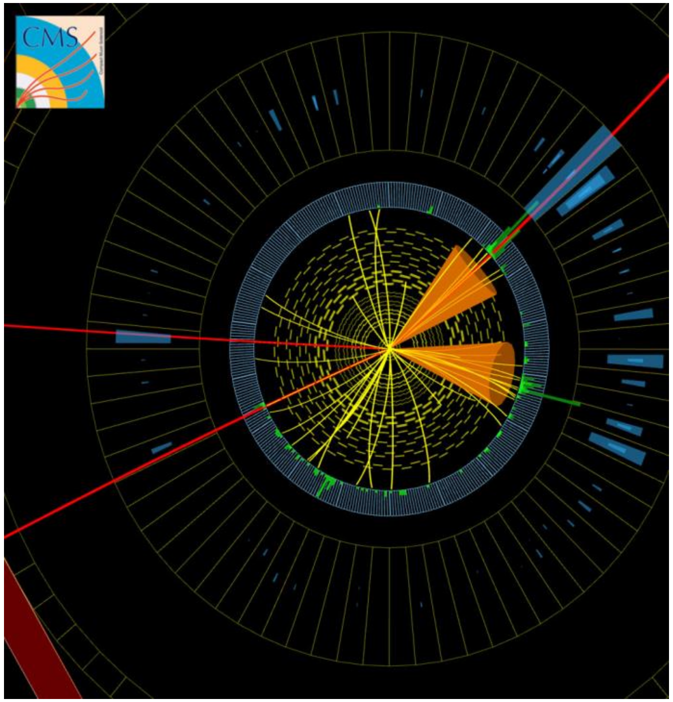
Study of vector boson + heavy flavour jets production in the CMS experiment



Displays of candidate VHcc events,
CMS Collaboration,
Report No. CMS-PHO-EVENTS-2019-006

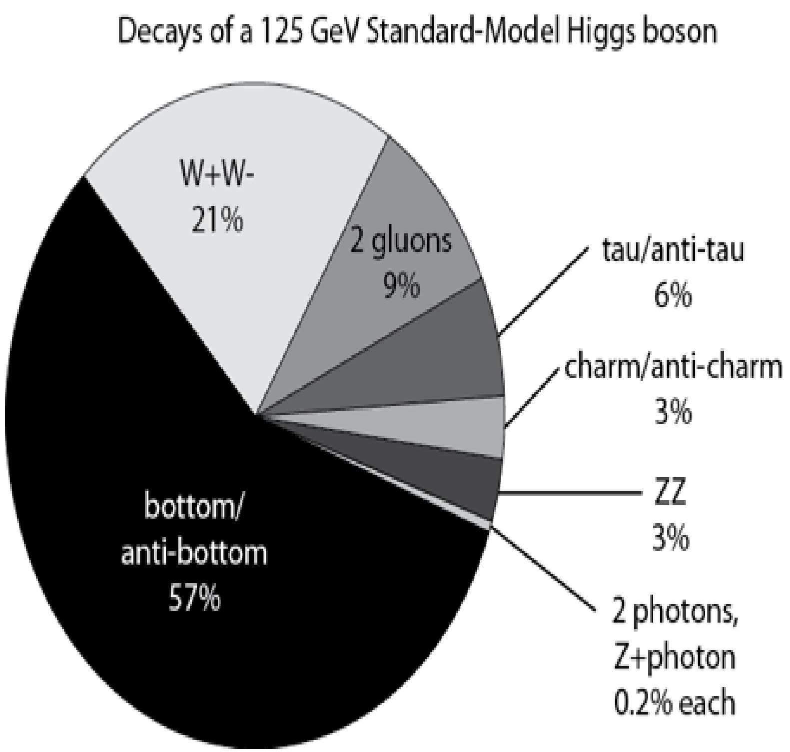
"Test of the standard theory of particle physics by precision measurement at the CMS experiment"

K. Márton, E. Bokor, V. Veszprémi

19 September 2022

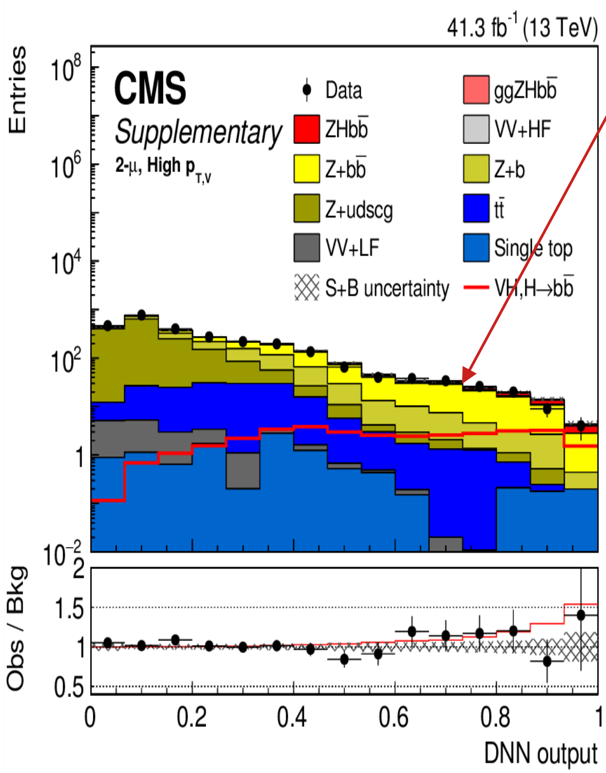
Vector boson + heavy flavour jets

- The associated production of vector bosons and jets in hadronic collisions is a significant background source in several measurements
  - standard model processes
  - Higgs boson studies
    - associated production of Higgs boson and W or Z at LHC
      - $H \rightarrow bb \sim 57\%$
      - $H \rightarrow cc \sim 3\%$
  - searches for physics beyond the SM
- → proper characterization of the V+jets productions and validation of their theoretical description is crucial



V + bb in Higgs measurements

[Phys. Rev. Lett. 121, 121801 (2018)](#)



- V+bb is the single most important background of VHbb measurement, especially at high p^{V}_{T}

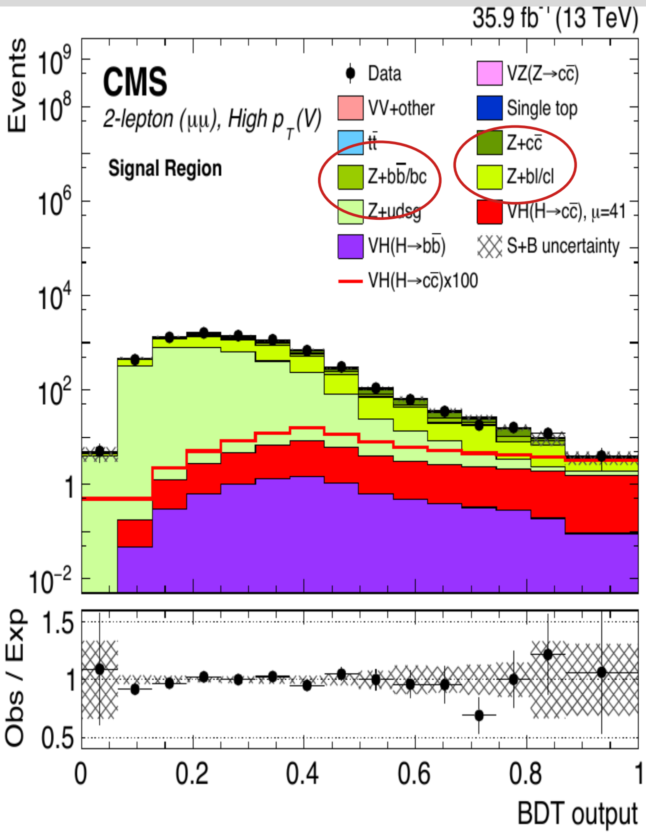
Table 10

The Higgs boson signal, background and data yields for each signal region category in each channel after the full selection of the multivariate analysis. The signal and background yields are normalised to the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties. An entry of "-" indicates that a specific background component is negligible in a certain region, or that no simulated events are left after the analysis selection.

Process	0-lepton		1-lepton		2-lepton			
	$p_T^V > 150$ GeV, 2-b-tag		$p_T^V > 150$ GeV, 2-b-tag		75 GeV $< p_T^V <$ 150 GeV, 2-b-tag		$p_T^V > 150$ GeV, 2-b-tag	
	2-jet	3-jet	2-jet	3-jet	2-jet	≥3-jet	2-jet	≥3-jet
$Z+ll$	17 ± 11	27 ± 18	2 ± 1	3 ± 2	14 ± 9	49 ± 32	4 ± 3	30 ± 19
$Z+cl$	45 ± 18	76 ± 30	3 ± 1	7 ± 3	43 ± 17	170 ± 67	12 ± 5	88 ± 35
$Z+\mathrm{HF}$	4770 ± 140	5940 ± 300	180 ± 9	348 ± 21	7400 ± 120	14160 ± 220	1421 ± 34	5370 ± 100
$W+ll$	20 ± 13	32 ± 22	31 ± 23	65 ± 48	< 1	< 1	< 1	< 1
$W+cl$	43 ± 20	83 ± 38	139 ± 67	250 ± 120	< 1	< 1	< 1	< 1
$W+\mathrm{HF}$	1000 ± 87	1990 ± 200	2660 ± 270	5400 ± 670	2 ± 0	13 ± 2	1 ± 0	4 ± 1
Single top quark	368 ± 53	1410 ± 210	2080 ± 290	9400 ± 1400	188 ± 89	440 ± 200	23 ± 7	93 ± 26
$t\bar{t}$	1333 ± 82	9150 ± 400	6600 ± 320	50200 ± 1400	3170 ± 100	8880 ± 220	104 ± 6	839 ± 40
Diboson	254 ± 49	318 ± 90	178 ± 47	330 ± 110	152 ± 32	355 ± 68	52 ± 11	196 ± 35
Multi-jet e sub-ch.	-	-	100 ± 100	41 ± 35	-	-	-	-
Multi-jet μ sub-ch.	-	-	138 ± 92	260 ± 270	-	-	-	-
Total bkg.	7850 ± 90	19020 ± 140	12110 ± 120	66230 ± 270	10960 ± 100	24070 ± 150	1620 ± 30	6620 ± 80
Signal (post-fit)	128 ± 28	128 ± 29	131 ± 30	125 ± 30	51 ± 11	86 ± 22	28 ± 6	67 ± 17
Data	8003	19143	12242	66348	11014	24197	1626	6686

[Phys. Lett. B 786 (2018) 59](#)

V + cc in Higgs measurements

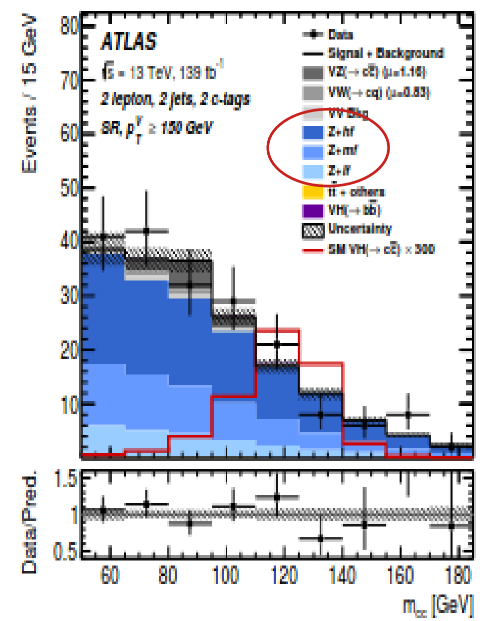


[Submitted to Physical Review Letters]

[Eur. Phys. J. C 82 (2022) 717]

Table 3: Summary of the background modelling systematic uncertainties considered. The values given refer to the size of the uncertainty affecting the yield of each background. Where the size of an acceptance systematic uncertainty varies between analysis regions, a range is displayed. Uncertainties in the shapes of the m_{cc} distributions are not shown below, but are taken into account for all backgrounds. CR and SR stand for control region and signal region.

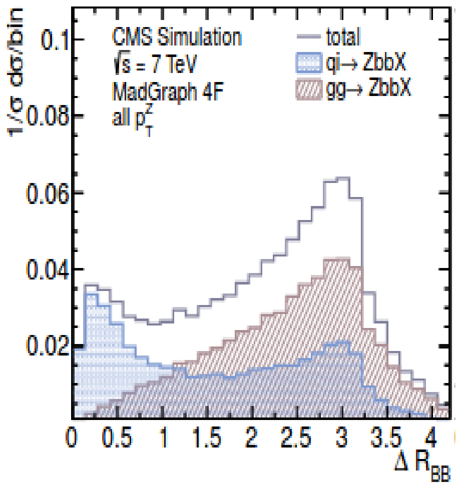
$VH(\to b\bar{b})$	
$WH(\to b\bar{b})$ normalisation	27%
$ZH(\to b\bar{b})$ normalisation	25%
Diboson	
$WW/ZZ/WZ$ acceptance	10%/5%/12%
p_T^V acceptance	4%
N_{jet} acceptance	7%–11%
Z + jets	
$Z+hf$ normalisation	Floating
$Z+mf$ normalisation	Floating
$Z+lf$ normalisation	Floating
$Z+bb$ to $Z+cc$ ratio	20%
$Z+bl$ to $Z+cl$ ratio	18%
$Z+bc$ to $Z+cl$ ratio	6%
p_T^V acceptance	1%–8%
N_{jet} acceptance	10%–37%
High-ΔR CR to SR	12%–37%
0- to 2-lepton ratio	4%–5%
W + jets	
$W+hf$ normalisation	Floating
$W+mf$ normalisation	Floating
$W+lf$ normalisation	Floating
$W+bb$ to $W+cc$ ratio	4%–10%
$W+bl$ to $W+cl$ ratio	31%–32%
$W+bc$ to $W+cl$ ratio	31%–33%
$W\to\tau\nu(+c)$ to $W+cl$ ratio	11%
$W\to\tau\nu(+b)$ to $W+cl$ ratio	27%
$W\to\tau\nu(+l)$ to $W+l$ ratio	8%
N_{jet} acceptance	8%–14%
High-ΔR CR to SR	15%–29%
$W\to\tau\nu$ SR to high-ΔR CR ratio	5%–18%
0- to 1-lepton ratio	1%– 6%

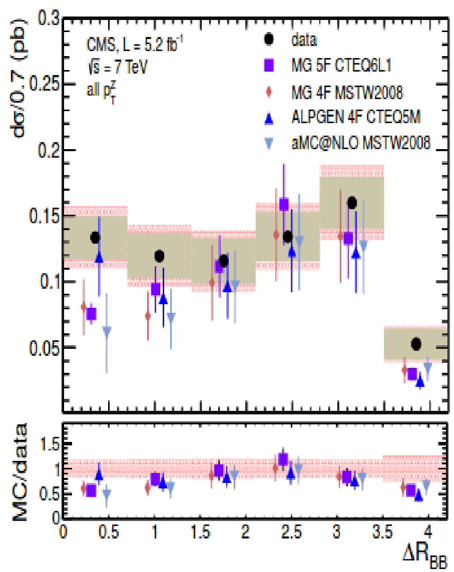


Vector boson + HF jets

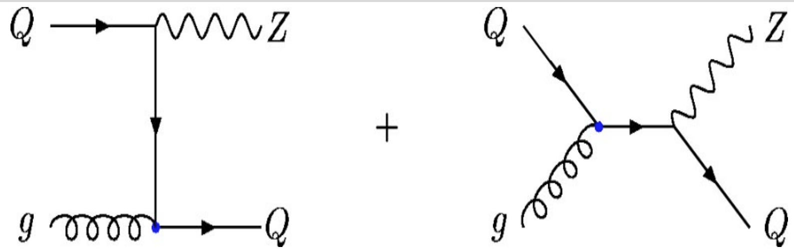
JHEP 12 (2013) 039

- V + 2 HF jets is used to study QCD in more details
  - ΔR_{bb} at small angle dominated by quark initiated process with gluon splitting in the final state
  - $\min(\Delta R_{Zb})$at high jet p_T probes the Z radiated from b-quark
- New input to the determination of the quark content of the proton → parton distribution functions are still an important source of systematic uncertainty at LHC
- Precise measurements of V + HF jets observables
  - improve the theoretical calculation of these processes
  - improve the modelling of V + HF jets events in the currently available MC generators



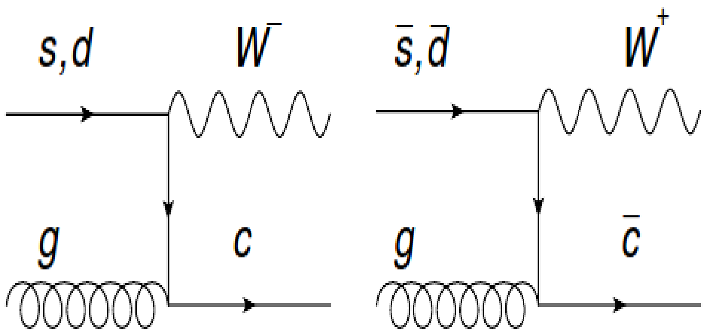


V + HF processes

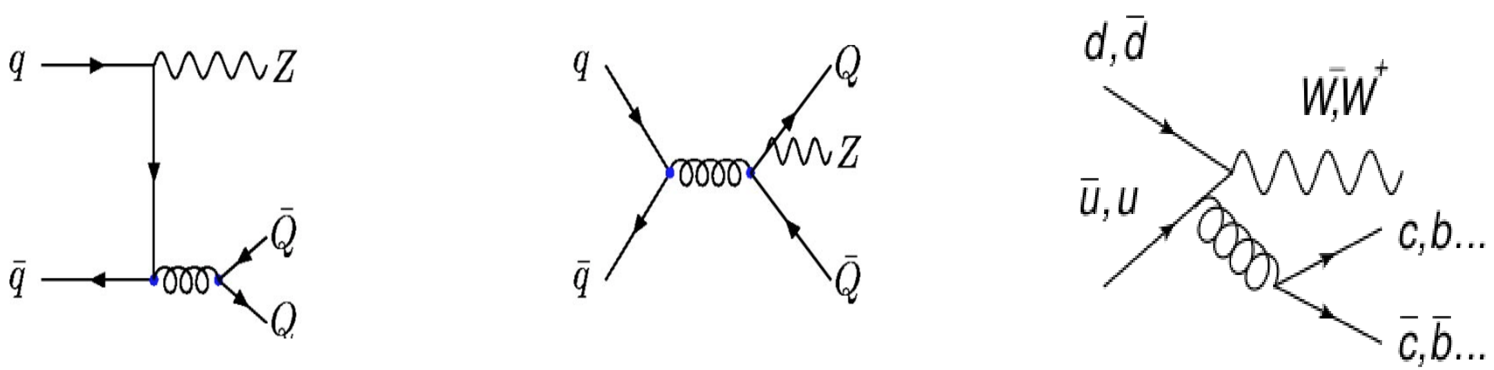


Z+c jet →observing the intrinsic charm quark (IC) component in the nucleon

Z+b → information on the b quark PDFs + helps to estimate the uncertainty coming from the PDF choice in the W boson mass measurement



W + c → direct access to the strange quark content of the proton at the W boson mass energy scale



V+bb, V+cc → probes the gluon splitting

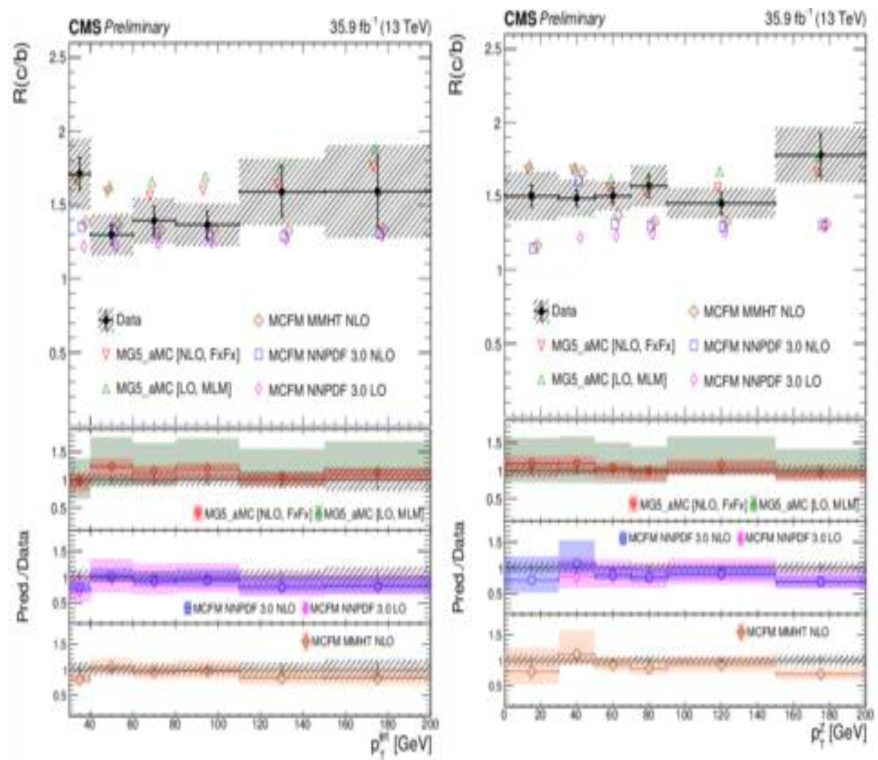
Latest V + HF publications

- W+c → CMS, 13 TeV, 35.9 fb^{-1} Eur. Phys. J. C 79, 269 (2019)
  - inclusive and differential cross section, ratio of W^+ + anti-c and W^- + c cross sections
- W+bb → CMS, 8TeV, 19.8 fb^{-1} Eur. Phys. J. C 77, 92 (2017)
- W+bb and W+cc → LHCb, 8TeV, 1.98 fb^{-1} Phys. Lett. B767 (2017) 110
- W+b and W+c ratios → LHCb, 7TeV+8TeV, 2.0fb^{-1} Phys. Rev. D 92, 052001 (2015)
  - σ(W+b) / σ(W+j), σ(W+c) / σ(W+j), σ(W^++j) / σ(Z+j), σ(W^-+j) / σ(Z+j)
- Z+c → CMS, 13 TeV, 35.9 fb^{-1} J. High Energ. Phys. 2021, 109 (2021)
- Z+b → CMS, 13 TeV, 137 fb^{-1} Phys. Rev. D 105, 092014 (2022)
  - Z+ ≥1b, Z+ ≥2b, ratio of Z + ≥2b and Z + ≥1b
- Z+c and Z+b ratios → CMS, 13 TeV, 35.9 fb^{-1} Phys. Rev. D 102, 032007 (2020)
  - σ(Z+c jets) / σ(Z+jets), σ (Z+b jets) / σ(Z+jets), σ(Z+c jets) / σ(Z+b jets)

→ **Z+cc** cross section and the ratio of **Z+cc / Z+bb** was never measured

Cross section ratio measurements

- By measuring cross section ratios, one can more precisely compare data with theoretical calculations
  - cancellations of several systematic uncertainties related to the jet, lepton, and luminosity measurements
  - a number of theory-related uncertainties are reduced as well, including those linked to the details of parton showering and hadronization



[Phys. Rev. D 102, 032007 (2020)]

Data and simulated samples

- pp collisions @ 13TeV, collected by the CMS experiment during Run2
- SingleElectron and SingleMuon datasets
- Integrated luminosity:
  - 2016 → 36.3 fb^{-1}
  - **2017 → 41.5 fb^{-1}**
  - 2018 → 59.8 fb^{-1}
  - Full Run2 → 137.6 fb^{-1}

Signal MC

- Drell-Yan (DY) with jets
  - samples with exclusive jet multiplicities (0, 1 or 2 jets)
  - sample generated with m_Z>50GeV
- → contain signal (Z+b, Z+c) and background (Z+light jet) events

Background MC

- W + jets
- tt
- Single top samples
- Diboson processes (WW, WZ, ZZ)
- ZH, H→bb or cc

Monte Carlo Samples

Sample	Number of Events	σ(×BR) (pb)
MC_2017-UL_DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	196 329 377	6404.0
MC_2017-UL_TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	249 133 364	750.5
MC_2017-UL_WW_TuneCP5_13TeV-pythia8	15 634 000	75.95
MC_2017-UL_WZ_TuneCP5_13TeV-pythia8	7 889 000	27.59
MC_2017-UL_ZZ_TuneCP5_13TeV-pythia8	2 706 000	12.17
MC_2017-UL_ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	13 882 000	3.549
MC_2017-UL_ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	69 921 000	69.09
MC_2017-UL_ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	129 903 000	115.3
MC_2017-UL_ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	5 674 000	34.97
MC_2017-UL_ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	5 649 000	34.91
MC_2017-UL_ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8	2 916 024	0.05
MC_2017-UL_ZH_HToCC_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8	2 848 348	0.05

Synchronisation with previous measurement

We compared the number of muon objects with the previous measurement after selections

- Using single muon run 2017 data
- Kinematic cut on P_T and η
- Id cut on level of muon object reconstruction
- Relative isolation from jets

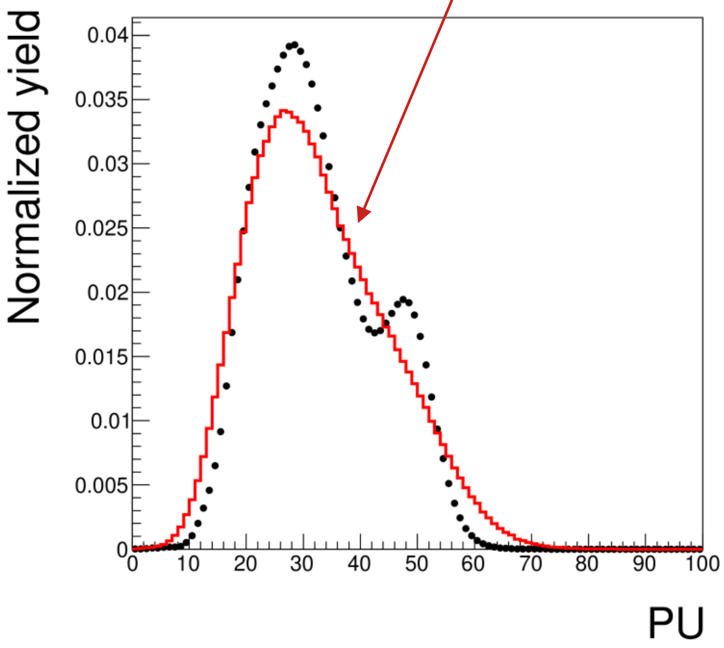
	Events	All Muons	Kinematics Cut	Id Cut	Rel. Isolation Cut
Zbb Analysis	**769 080 720**	**883 128 000**	**619 407 000 (70.1%)**	**598 434 000 (96.6%)**	**469 600 000 (78.5%)**
This Analysis	**751 863 031**	**1 018 982 218**	**707 671 410 (69.4%)**	**619 377 005 (84.6%)**	**483 898 685 (78.1%)**

Discrepancies in the numbers are due to different data reconstructions

Pileup re-weighting

- Single bunch-crossing contains multiple pp collisions beyond the triggered hard-scattering (pileup)
- Jets, missing energy, and other objects are corrected for the extra particles
- MC events need to be re-weighted to match data
- Estimated effect of associate systematic uncertainty is around ***2-3% after correction***
  - *Still to be done in our analysis...*

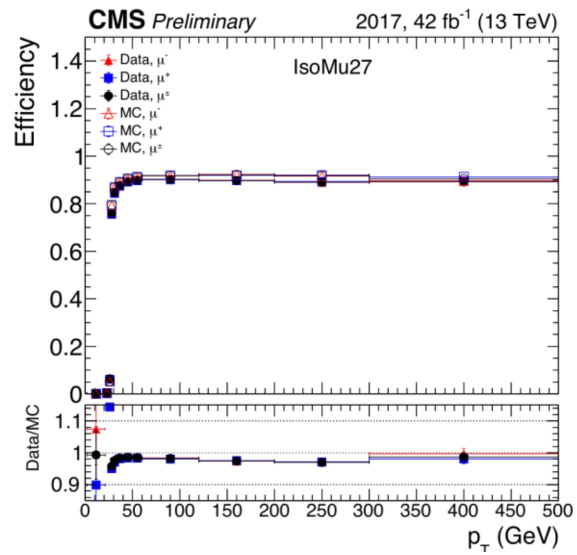
PU is simulated by mixing a number of minimum bias interactions into main event

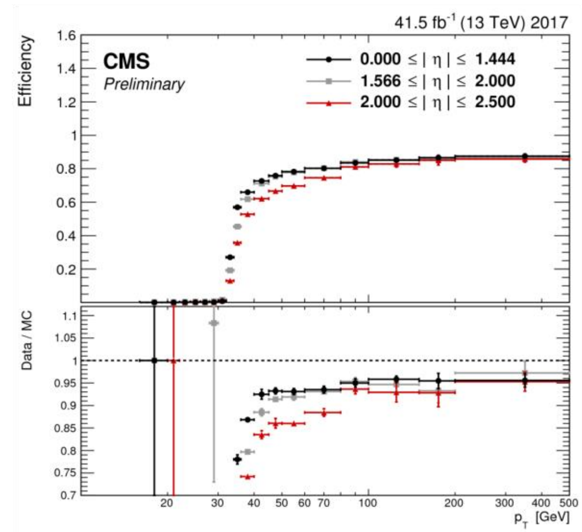


Expected PU is computed from bunch-by-bunch luminosity

Trigger

- The pp interaction rate is ~1 GHz of which about ~1 kHz is feasible to record
- HLT_IsoMu27 used for muons, selecting for a single isolated muon with pT>27 GeV
- HLT_Ele32_WPTight_Gsf_L1DoubleEG_v used for electrons selecting for ET>32 GeV electrons
- Trigger selection efficiency is generally highest and flat beyond the offline selection of leading leptons
- Trigger is emulated in MC, but different efficiency needs to be corrected for by MC scaling
  - No such correction is applied so far, introducing ~0-3% **pT and eta dependent** discrepancy in muons
  - For two muons this becomes ~0-0.7%
  - And ~5-10% in electrons, especially significant at low ET
  - For two electrons this becomes up to 4%



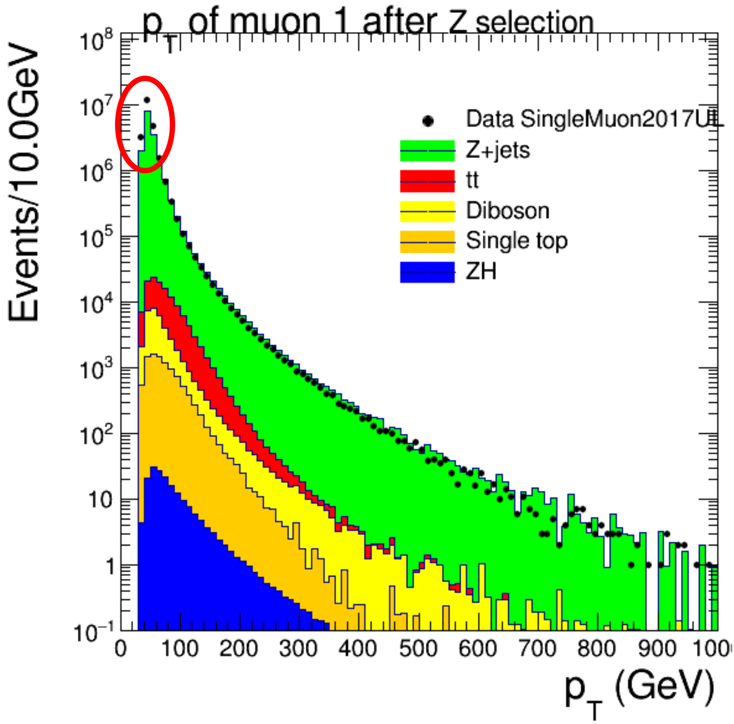


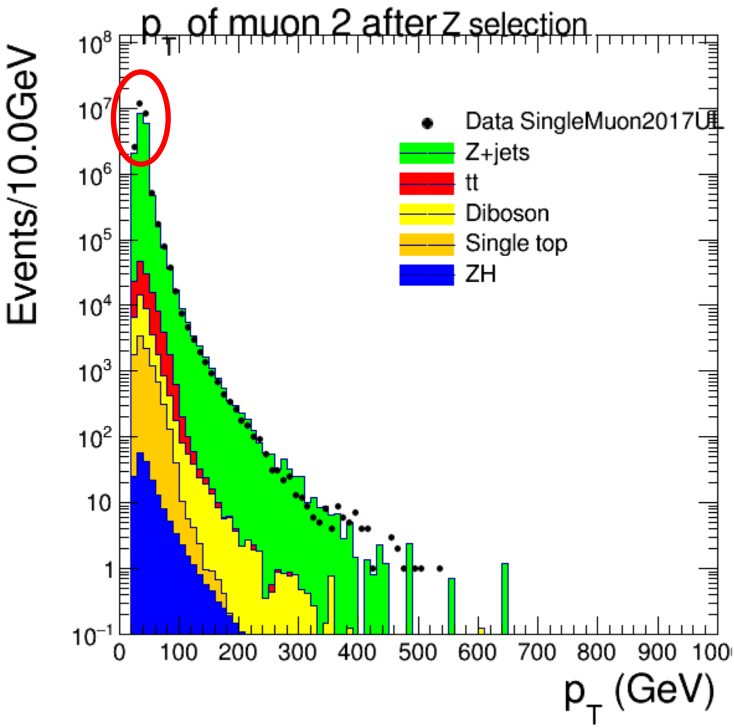
Object Definitions and Control Plots

Muon Definition

Muon selection:

- **Kinematics**: P_T > 35 GeV (lead), 25 GeV (sub-lead) ; |η| < 2.4
- **Muon ID:** Medium ID: Loose muon with track-quality and muon-quality requirements
- **Relative Isolation:** In require < 0.2

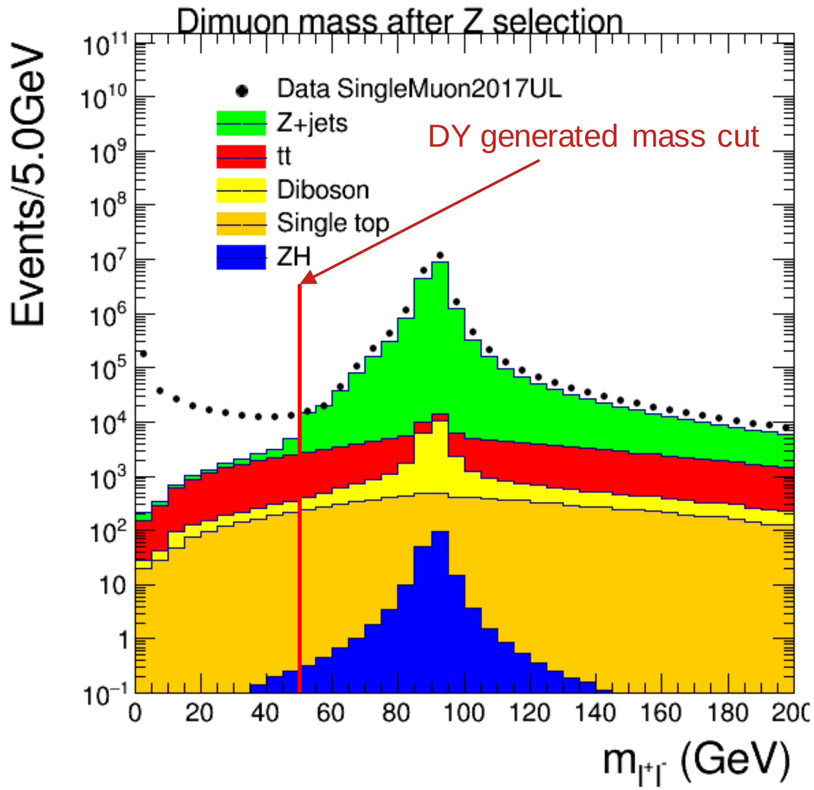


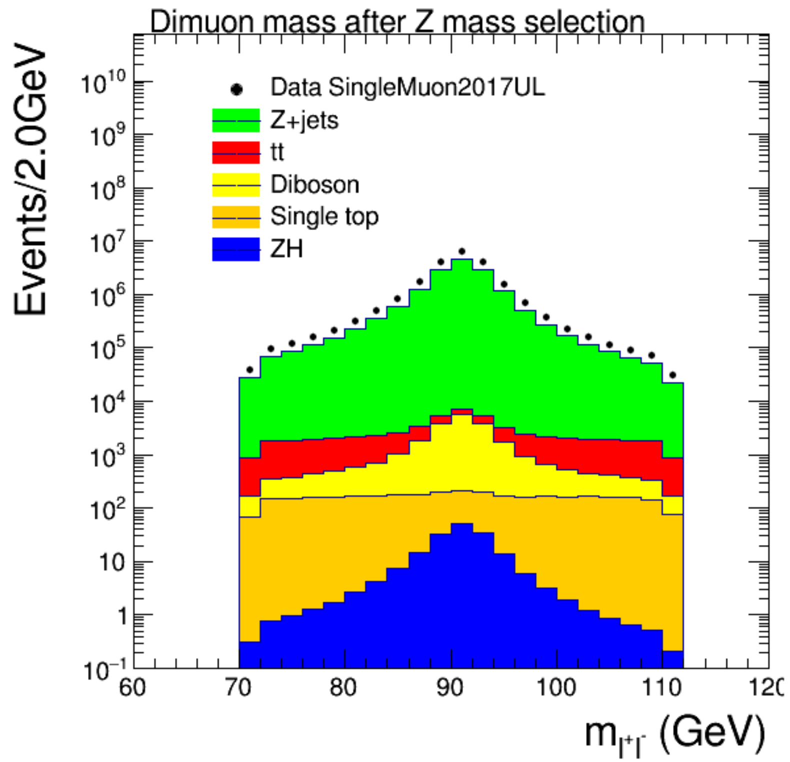


Dimuon (Z) mass

Z ($\mu^+ \mu^-$) candidates: Opposite charge, Leading and Sub-Leading muons

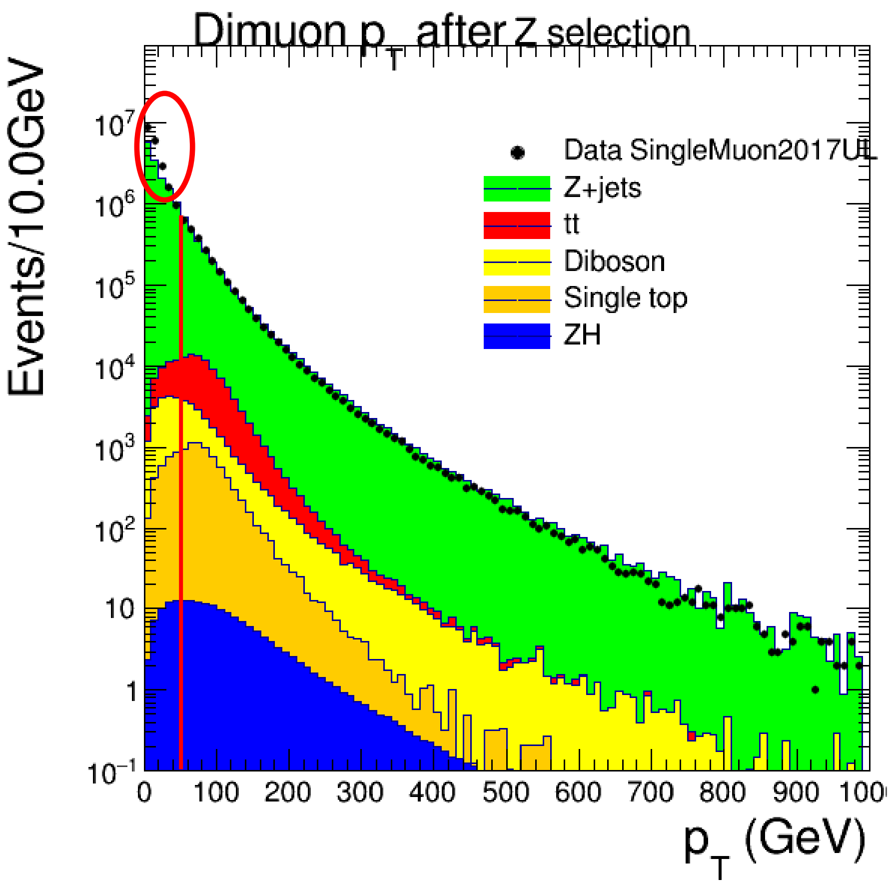
- Considered reconstructed mass cut: 71 GeV << 111 GeV
- Drell-Yan (Z/γ + jets) process generation: > 50 GeV
- Cut on generated mass → possible explanation of discrepancies at lower P_T

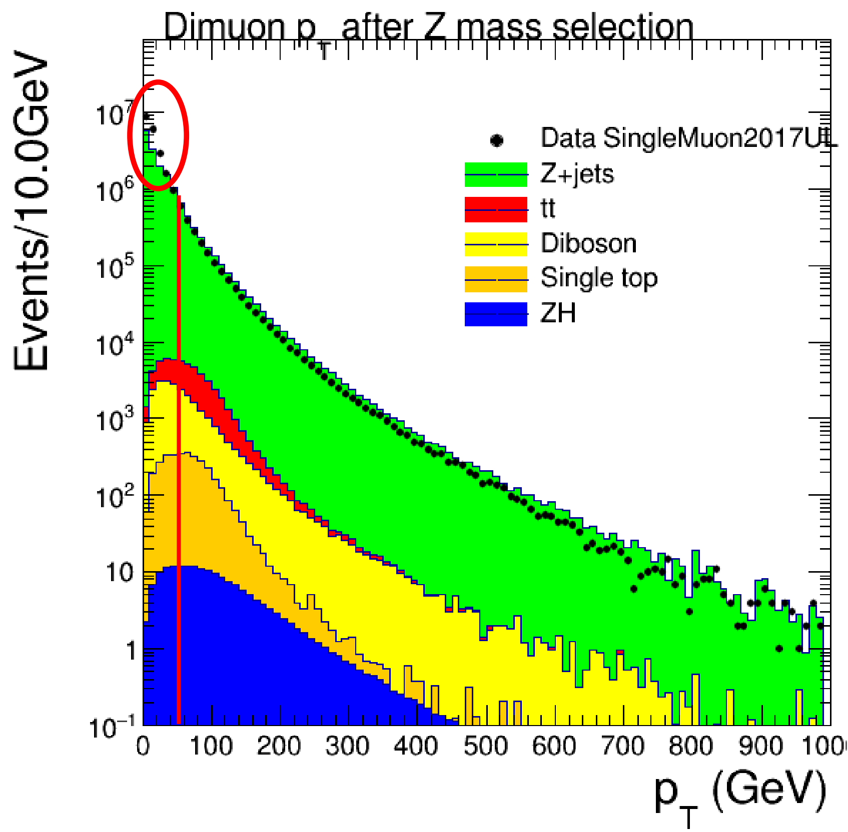




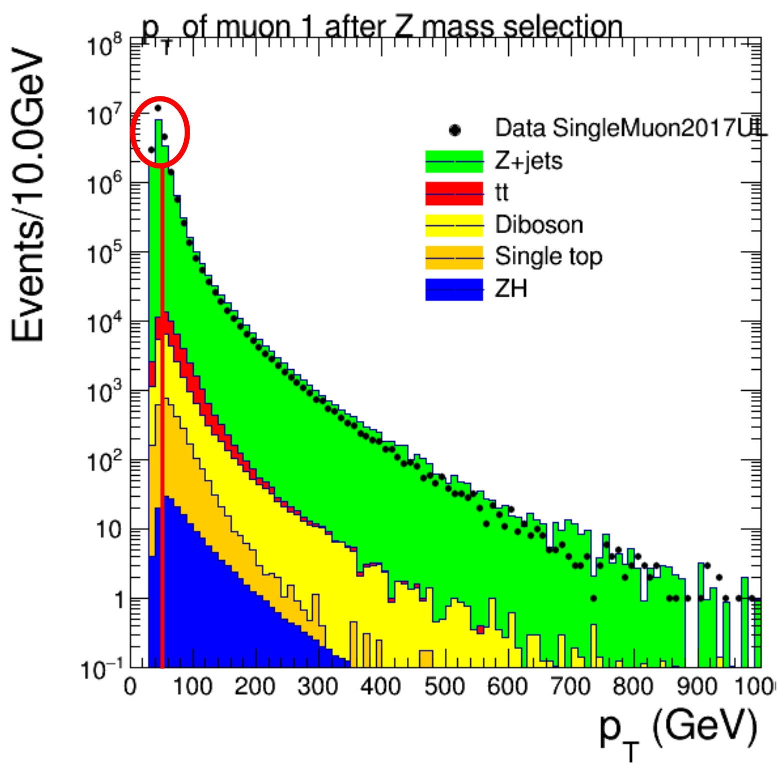
Dimuon (Z) transverse momentum

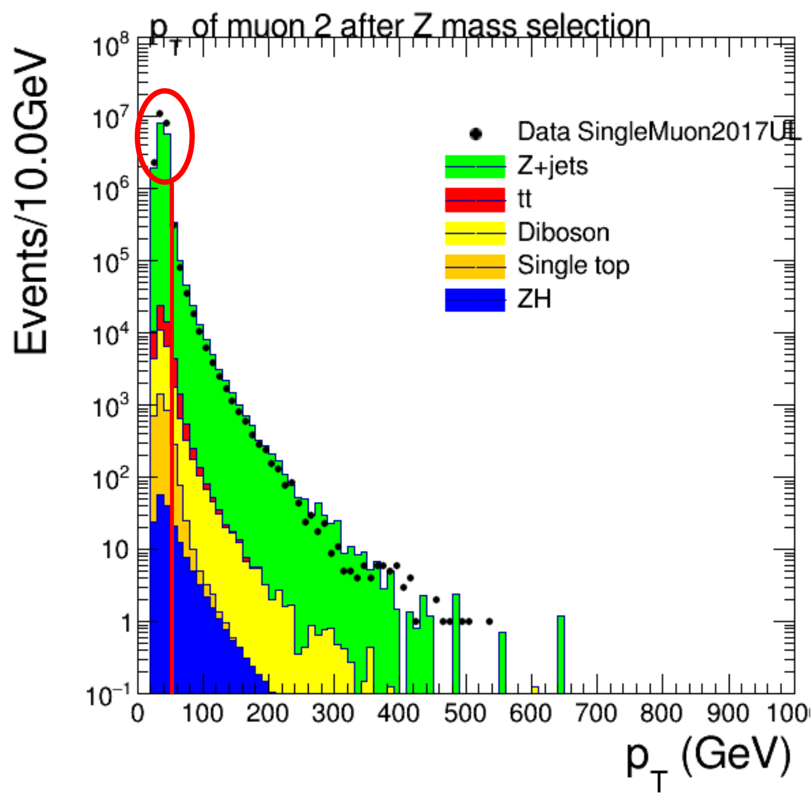
- Data and simulation are in agreement in several magnitudes
- Discrepancy at lower momenta: cut in generated < 50 GeV
- Must examine simulation with lower as well





Muon P_T After Cut on Dimuon Mass



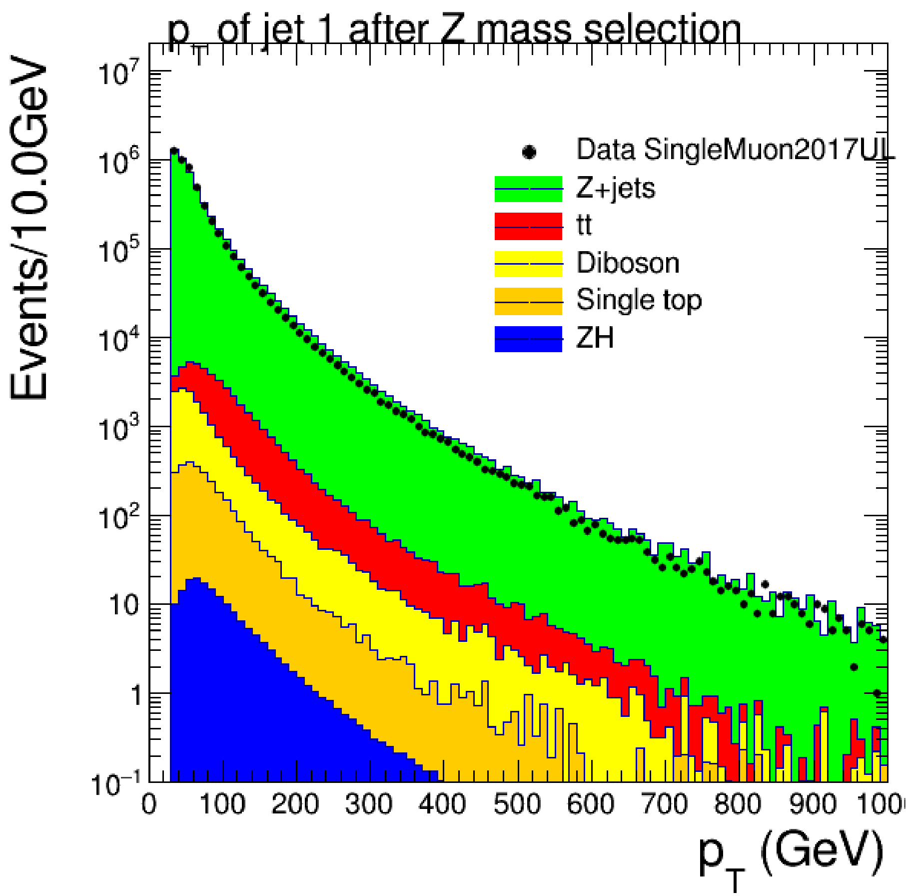


Jet Definition

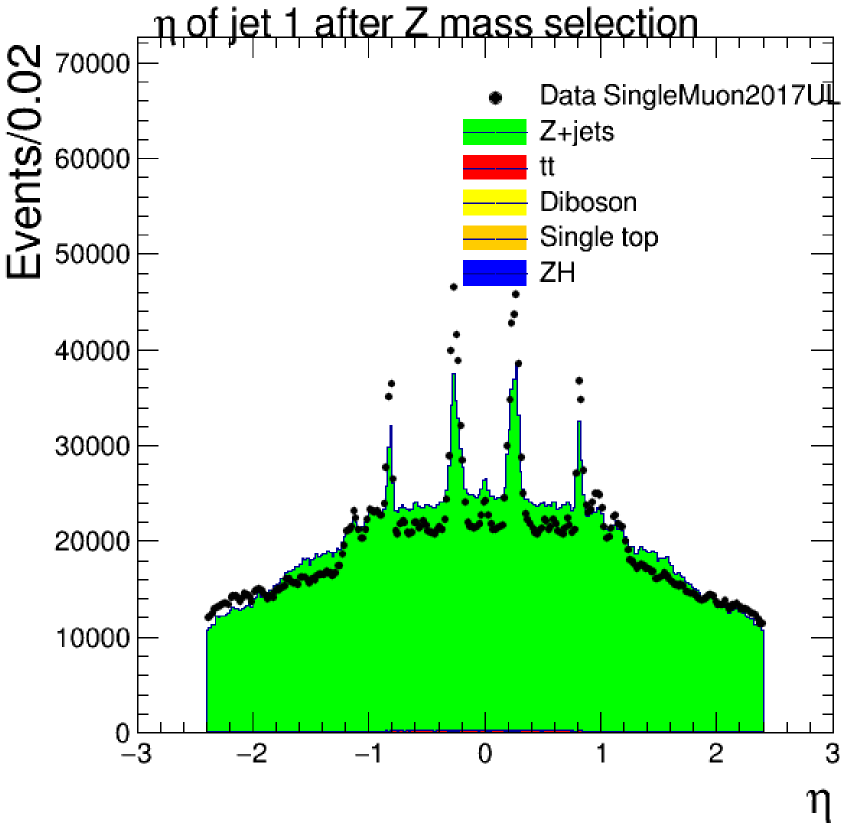
- **Kinematics selection**: $P_T > 30$ GeV; $|\eta| < 2.4$
- **Lepton Cross Cleaning:** Discard if overlap:
  Tight ID, $P_T > 25$ GeV muon or electron in
- **ID selection:** Tight ID
- **Pileup ID selection:** Jets from PU identified by

ID discriminator > 0.61 (if $30 < P_T \leq 50$ GeV)

Tight Jet ID Definition	
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

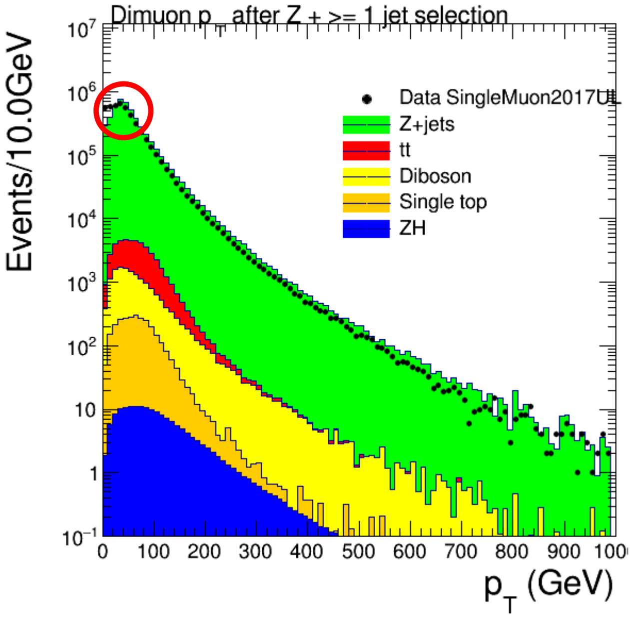


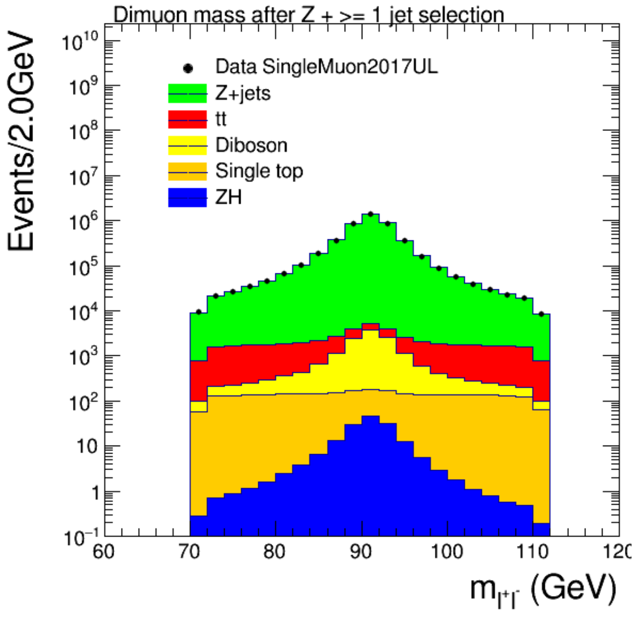
Jet η After Cut on Dimuon Mass



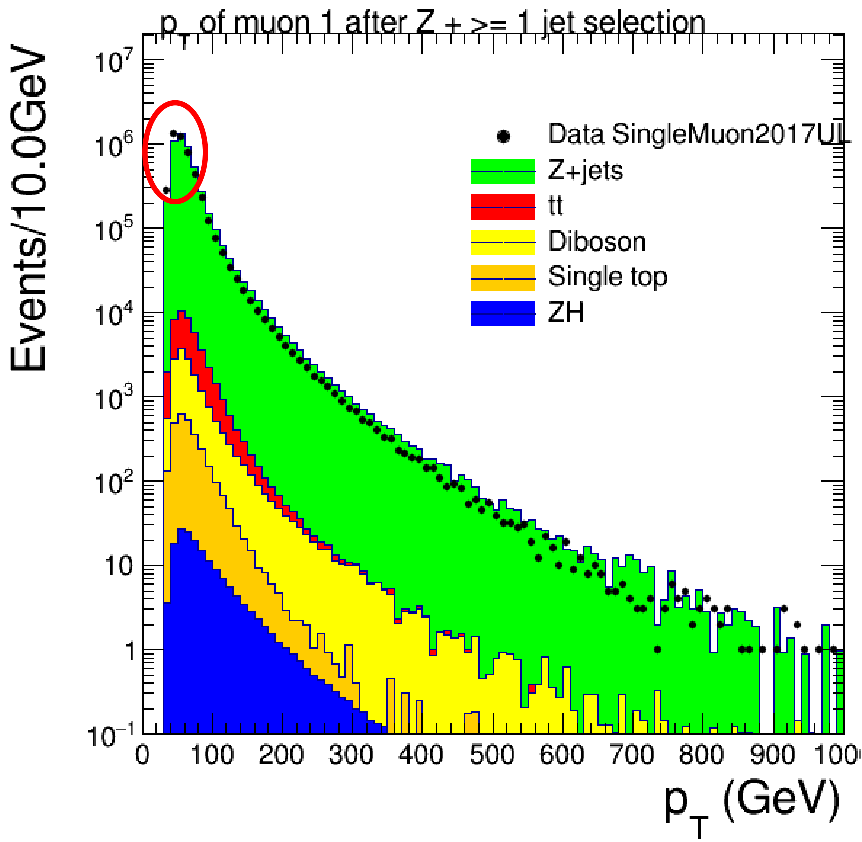
Dimuon P_T and Mass After >= 1 Jet Selection

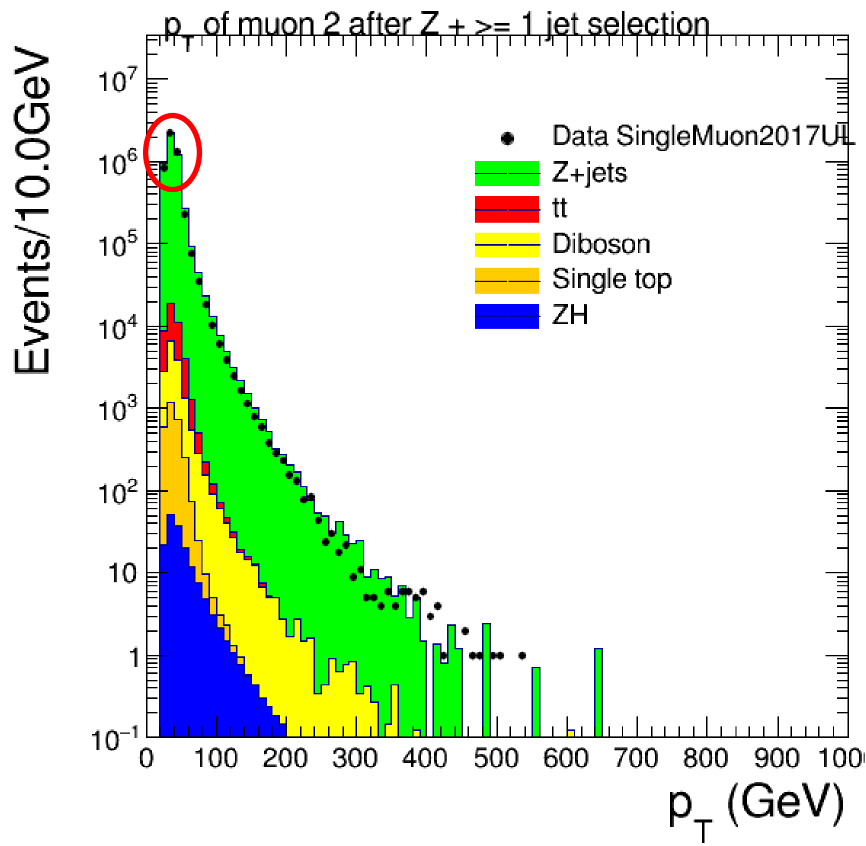
	Previous Analysis	This Analysis
DY	4 085 659	4 521 270
tt	47 497	29 548
Single t	4 376	2 678
VV	25 340	12 959
ZH	230	165
MC	4 163 103	4 566 620
Data	4 362 308	4 742 969
Data/ MC	**1.05**	**1.04**



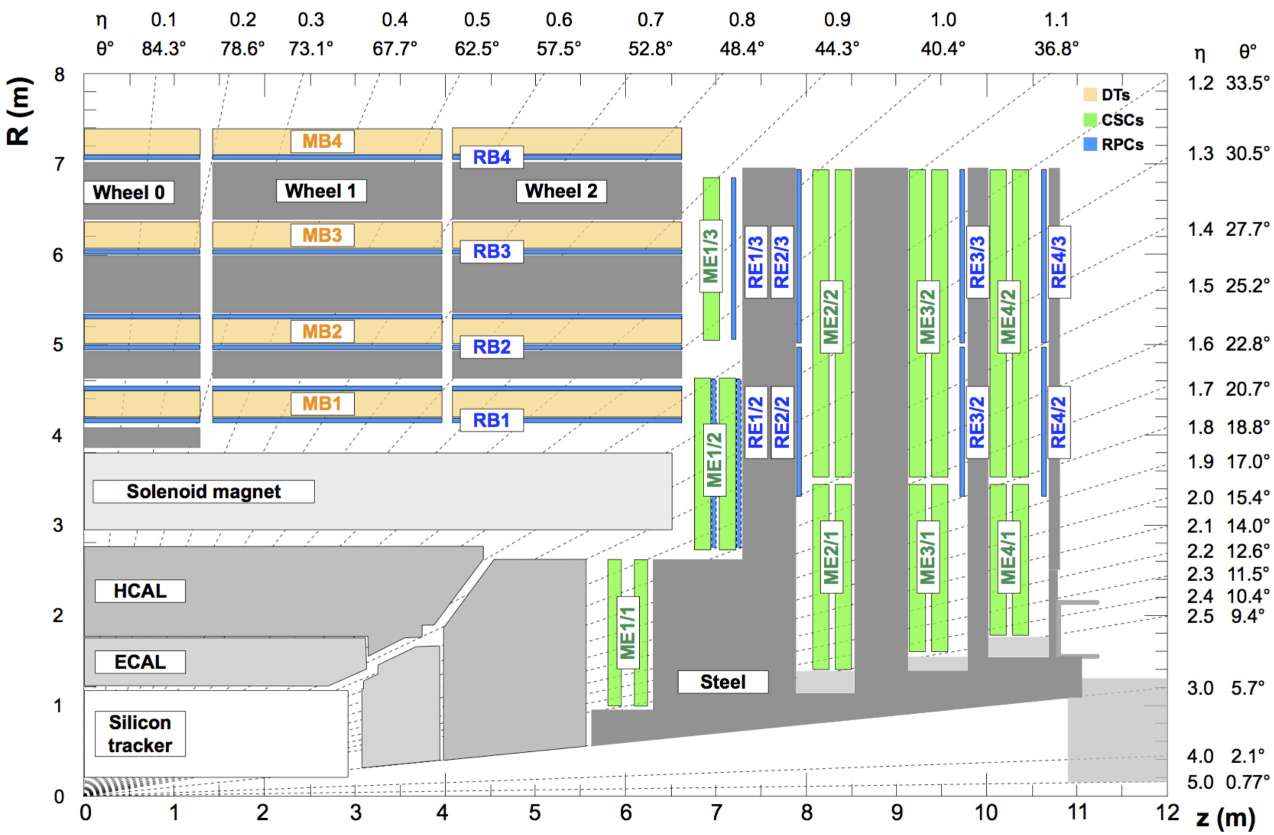


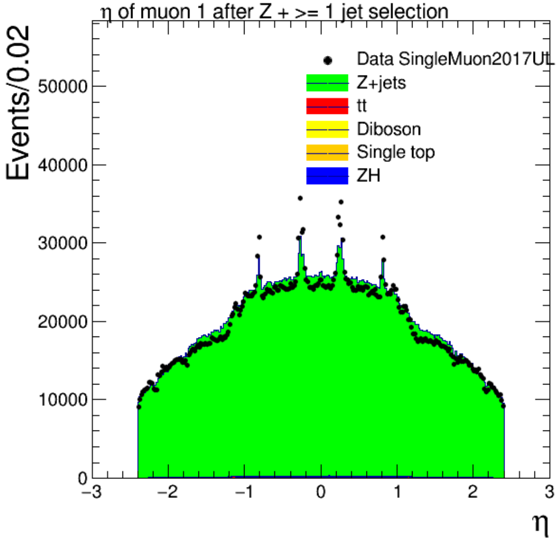
Muon P_T After >= 1 Jet Selection

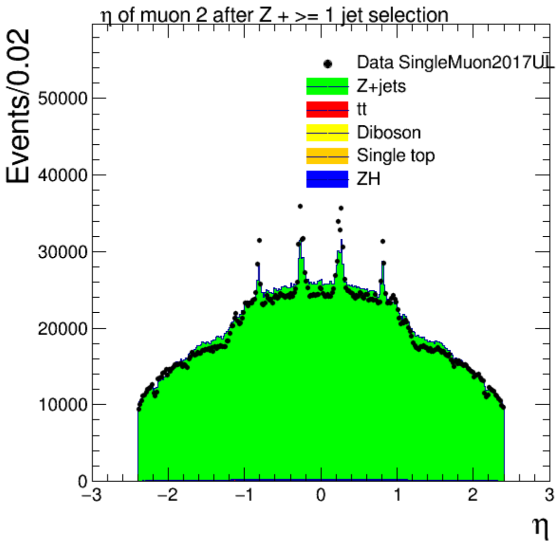




Muon η After >= 1 Jet Selection



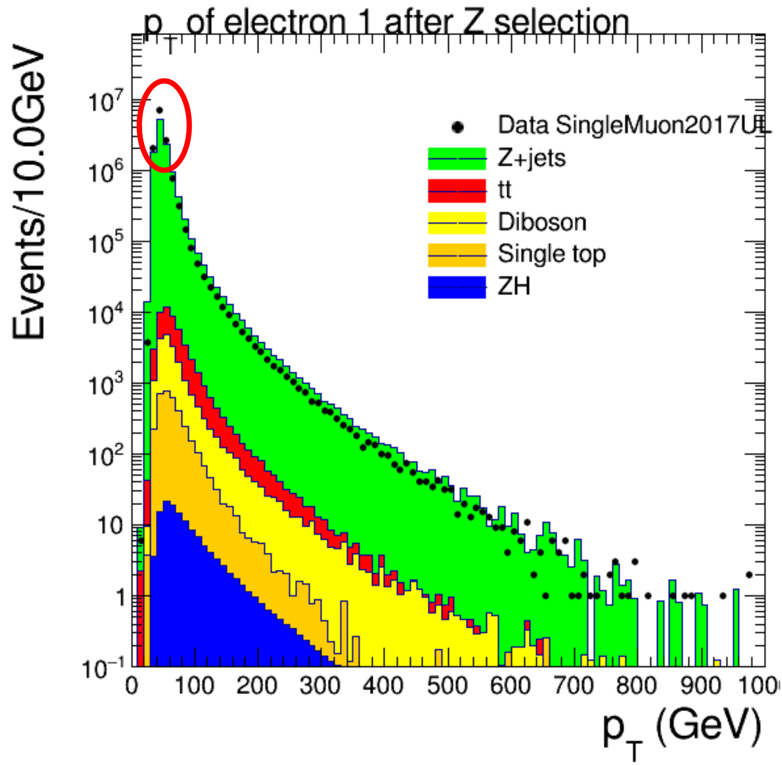


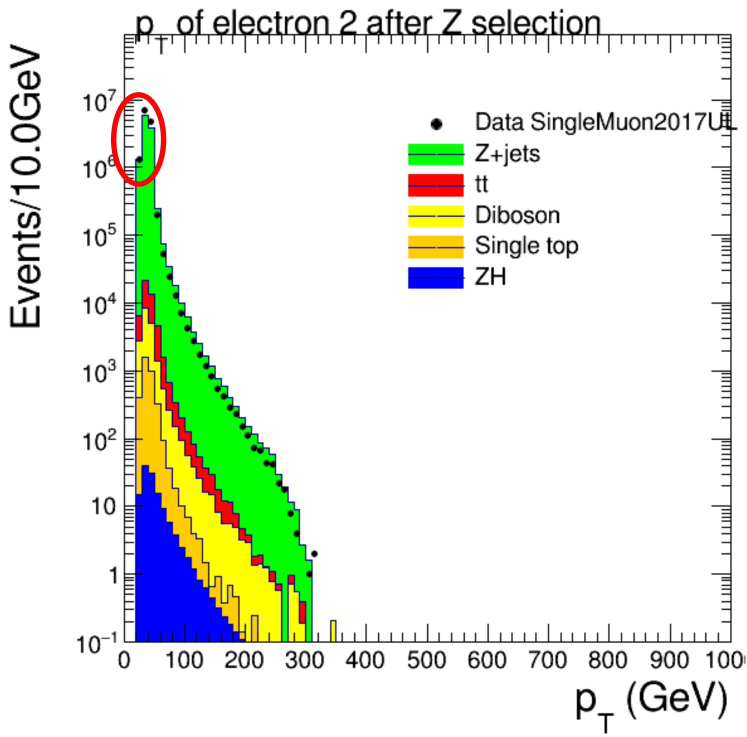


Electron Definition

Electron selection:

- **Kinematics**: $P_T > 25$ GeV (sub-lead)
- **Electron ID:** Tight ID **(**94X-V2) corresponding to ~70% ID efficiency
- **Impact parameter:**
  - If $|\eta| < 1.48$: $|d_z| <0.1$ cm and $|d_0| < 0.05$ cm (barrel region)
  - If $1.48 < |\eta| < 3.0$: $|d_z| <0.2$ cm and $|d_0| < 0.1$ cm (endcap region)

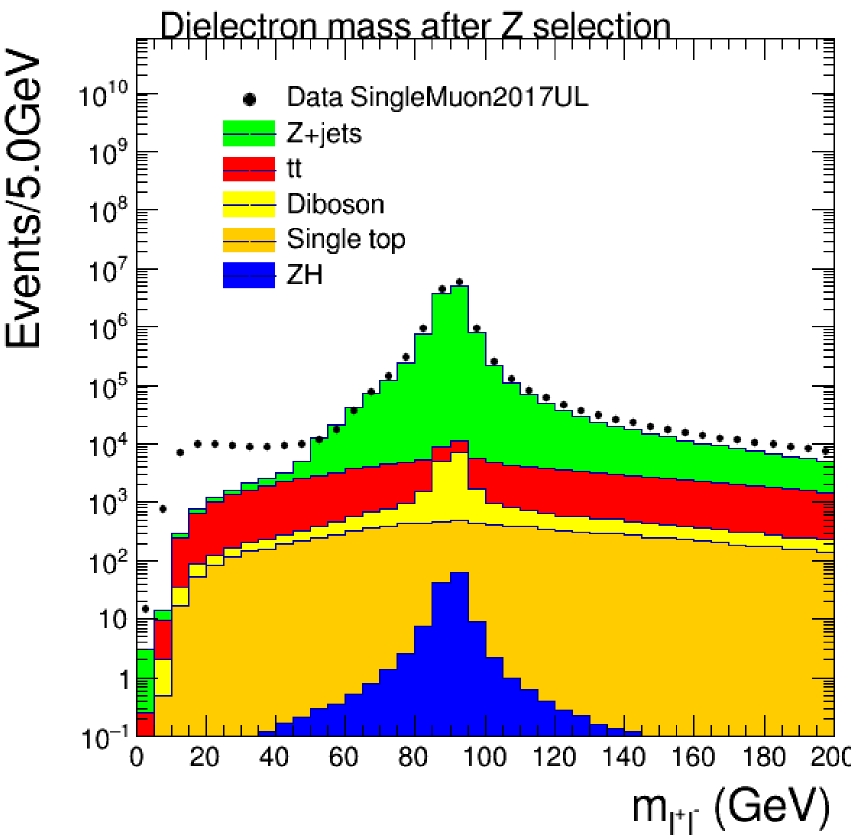


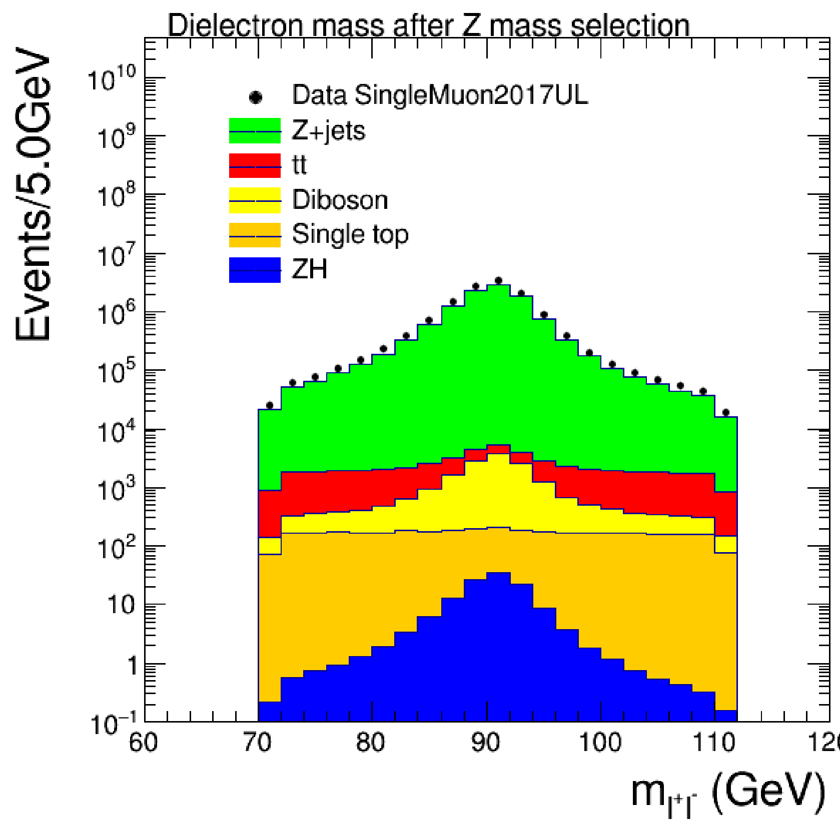


Dielectron (Z) mass

Z (e^+e^-) candidates: Opposite charge, Leading and Sub-Leading electrons

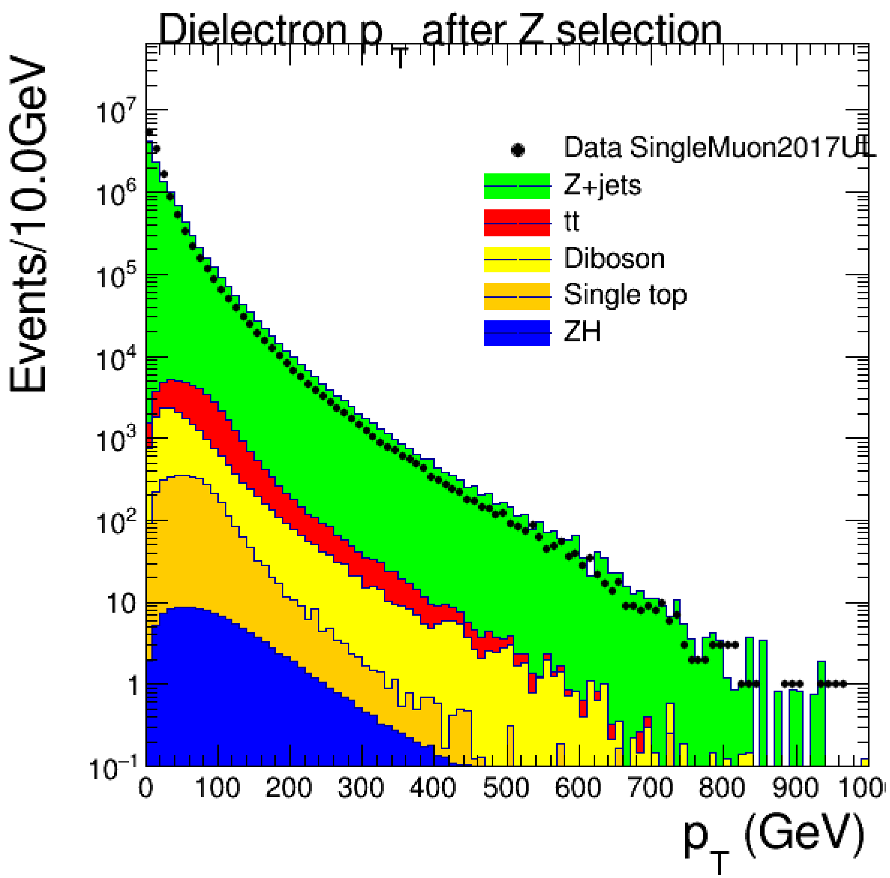
- Considered reconstructed mass cut: 71 GeV << 111 GeV
- Drell-Yan (Z/γ + jets) process generation: > 50 GeV



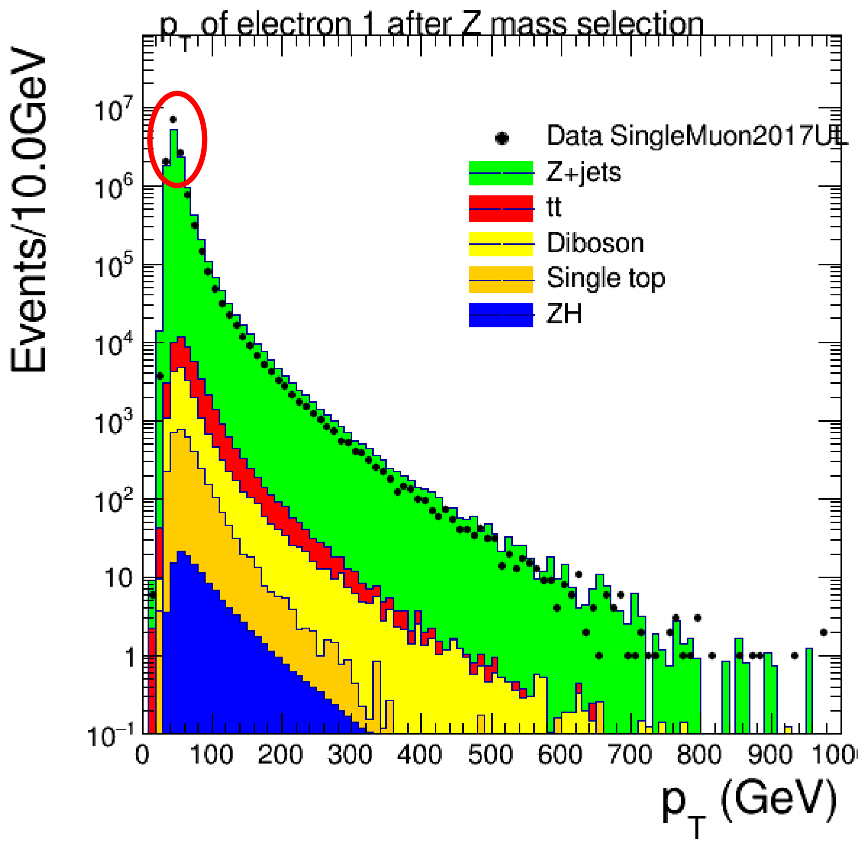


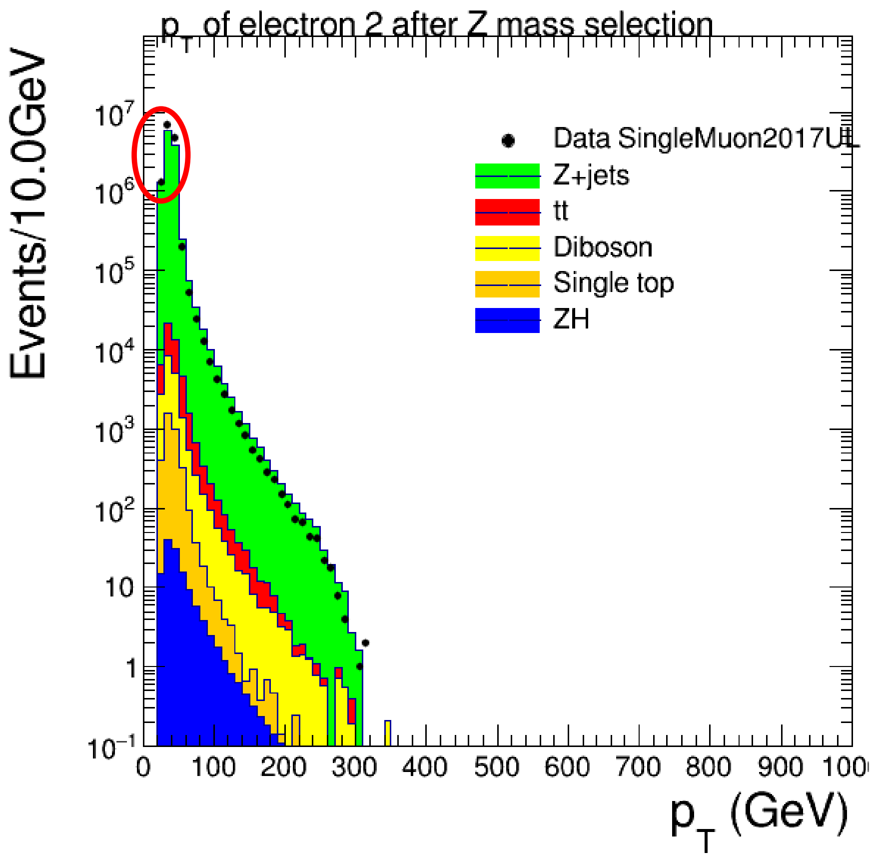
Dielectron (Z) transverse momentum

- Data and simulation are in agreement in several magnitudes
- Higher number of data at lower momenta: cut in generated < 50 GeV
- Must examine simulation with lower as well

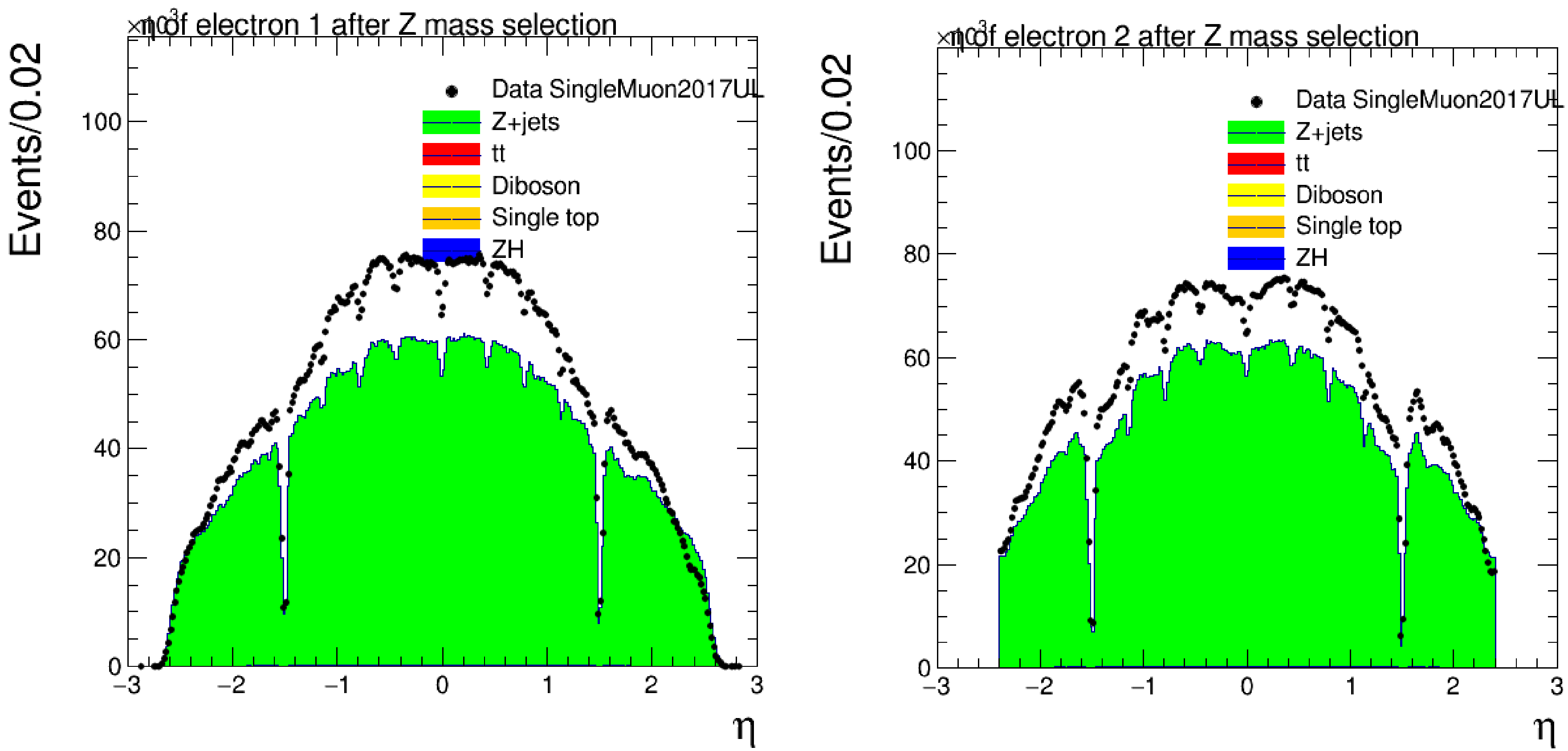


Electron P_T After Cut on Dimuon Mass

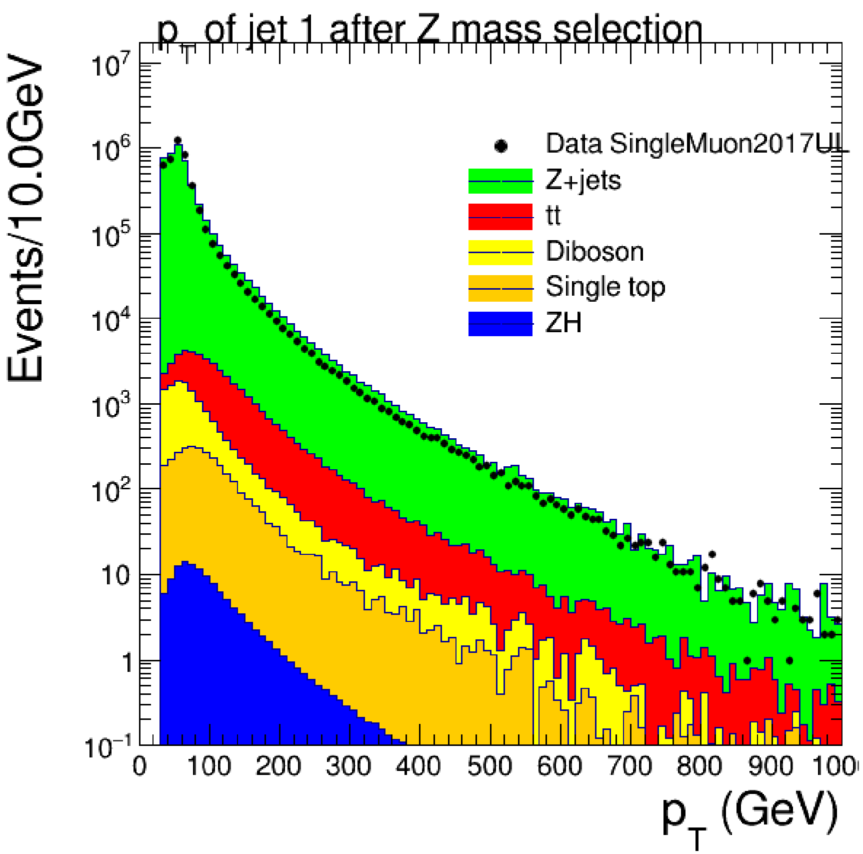


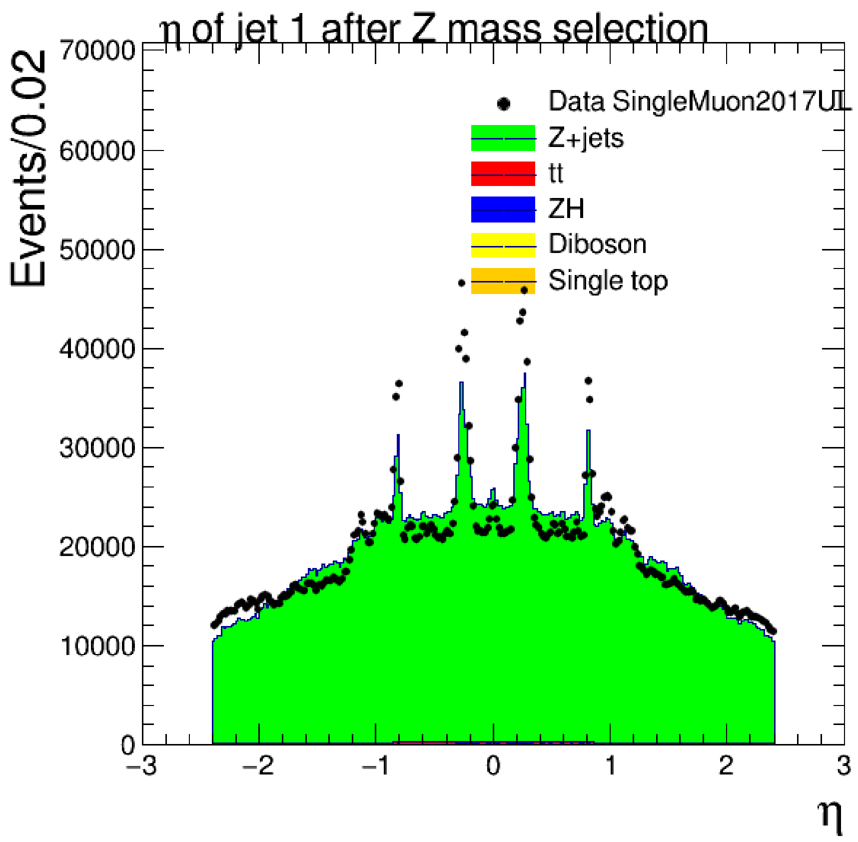


Electron η After Cut on Dimuon Mass

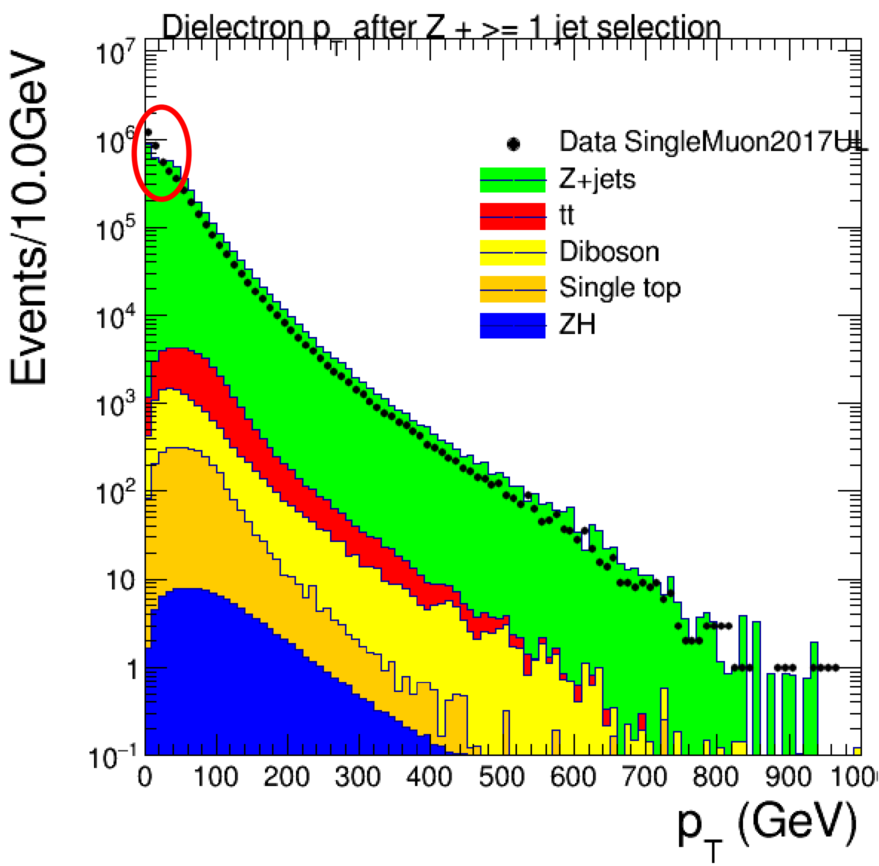


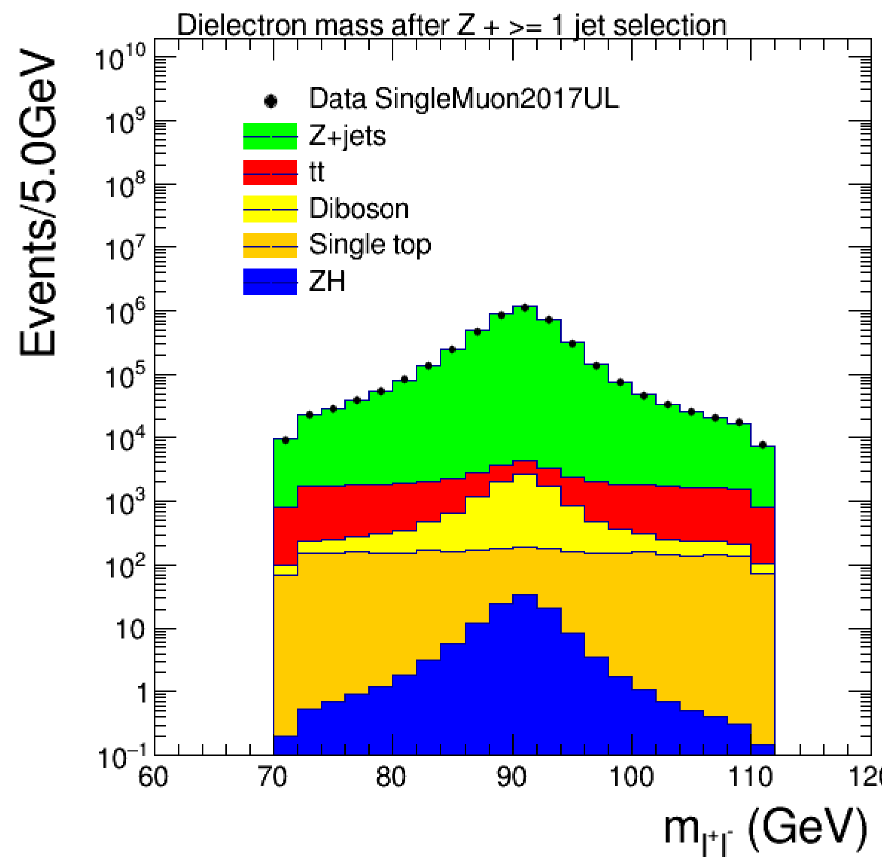
Jet P_T and η After Cut on Dimuon Mass



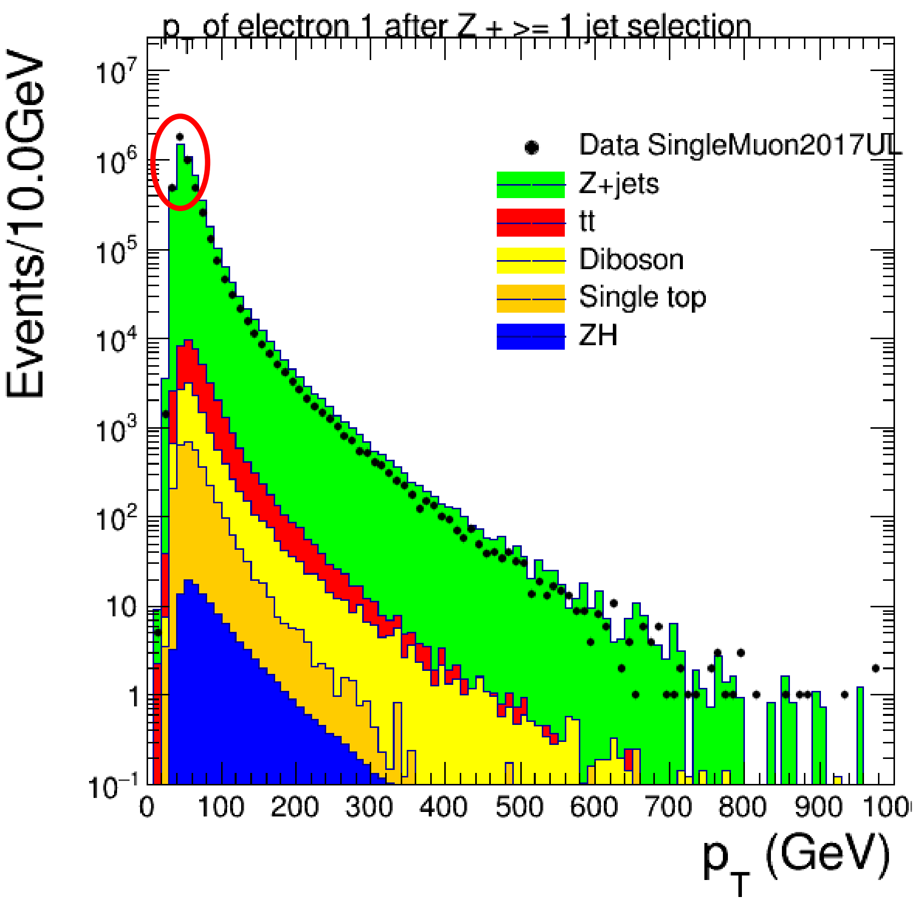


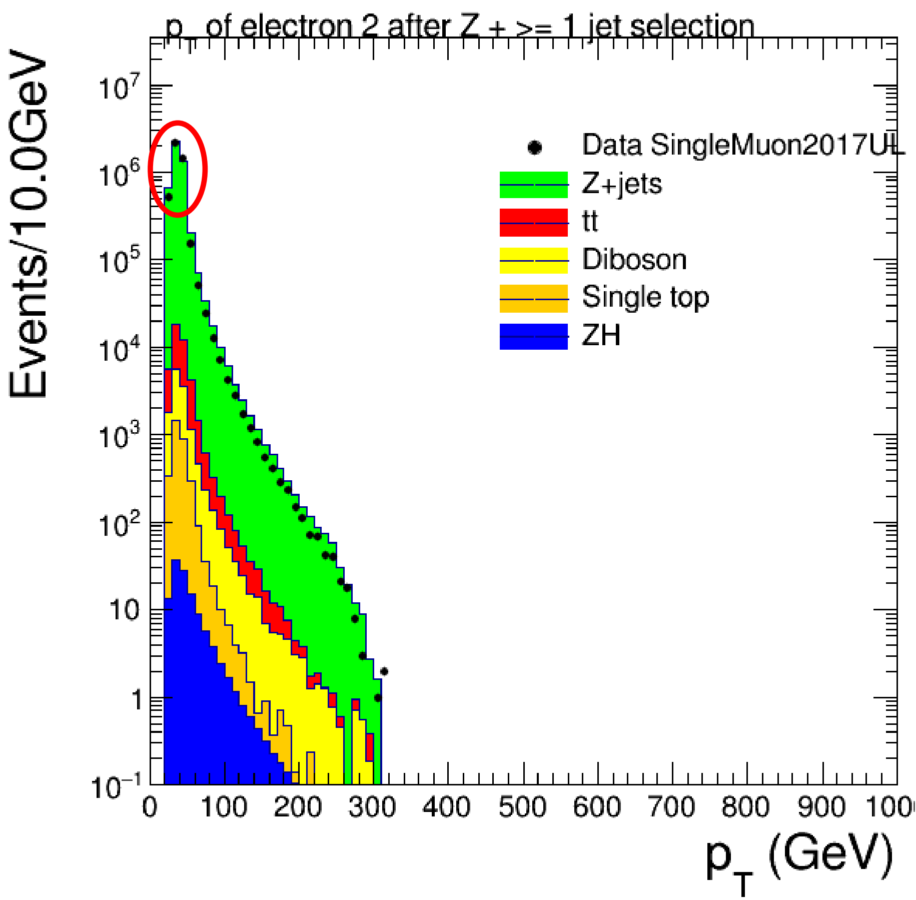
Dielectron P_T and Mass After >= 1 Jet Selection



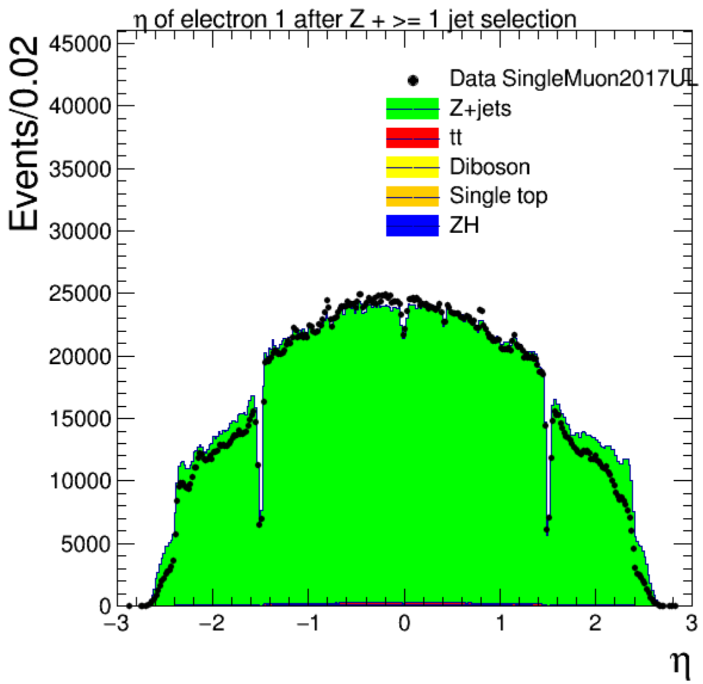


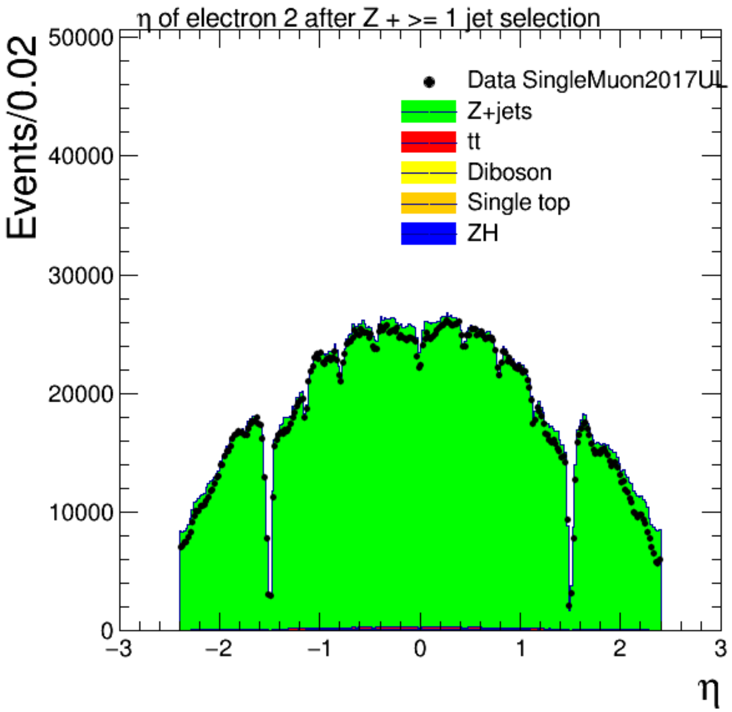
Electron P_T After >= 1 Jet Selection





Electron η After >= 1 Jet Selection





Summary

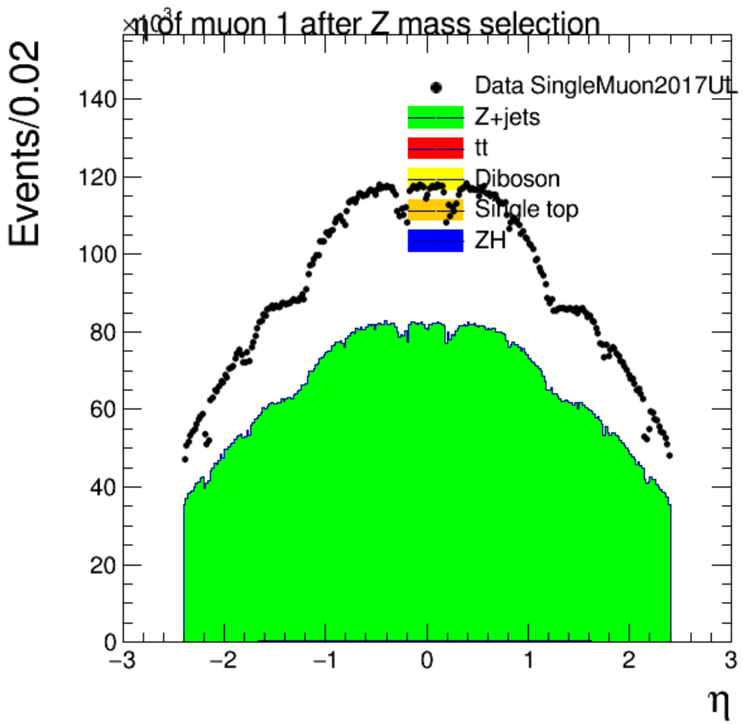
- The importance of the V + h.f. analyses have been outlined
- Started out addressing so far unpublished measurements of Z + h.f. processes by synchronizing the latest Z + b cross-section measurement
- Using the final re-reconstruction of the Run 2 data, we have established the basic selections and achieved good agreement → first milestone reached
- Few checks need to be performed to achieve better agreement with Monte Carlo simulation (trigger and PU scaling), but we have set up all tools

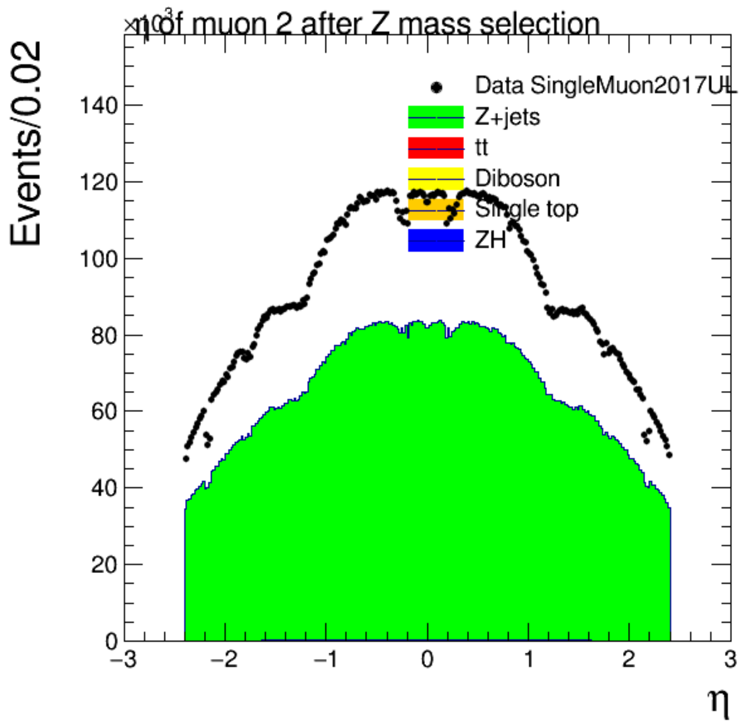
Next steps

- Define control regions to better control main background normalization and work out background subtraction
- Re-measure Zbb cross sections, differential cross-sections as function of Z and jet momenta and relative angles
- Study details of Z-production modes
- Compare performance of available MC generators
- Repeat the same measurements with c-tagging
- Work out analysis for Z+cc/Z+bb ratio and address details of gluon splitting in the context of V + h.f.

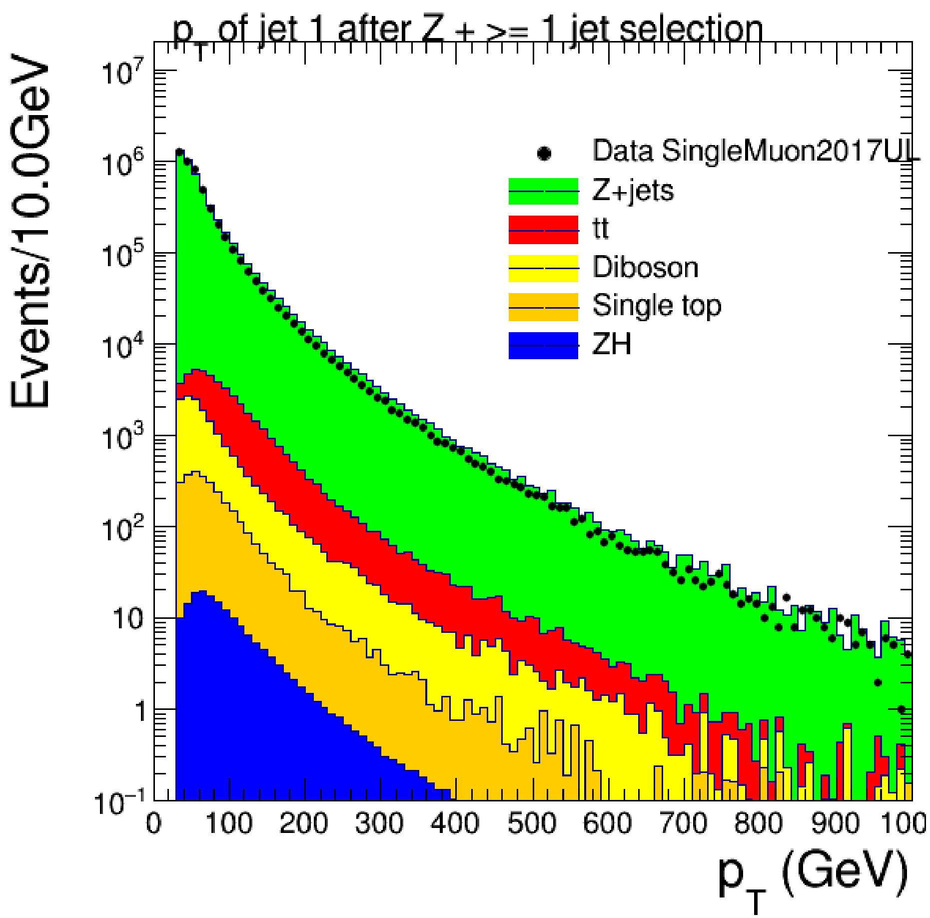
Backup

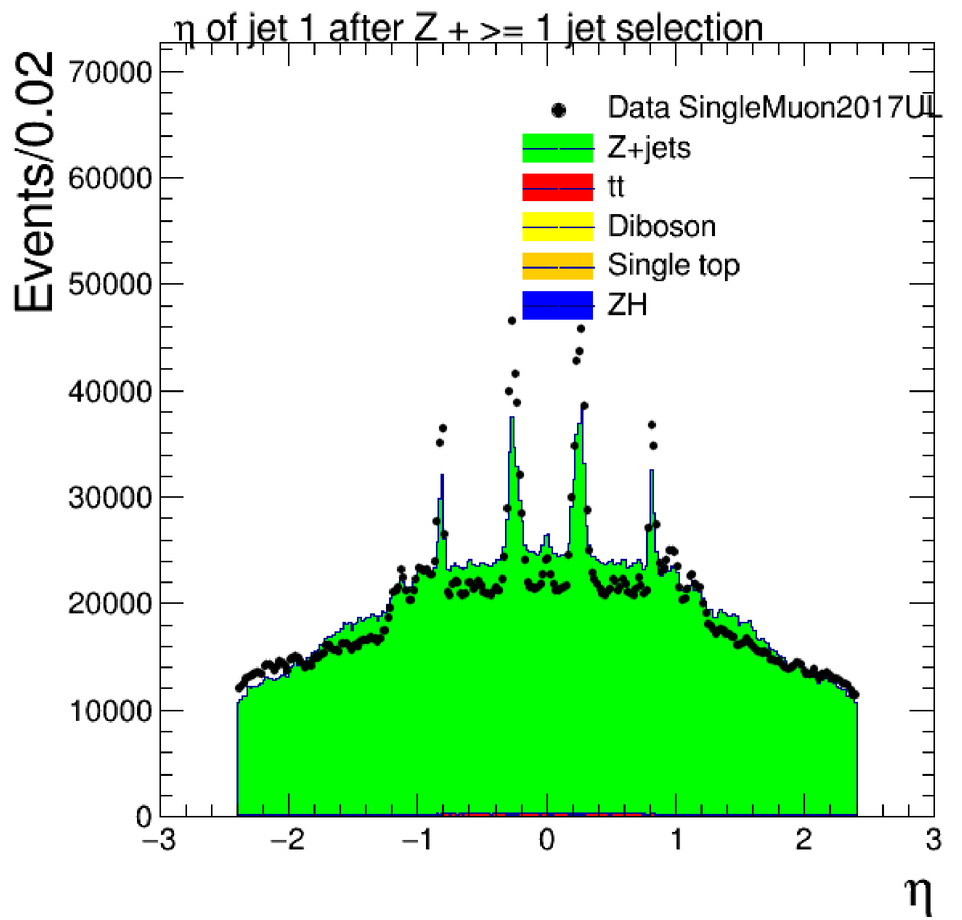
Muon η After Cut on Dimuon Mass





Jet P_T and η After >= 1 Jet Selection (Muon)





Jet P_T and η After >= 1 Jet Selection (Electron)