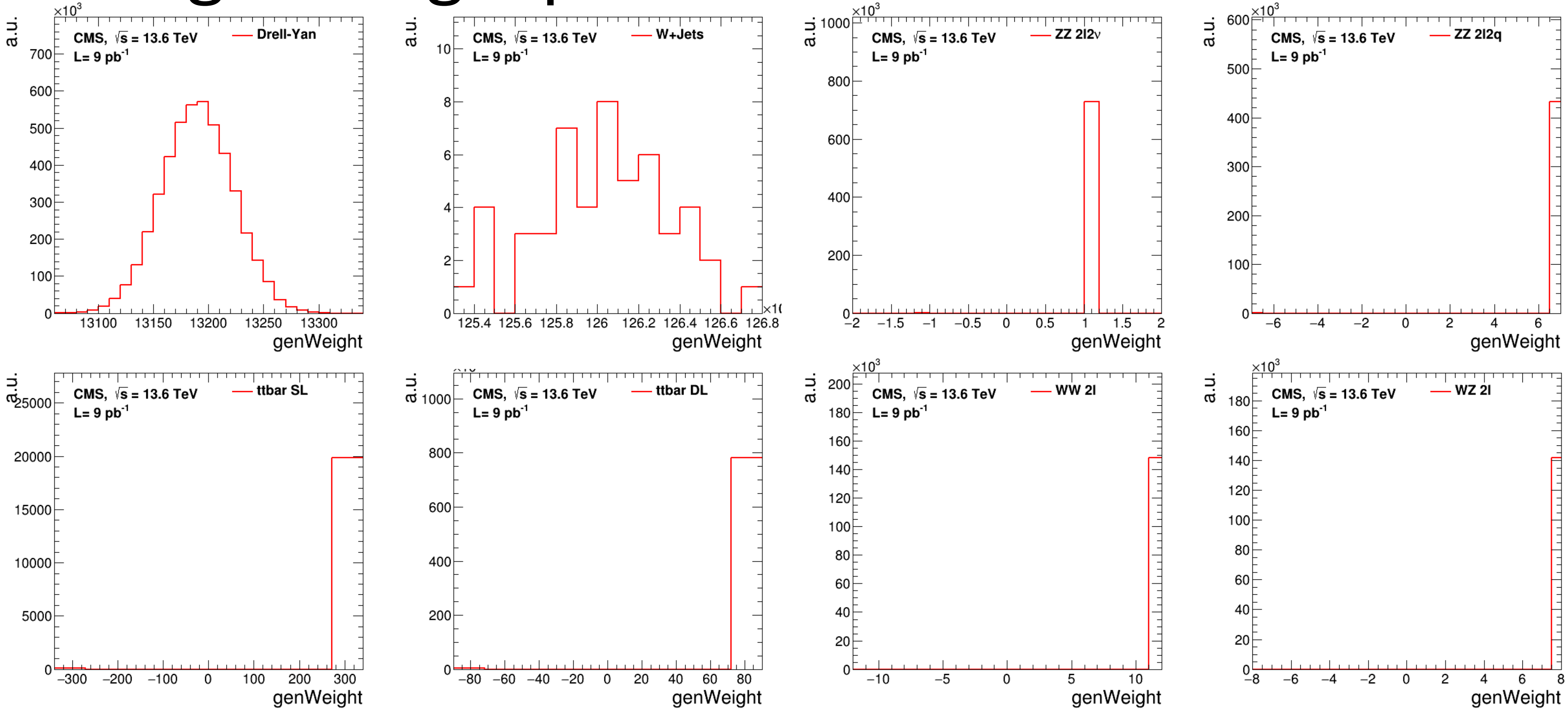
- Difference exists in p_T
- Good matching in ϕ and η.
- Might be solved by Rochester momentum scale correction.

Subeading muon kinematic variables



- Similar behavior to the leading Muon.

MC genWeight plot



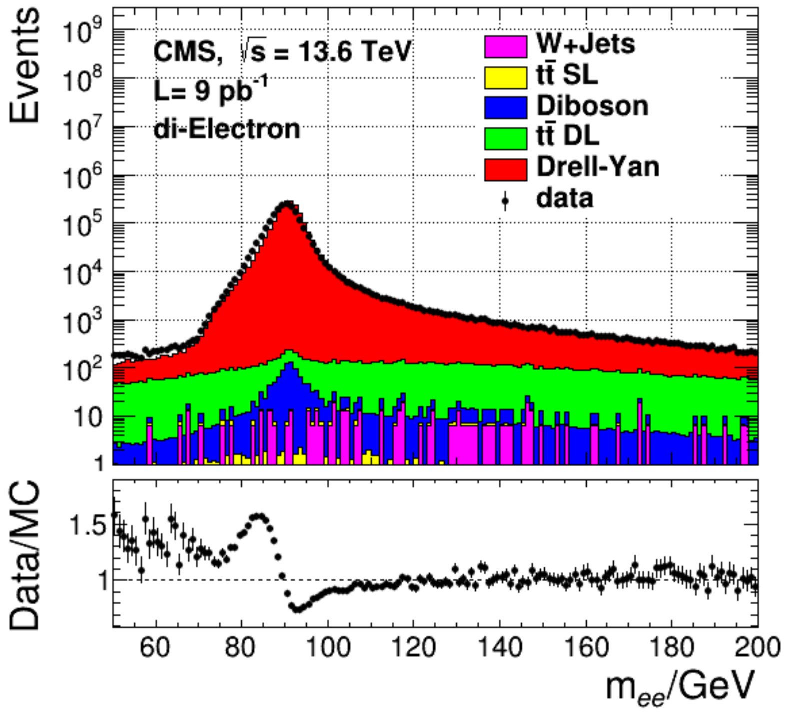
- Drell-Yan and W+jets genWeight are all positive, with small variations.
- Diboson and $t\bar{t}$ have a small number of events with negative weight.

Electron selection

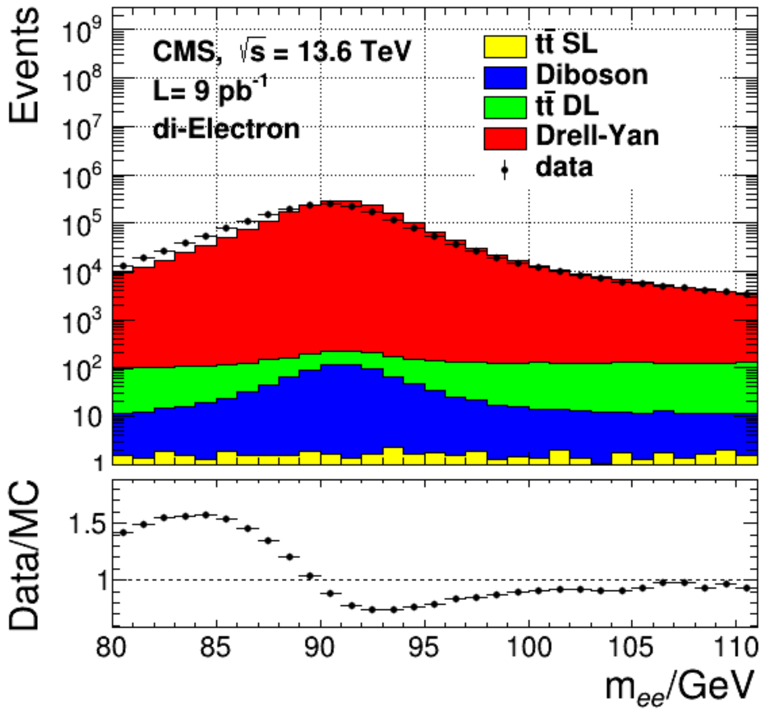
- Good Electron (positron) requirement:
  - Tight cut-based ID
  - Kinematic cut: $p_T > 35\mathrm{GeV}$, $|\eta| < 2.4$
  - Exclude barrel-endcap transition region: $|\eta_{SC}| < 1.442$ or $|\eta_{SC}| > 1.566$
    - Super-Cluster η_{SC} by $\eta + \Delta\eta_{SC}$
- Also, first look at the region with exact 2 good electrons in p_T descending order.

Di-electron channel

m_{ee} of data & MC

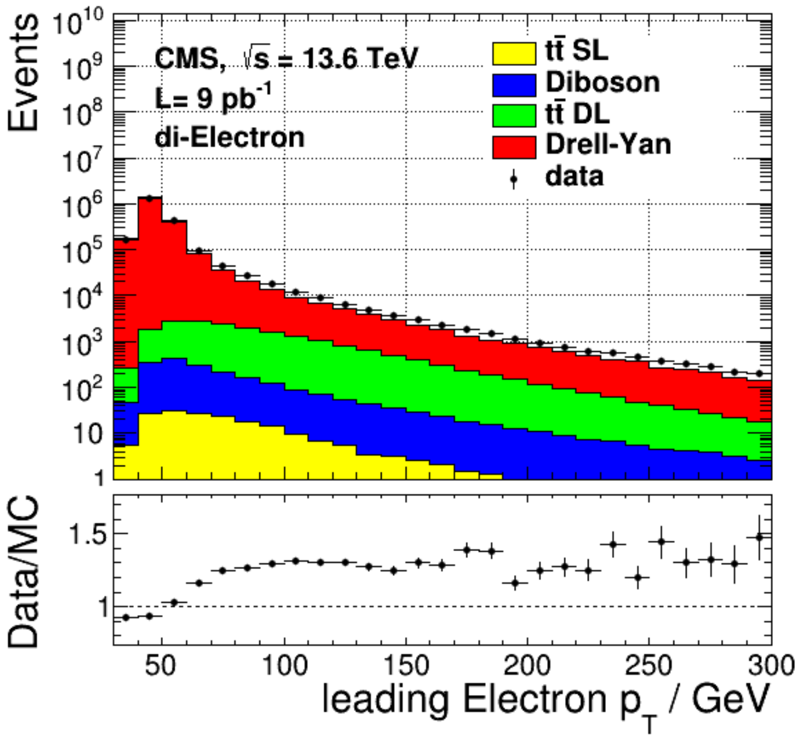


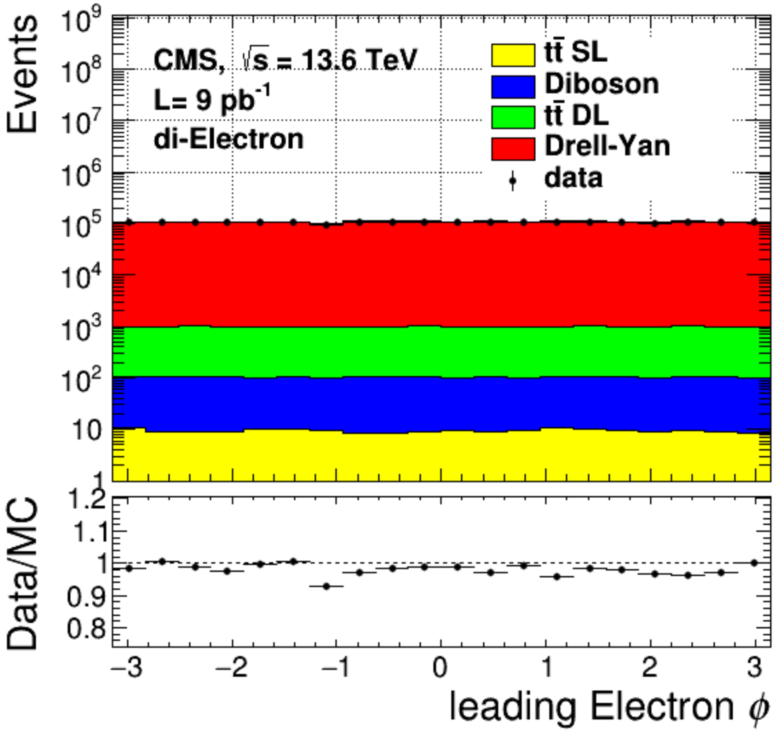
Zoom-in

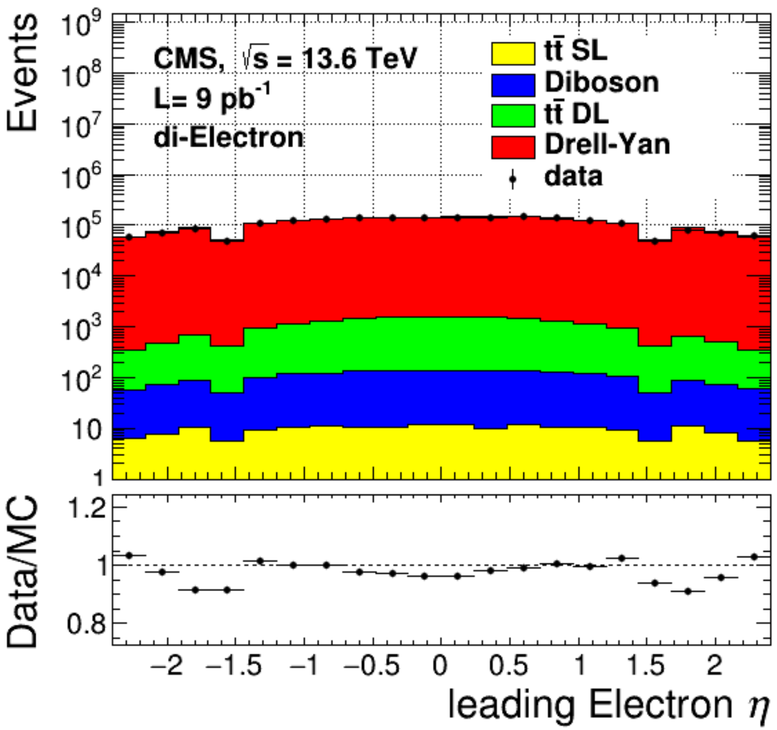


- The issue of Z peak shift is even stronger in the ee channel.

Leading electron kinematic variables

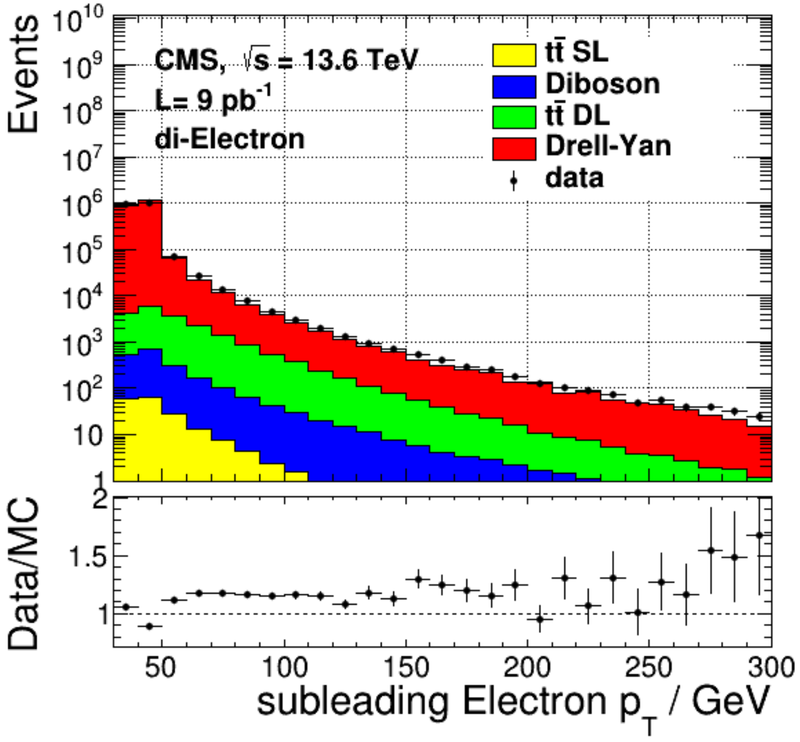


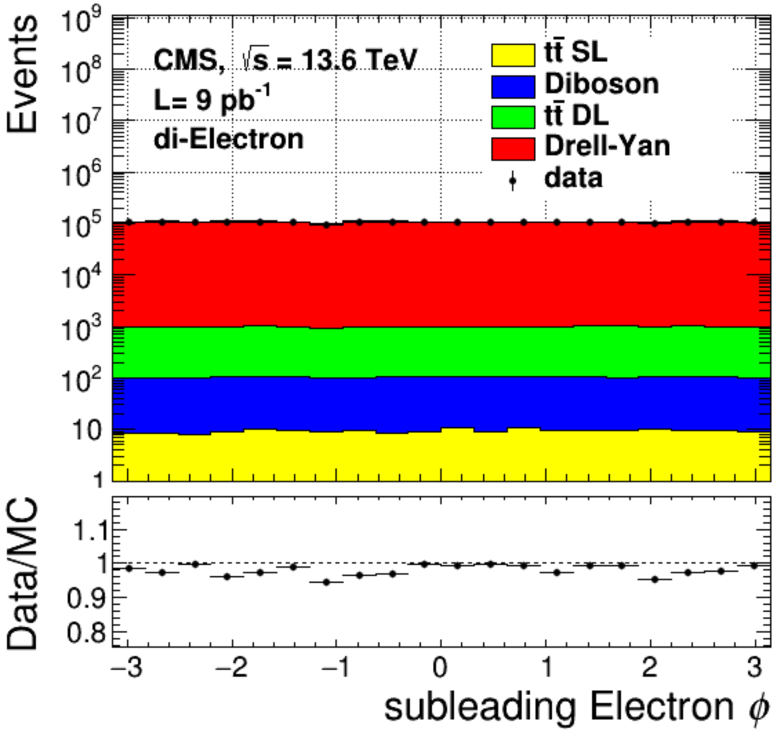


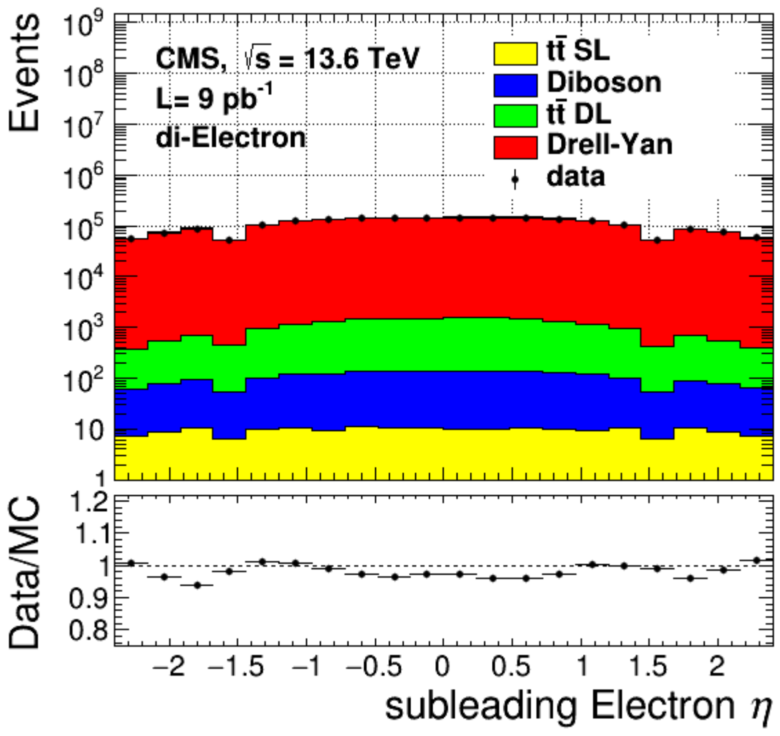


- The similar p_T shifting as di-Muon.
- Good matching in ϕ.
- η has the impact of maybe electron η selection, so it might need scale factors here.

Subleading electron kinematic variables







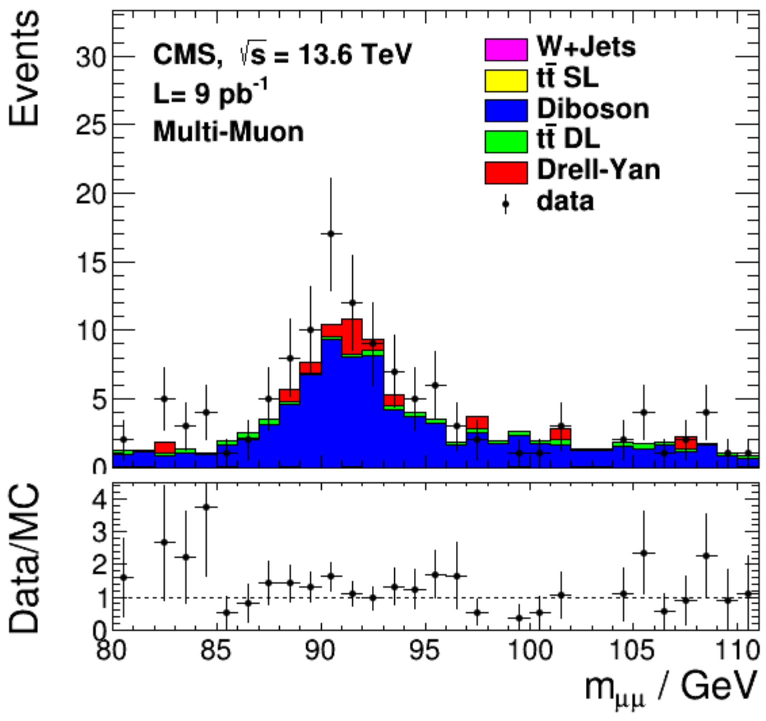
- Also behavior similar to the leading electron.

Multiple leptons

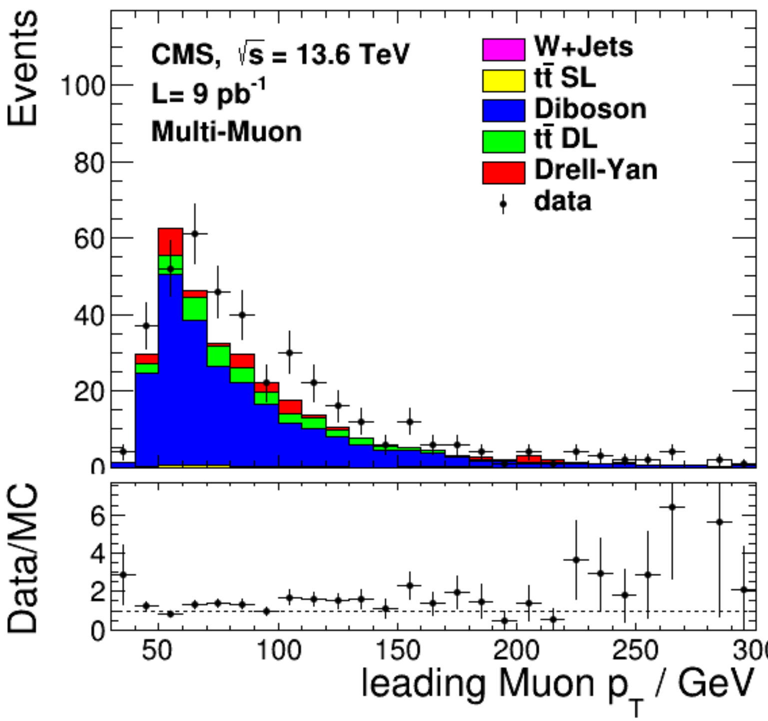
- Multi-muon using the Muon0 and Muon1 datasets.
- Multi-electron using the EGamma0 and EGamma1 datasets.
- Require selected “GoodMuon” or “GoodElectron” at least 3.
- Replace the Diboson samples for multiple leptons contribution:
    - WZ_TuneCP5_13p6TeV_pythia8
    - ZZ_TuneCP5_13p6TeV_pythia8
    - WW_TuneCP5_13p6TeV_pythia8
- These samples have relatively low statistics and thus are not used for di-lepton region comparison.

Multi-muon data-MC comparison

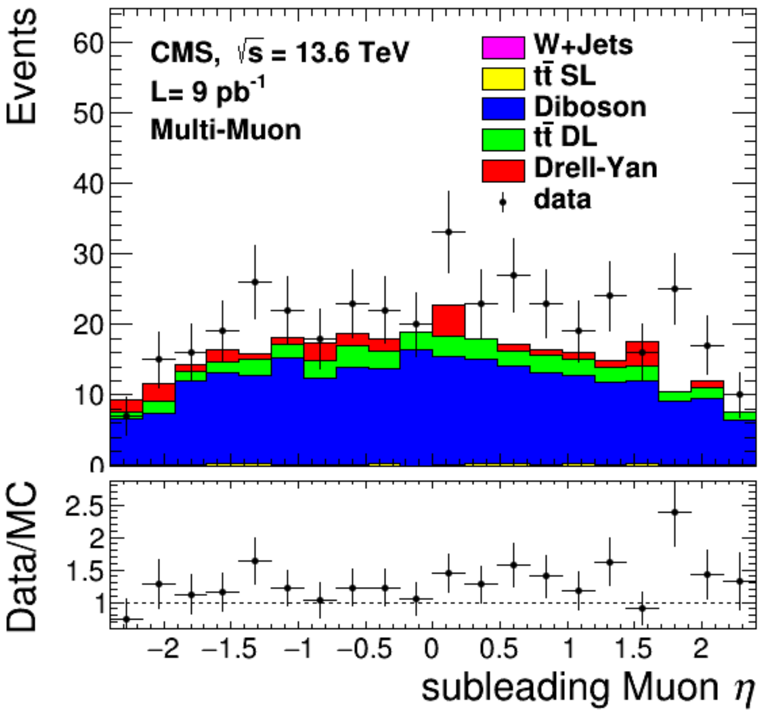
$m_{\mu\mu}$ in [50, 300] GeV



Laeding p_T comparison



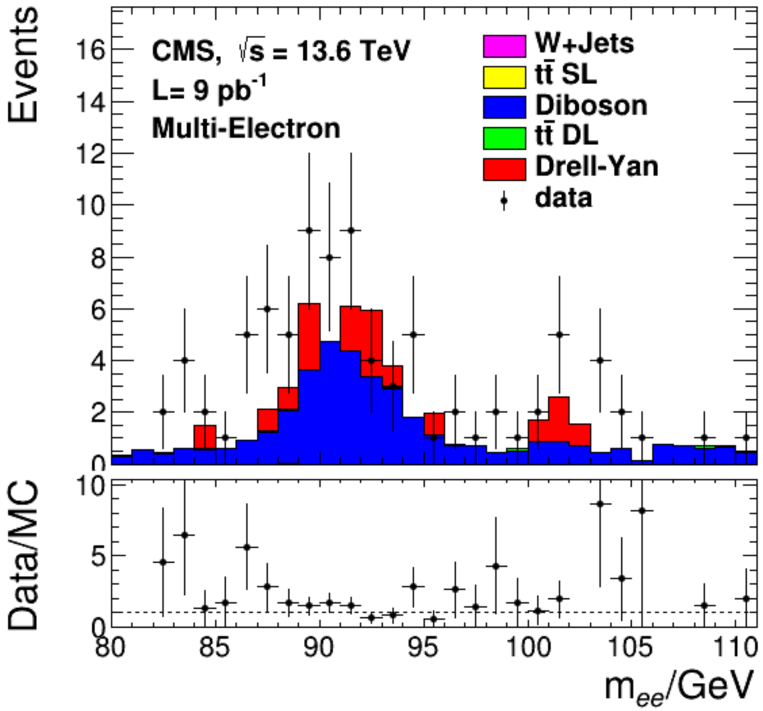
Laeding η comparison



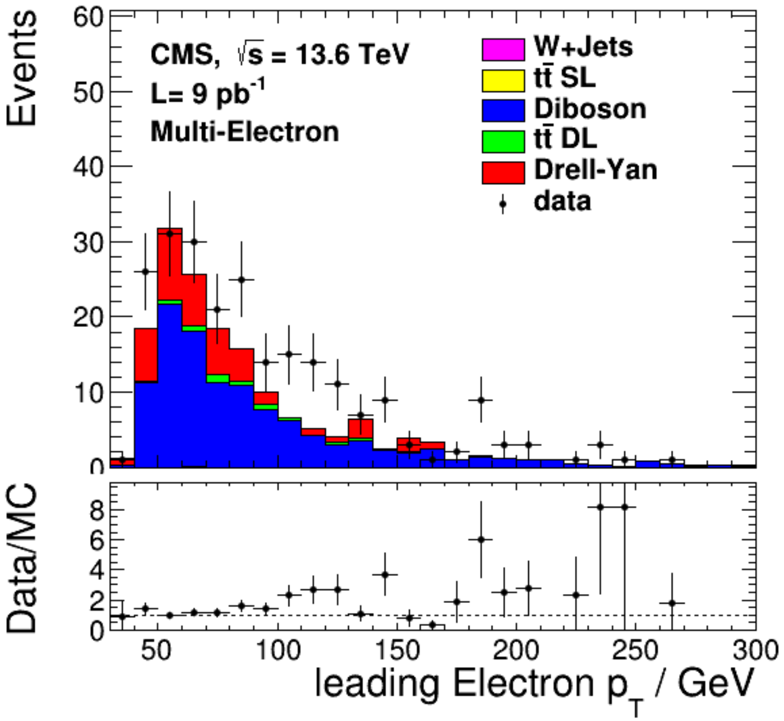
- Data exceeds MC contribution.
- Mis-identified lepton not added, which should also have significant contribution.

Multi-electron data-MC comparison

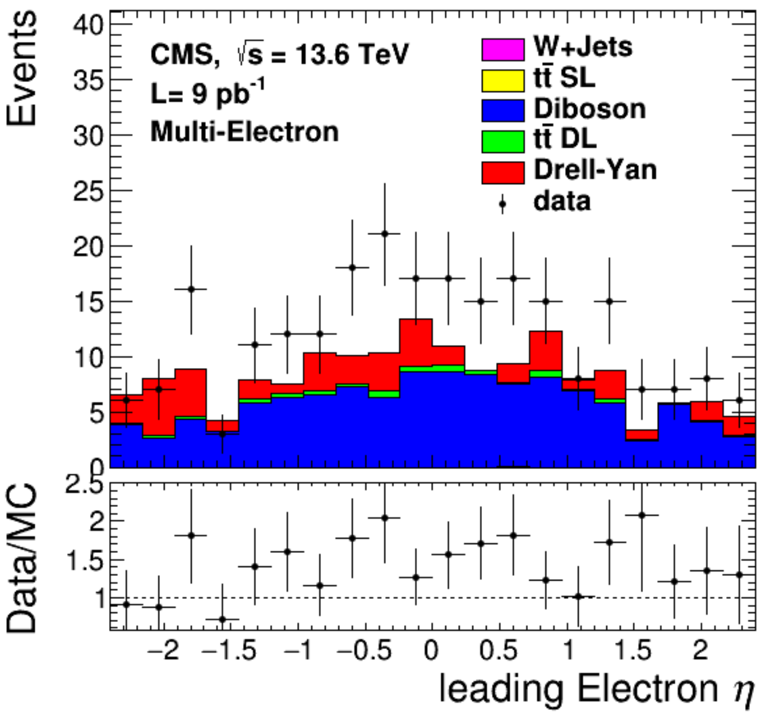
m_{ee} in [50, 300] GeV



Laeding p_T comparison



Laeding η comparison



- The similar gap as in the multi-Muon scenario.

Summary

- The data-MC comparison in di-Muon and di-Electron regions showed a shift in $\boldsymbol{m_{\mu\mu}}$ or $\boldsymbol{m_{ee}}$ peaks.
- The lepton p_T has a shift correspondingly. The lepton η and Muon ϕ matching is good. The electron ϕ matching has a overshooting at the transition region.
- The multi-Muon and multi-Electron regions showed a gap between data and MC, which might be contribution from misidentified leptons.