Institute of High Energy Physics
Chinese Academy of Sciences

Evidence for vector boson scattering in semileptonic lνqq final states in proton-proton collisions at $\sqrt{s}$= 13 TeV with CMS

Ram Krishna Sharma
On Behalf of CMS Collaboration

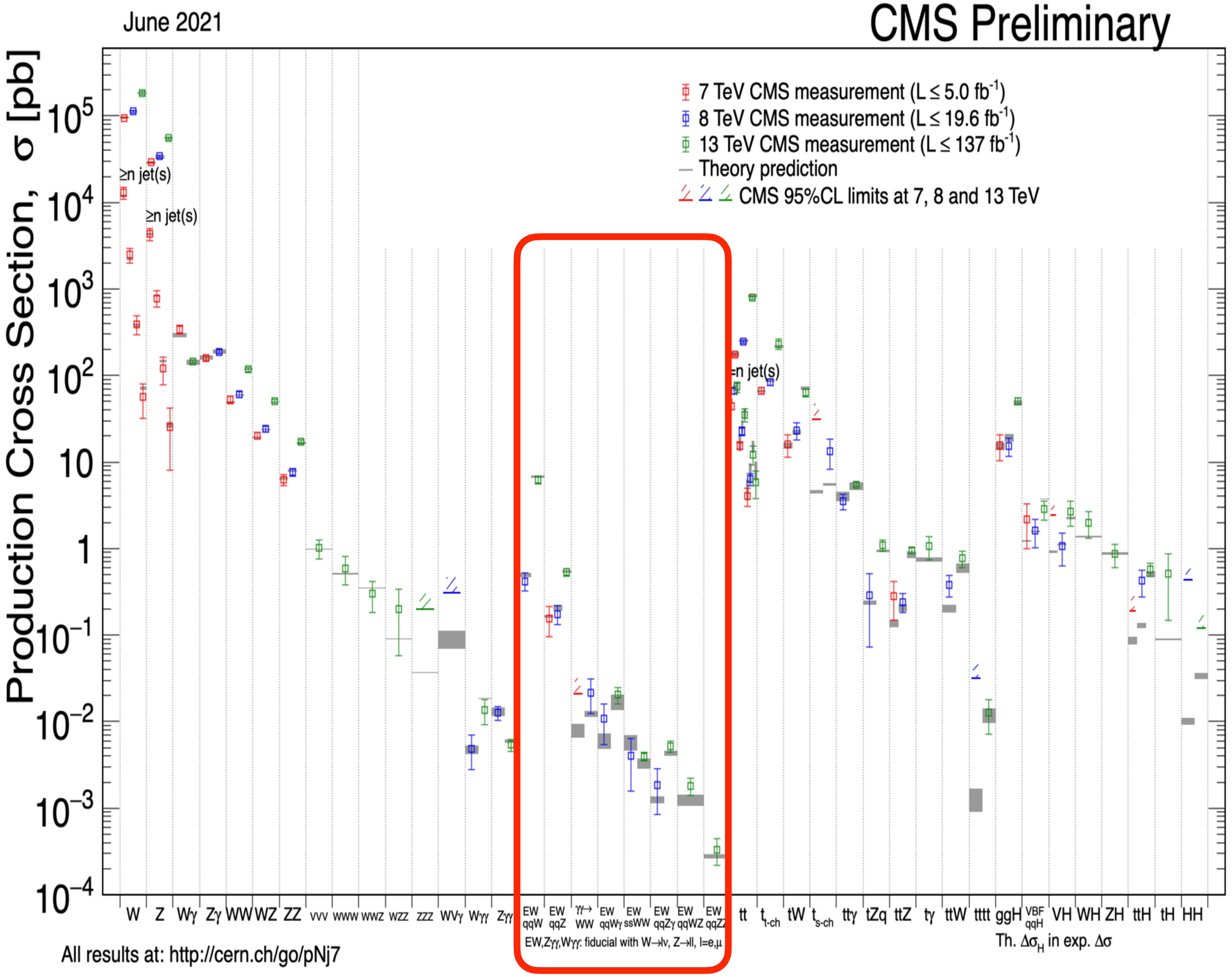
China LHC Physics Workshop (CLHCP - 2021)
Tsinghua University and Nanjing Normal University
25-28 November 2021

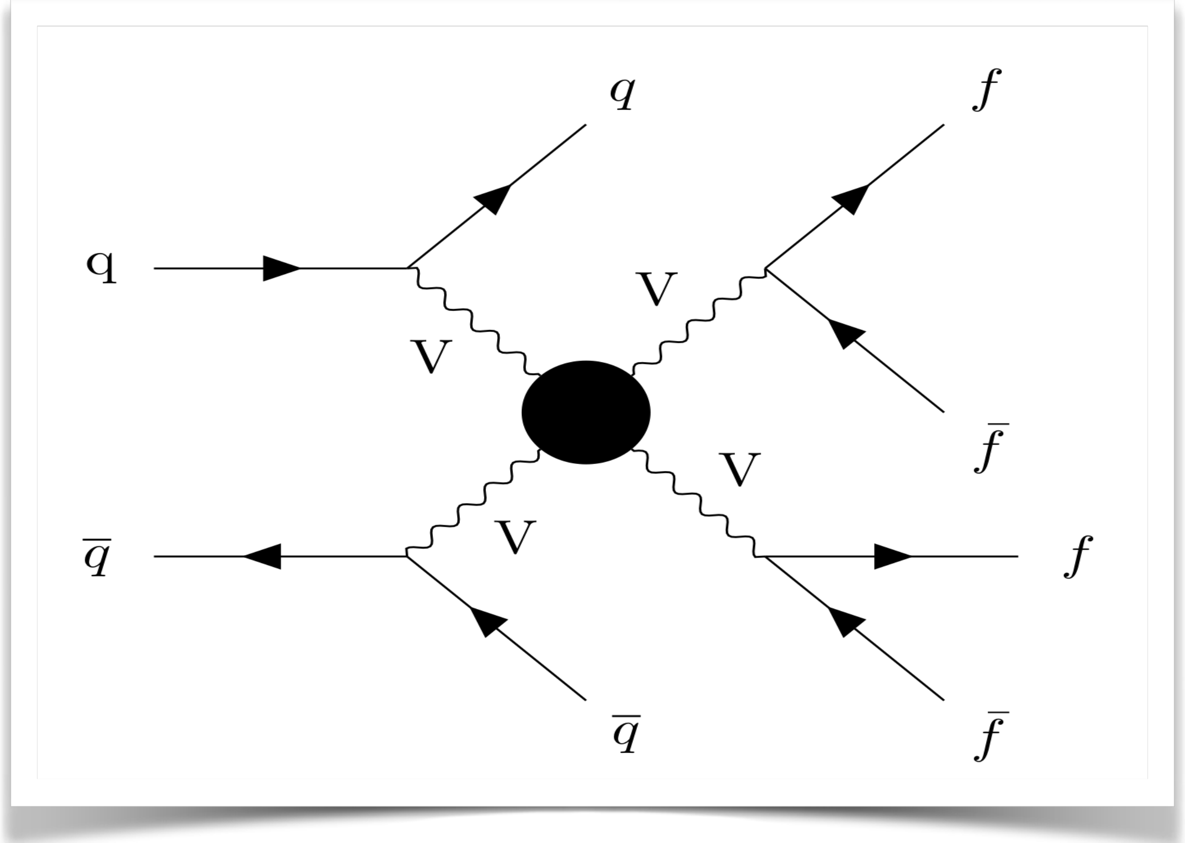
Background reference: https://www.symmetrymagazine.org/article/six-fabulous-facts-about-the-standard-model

Vector Boson Scattering

- Tested and measured SM cross-sections over 10 orders of magnitude
- Discovered Higgs
- **Still need to understand the mechanism** behind the Electroweak Symmetry Breaking Mechanism (**EWSB**).

- Without Higgs, Vector Boson Scattering (VBS) cross-section violate unitarity at the TeV scale.
  - VBS: the radiated gauge bosons by quarks from the two protons and interact with each other and decay afterwards.
  - VBS measurement gives way to probe the EWSB.
  - Rare process in SM, requires good discrimination against enormous backgrounds.
    - With advancement in ML techniques, provide way to improve signal discrimination
  - First evidence of VBS in WV semi-leptonic channel at LHC.★
  - Also, longitudinal polarised WW scattering is probed first time at CMS.

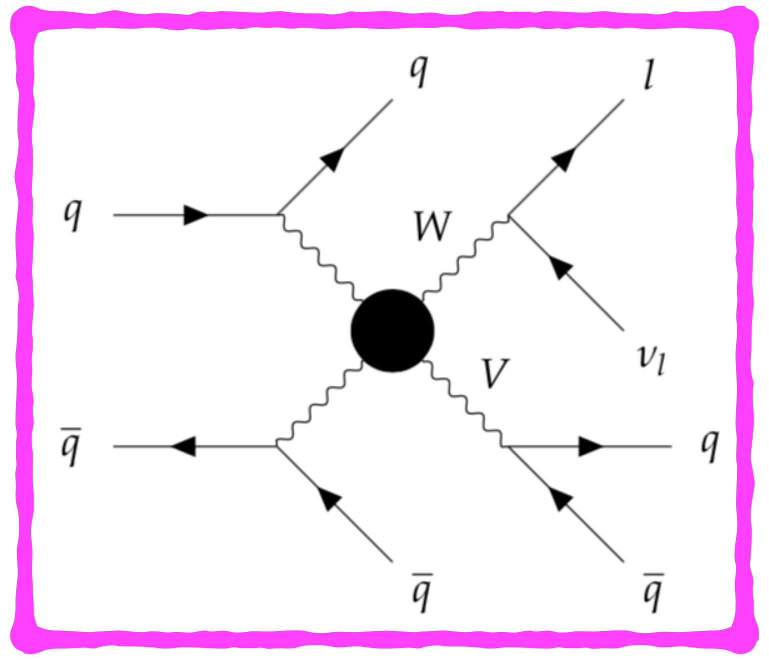


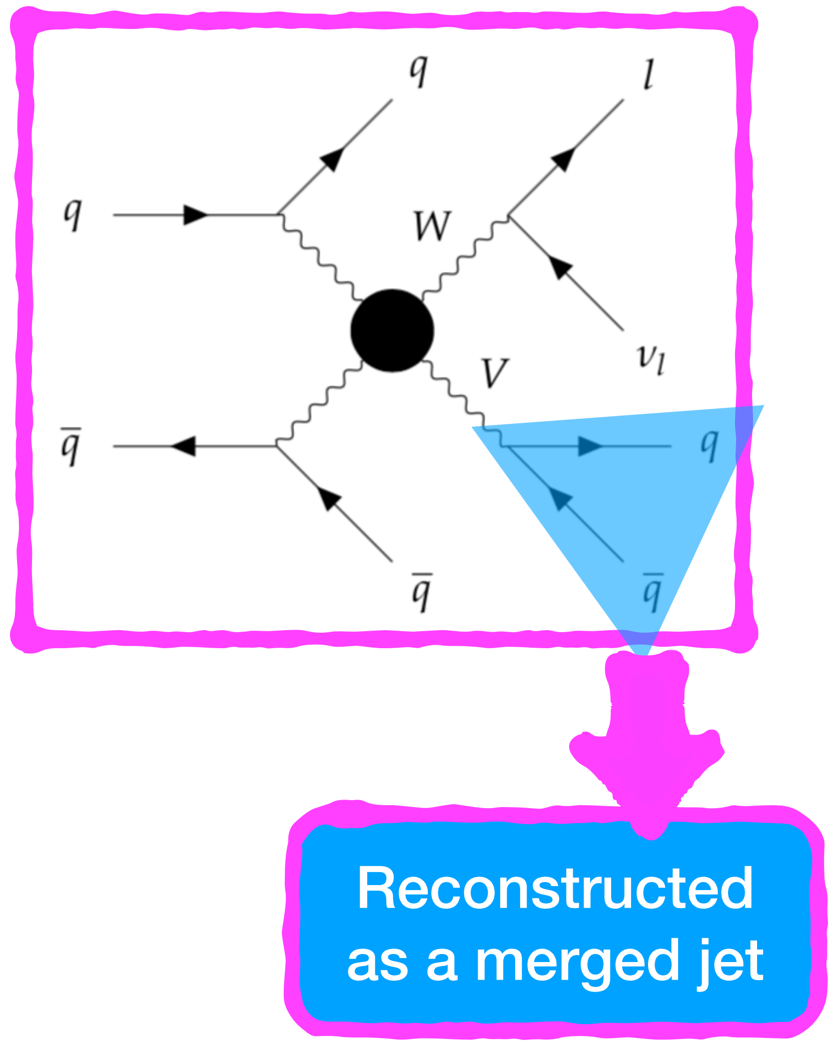


VBS Diagram

Motivation of WV Channel (V = W/Z)

- WV production in association with two jets
  - Semi-leptonic final state with a boosted & resolved hadronic W/Z
- Benefits:
  - **larger branching** ratio than same sign analysis WW final state.
  - **Full WW invariant mass reconstruction** (neutrino pz calculation by constraining W-boson mass)
  - **Contribution from all possible QGC and TGC vertex** (for WVjj process):
    - γWW, ZWW, WWWW, ZZWW, and WWγγ





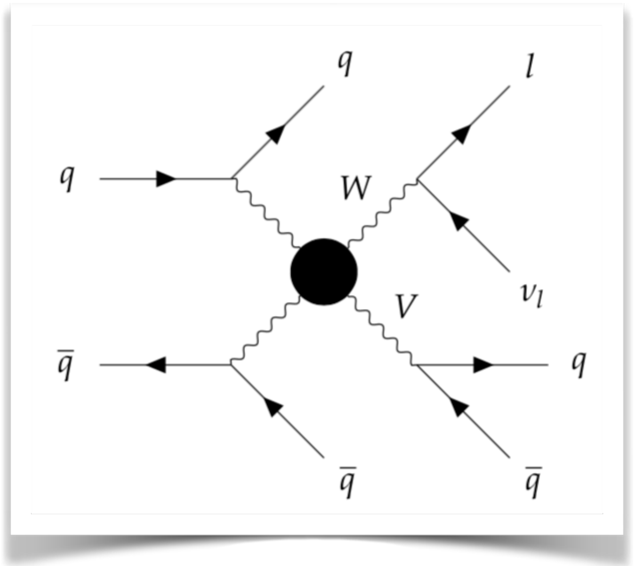
Signal and Background Process

- WVJJ (EWK) : Electroweak production of WVJJ.
- WVJJ (QCD initiated): Irreducible background for analysis.
- **W+Jets: Most dominating background.**
- Drell-Yan: Z/Gamma decays to l+l- and we mis-measure one lepton because of acceptance or inefficiency effects, gives missing energy.
- tt Jets: Top quark always decays to one b-quark and one W boson. So, $t\bar{t} \rightarrow bWbW \rightarrow bl\nu l\nu$, if we mis-measure one lepton and b quark form jets.
- Single top production: Here $t \rightarrow bW \rightarrow bl\nu$, and 3 jets is reconstructed.

Event Selection

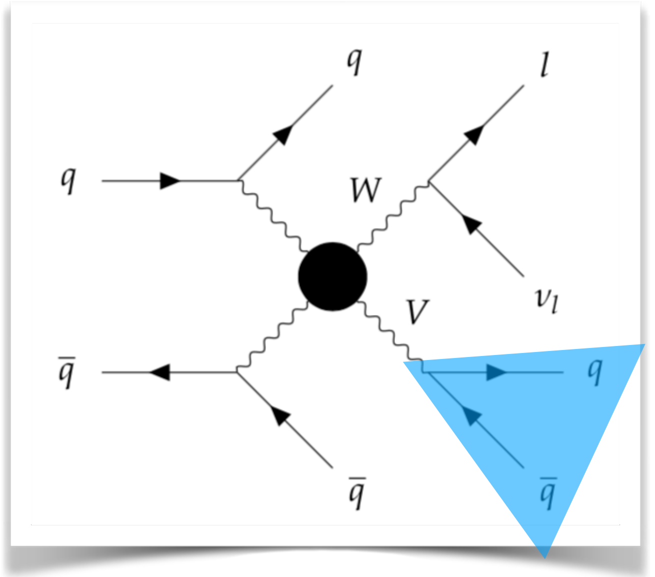
Event selection:

- **Two VBS jets**
- **Two vector bosons WV:**
  - Leptonic decays of W bosons into electron and muons with associated neutrinos
  - V (= W/Z) always decay hadronically. It has two categories:
    - Boosted category: Reconstructed as a fat jet having radius parameter of 0.8
    - Resolved category: Reconstructed as two resolved jets having radius parameter of 0.4



Resolved Category

Two jets, pT>30 GeV



Boosted Category

One fat jet pT > 200 GeV

- VBS Topology:
  - **High pseudo-rapidity gap** between VBF jets: $\Delta\eta_{jj} > 2.5$
  - **Larger di-jet invariant mass**: $M_{VBS\text{-}jj} > 500$ GeV

- Additional requirement to enhance VBS:
  - Zeppenfeld Variable : $Z = \dfrac{\eta - \dfrac{\eta_{j1} + \eta_{j2}}{2}}{|\eta_{j1} - \eta_{j2}|}$
  - Centrality:

$$\xi_V = min\{\Delta\eta_-, \Delta\eta_+\}$$

where,

$$\Delta\eta_- = min\{\eta(V_{had}), \eta(V_{lep})\} - min\{\eta_{j1}, \eta_{j2}\},$$
$$\Delta\eta_+ = max\{\eta_{j1}, \eta_{j2}\} - max\{\eta(V_{had}), \eta(V_{lep})\}$$

- **Signal extraction using DNN**

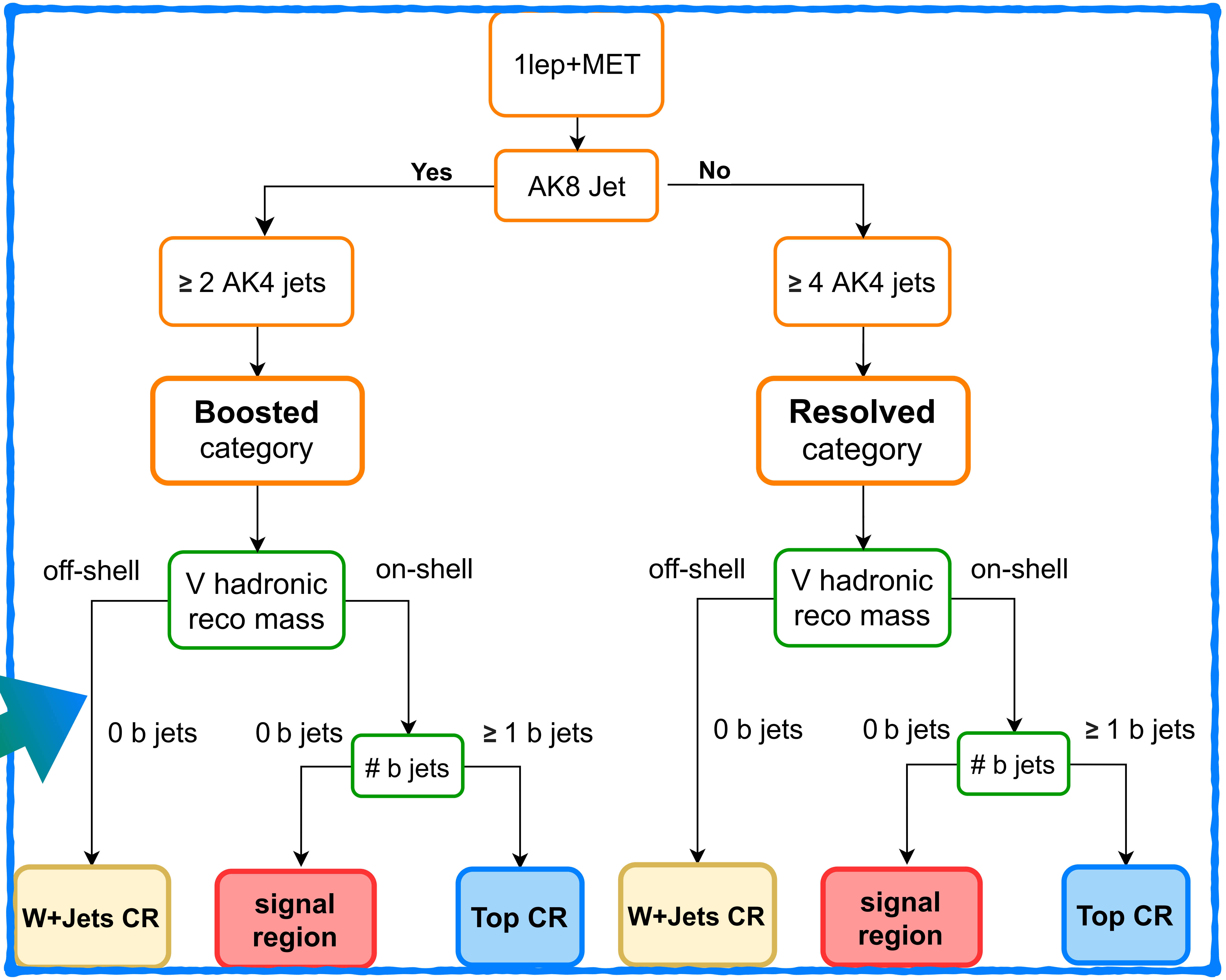
Analysis Phase Space

- Control region (CR): Region orthogonal to the signal region
  - **W+jet CR:** $m_W < 65\ GeV\ or\ m_W > 115\ GeV$
  - **Top CR: requires $\geq$ 1 b-jets**
- Split according to leptons flavour
- Final fit combining all regions

Object selection

Phase space selection

Simultaneous fit regions



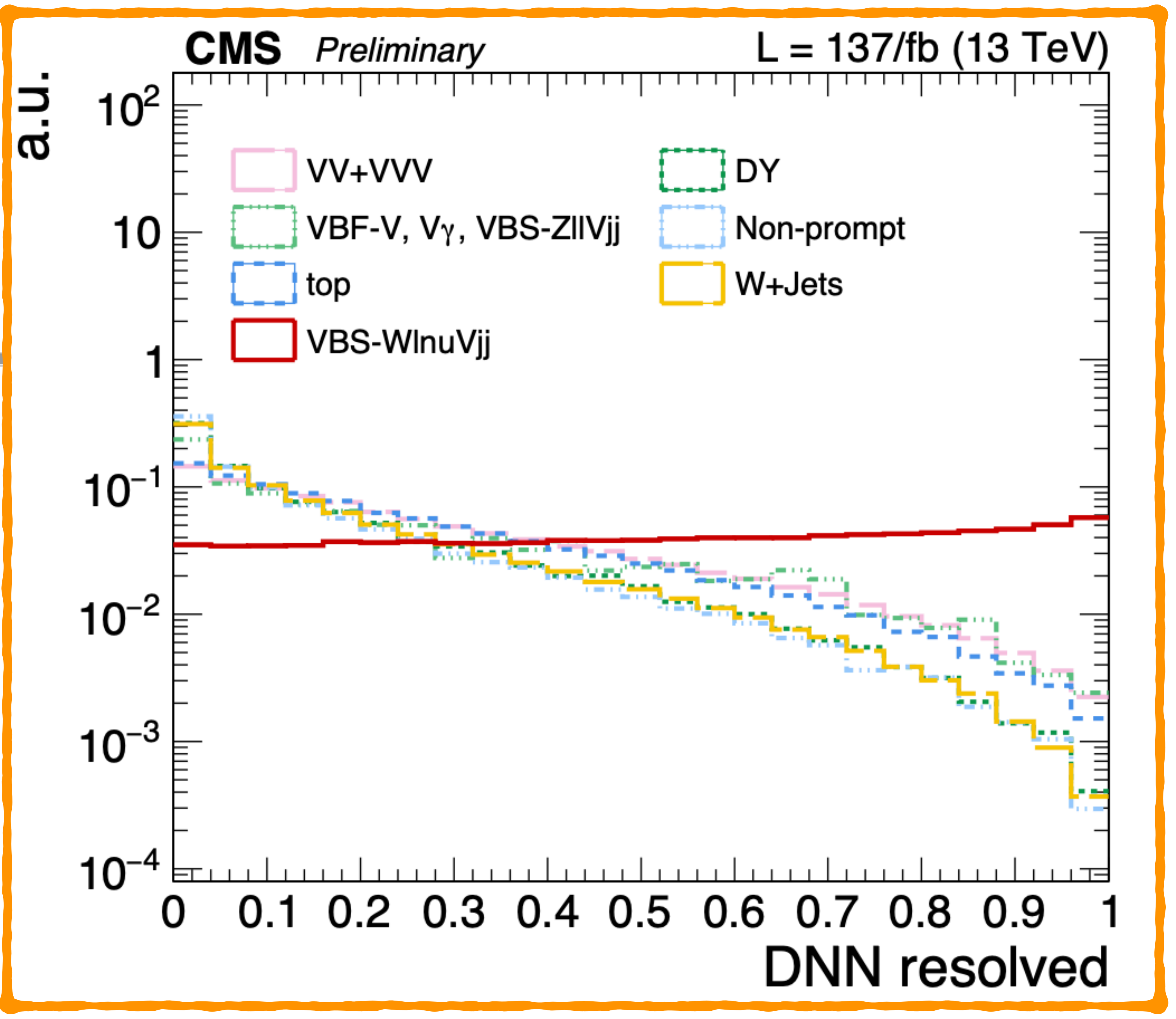
Deep Neural Network (DNN) Training

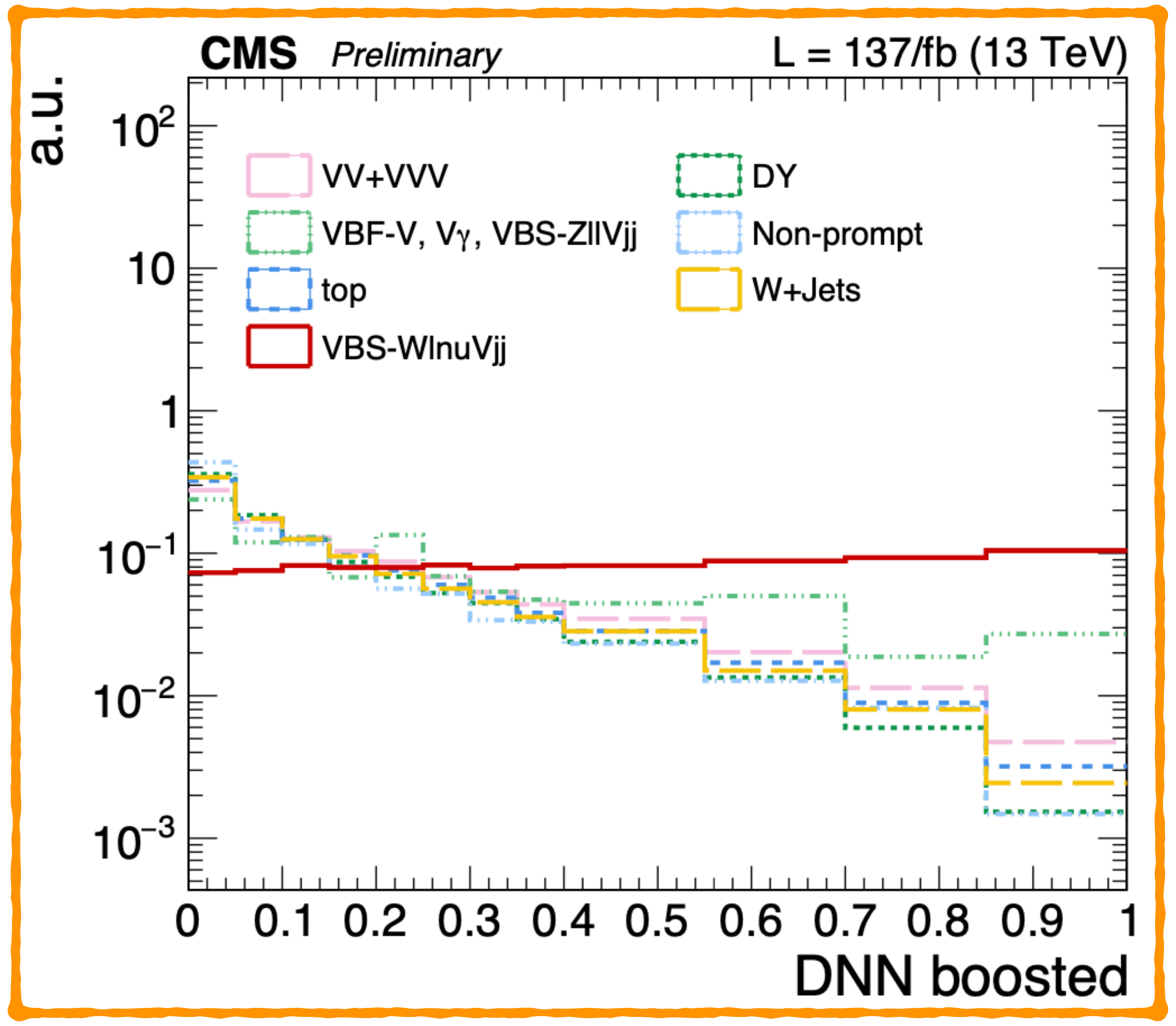
- **Binary DNN** trained with VBS as signal vs all backgrounds
- Separate model for resolved and boosted categories:
  - Resolved category: Fully connected DNN having 64-64-64-64 nodes with 16 input variables
  - Boosted category: Fully connected DNN having 64-32-32-32 nodes with 13 input variables
- To avoid overtraining dropout layers and L2 weight regularisation was used.

Variable	Resolved	Boosted
Lepton pseudorapidity	✓	✓
Lepton transverse momentum	✓	✓
Zeppenfeld variable for the lepton	✓	✓
Number of jets with $p_T > 30$ GeV	✓	✓
VBS leading tag-jet p_T	-	✓
VBS trailing tag-jet p_T	✓	✓
Pseudorapidity interval between VBS tag-jets	✓	✓
Quark Gluon discriminator of the highest p_T jet of the VBS tag-jets	✓	✓
Azimuthal angle distance between VBS tag-jets	✓	✓
Invariant mass of the VBS tag-jets pair	✓	✓
p_T of jets from V_{had}	✓	-
Pseudorapidity difference between V_{had} jets	✓	-
Quark Gluon discriminator of the V_{had} jets	✓	-
V_{had} p_T	-	✓
Invariant mass of the V_{had}	✓	✓
Zeppenfeld variable for the V_{had}	-	✓
V_{had} centrality	-	✓

DNN Input variables

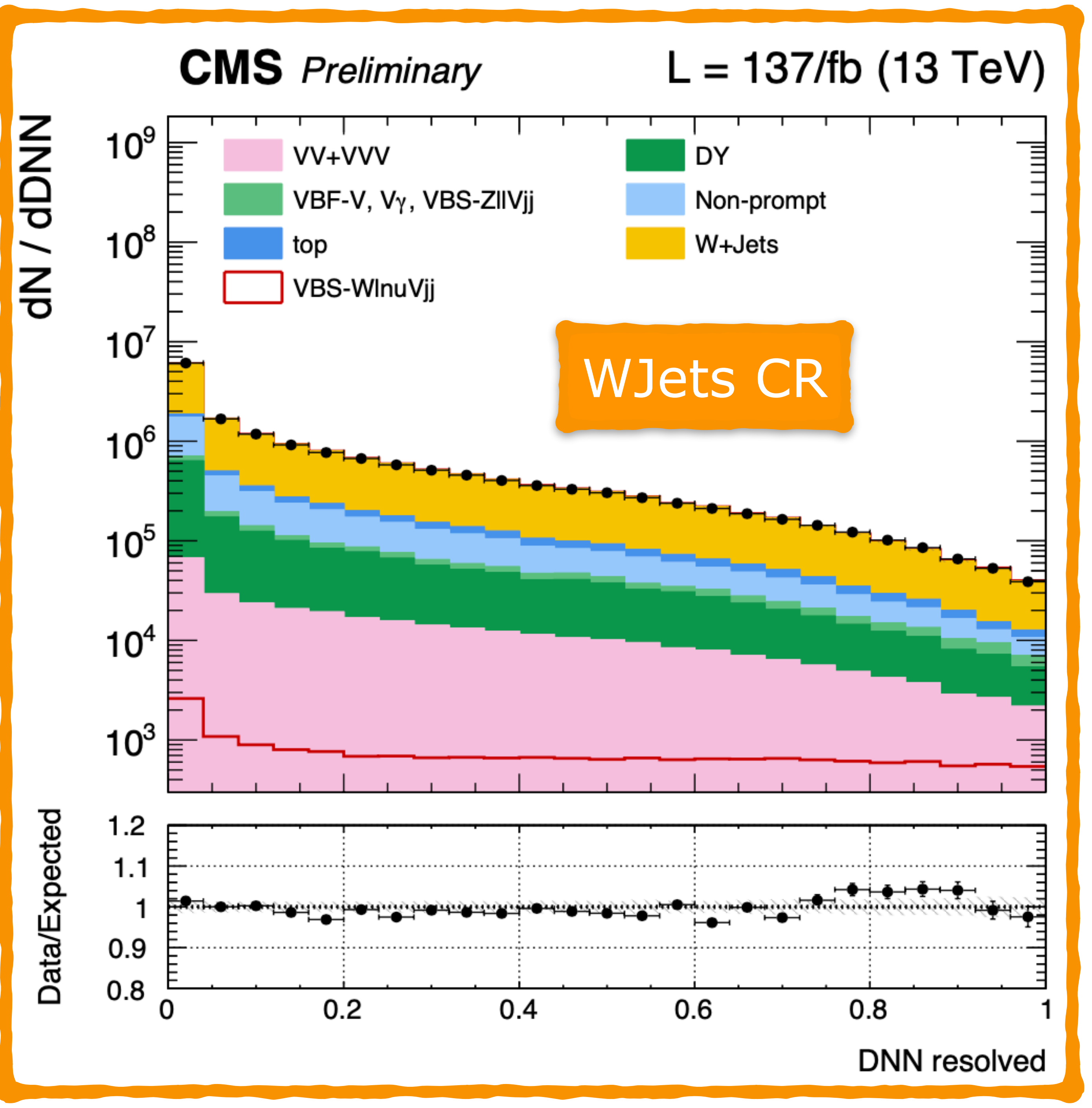
DNN Score Distribution

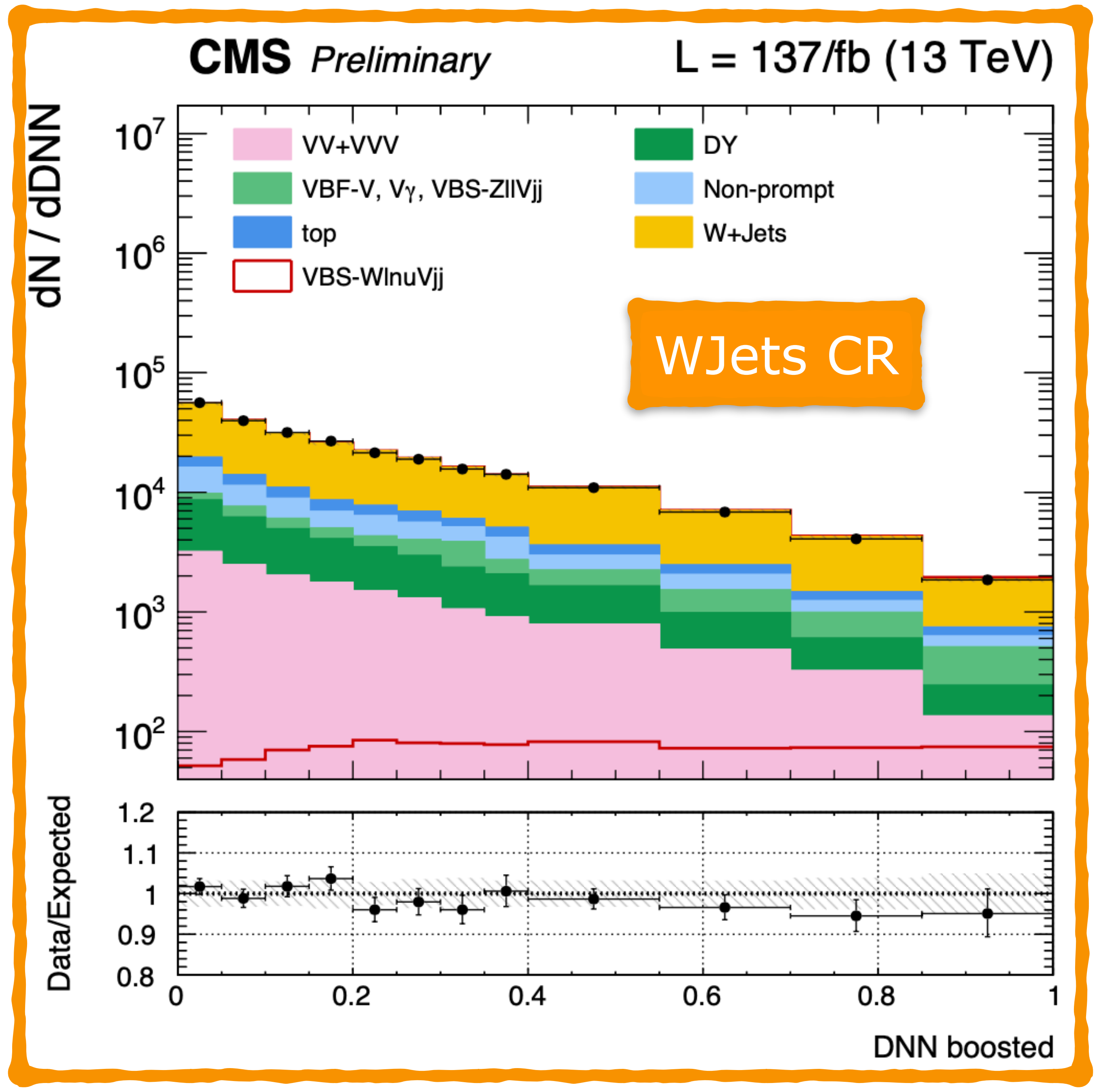




Background Estimation

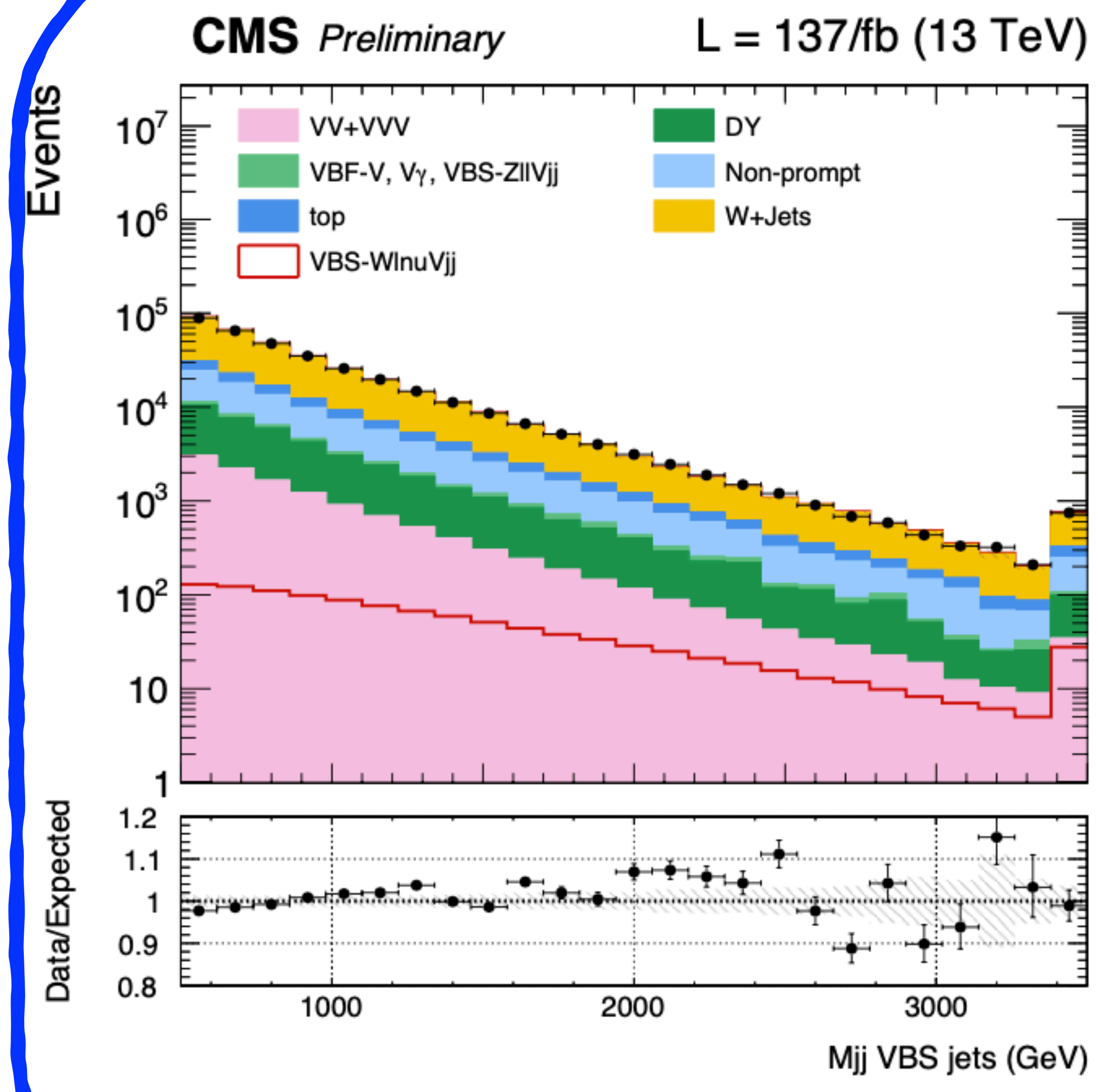
- QCD multijet background: Estimated using data-driven method
  - Tight/Loose efficiency for fake (prompt) lepton measured in QCD (Drell-Yan) enriched data sample
  - Construct relation between # prompt/ fake leptons and # passing/failing tight ID
  - Weight events in Loose data control region by (probability to have at least one non-prompt lepton) x (probability to still pass tight selection)
- W+Jets:
  - Mismodelling of the jet pT spectrum for W+many-jets sample → data-driven differential corrections
  - W+Jets contribution taken from MC but corrected in a data driven way.
  - Fit their normalization in the global fit in the W+jets CR
  - After the data driven estimation → Predictions and data agree within uncertainties
- Top background: Shape from MC, normalisation extracted from top CR in the final fit to the data



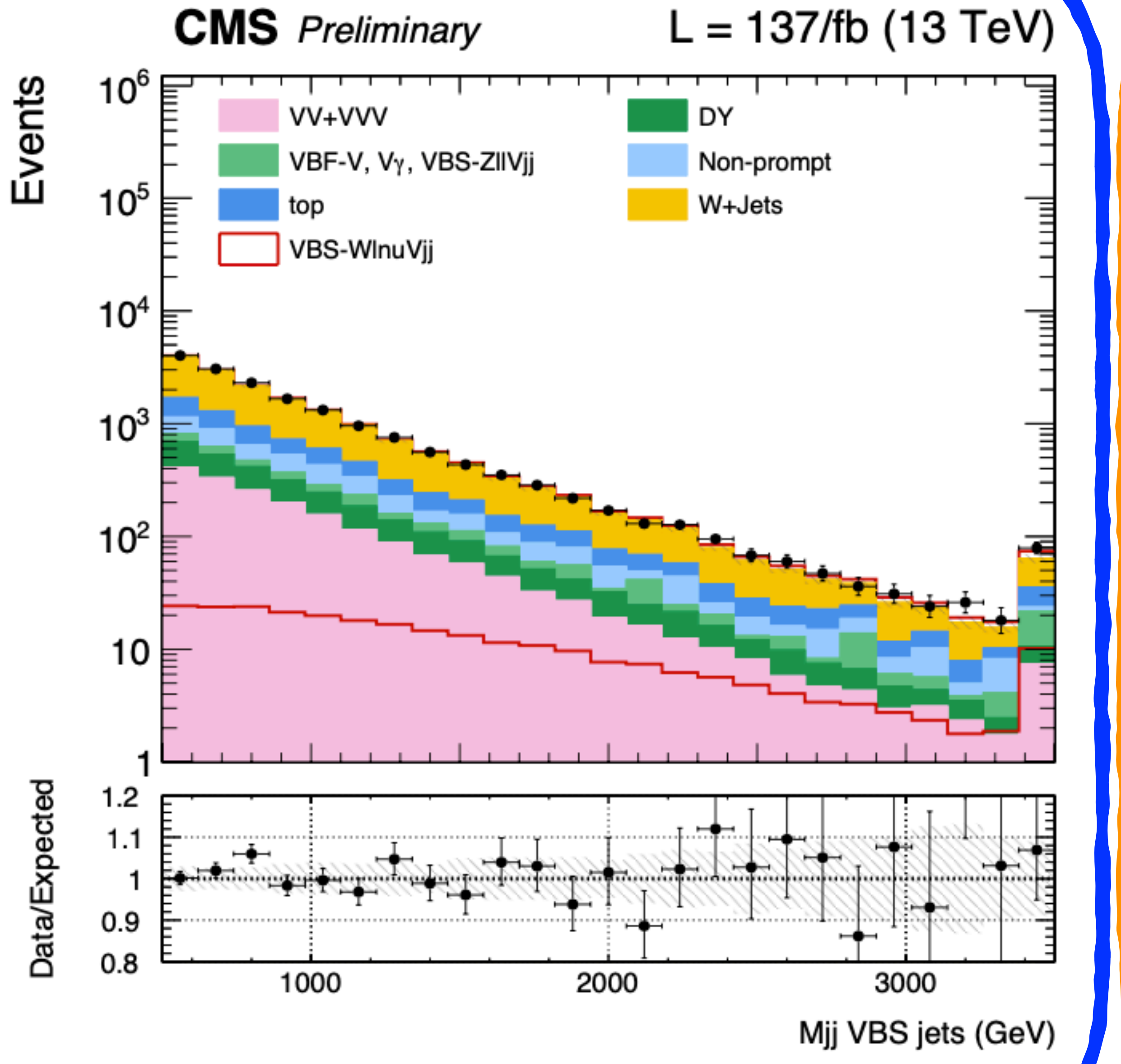


Signal Region

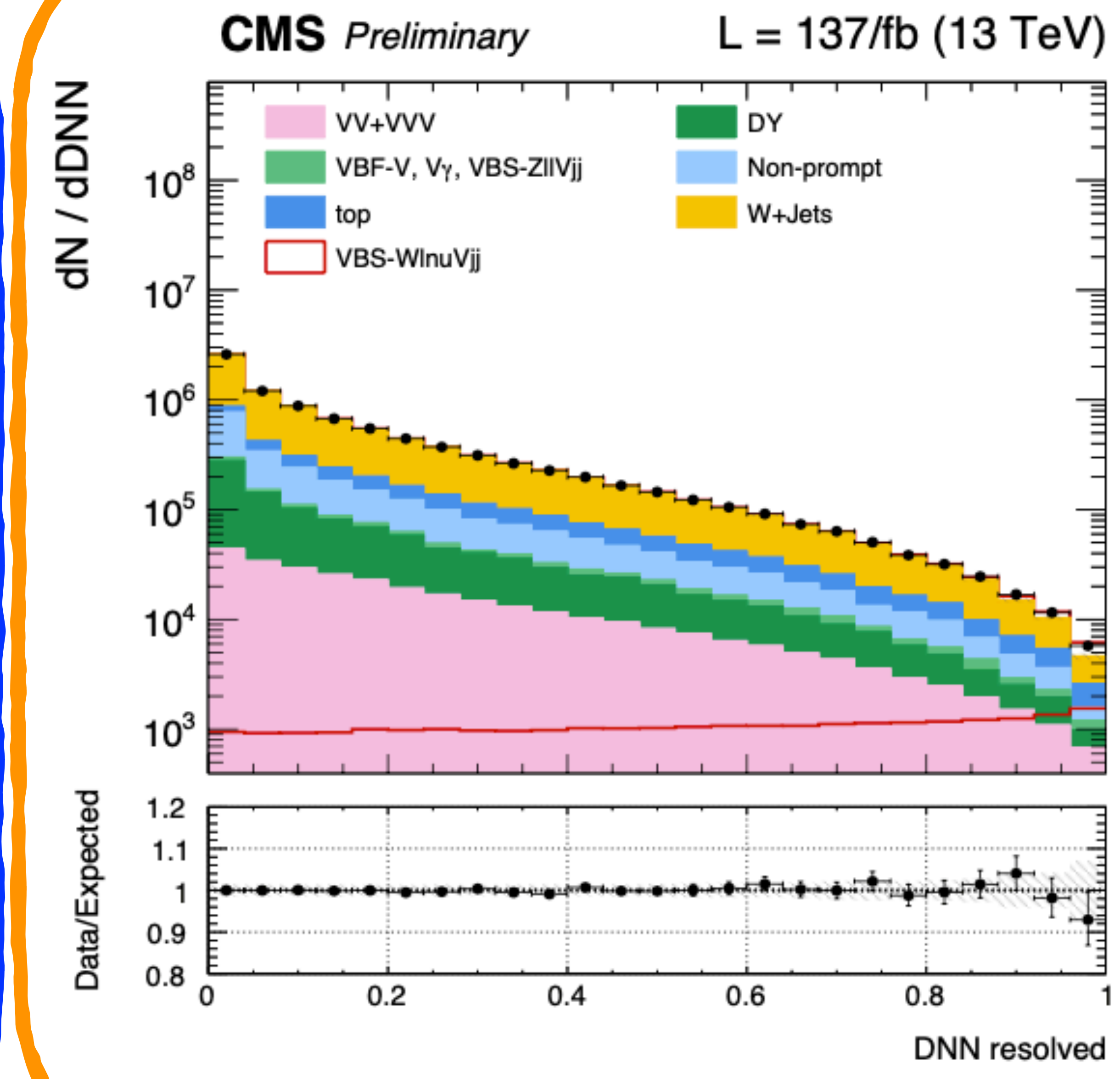
- Signal region distribution for one of highest ranked variable, di-jet invariant mass of VBF jets and DNN score, for both resolved and boosted case



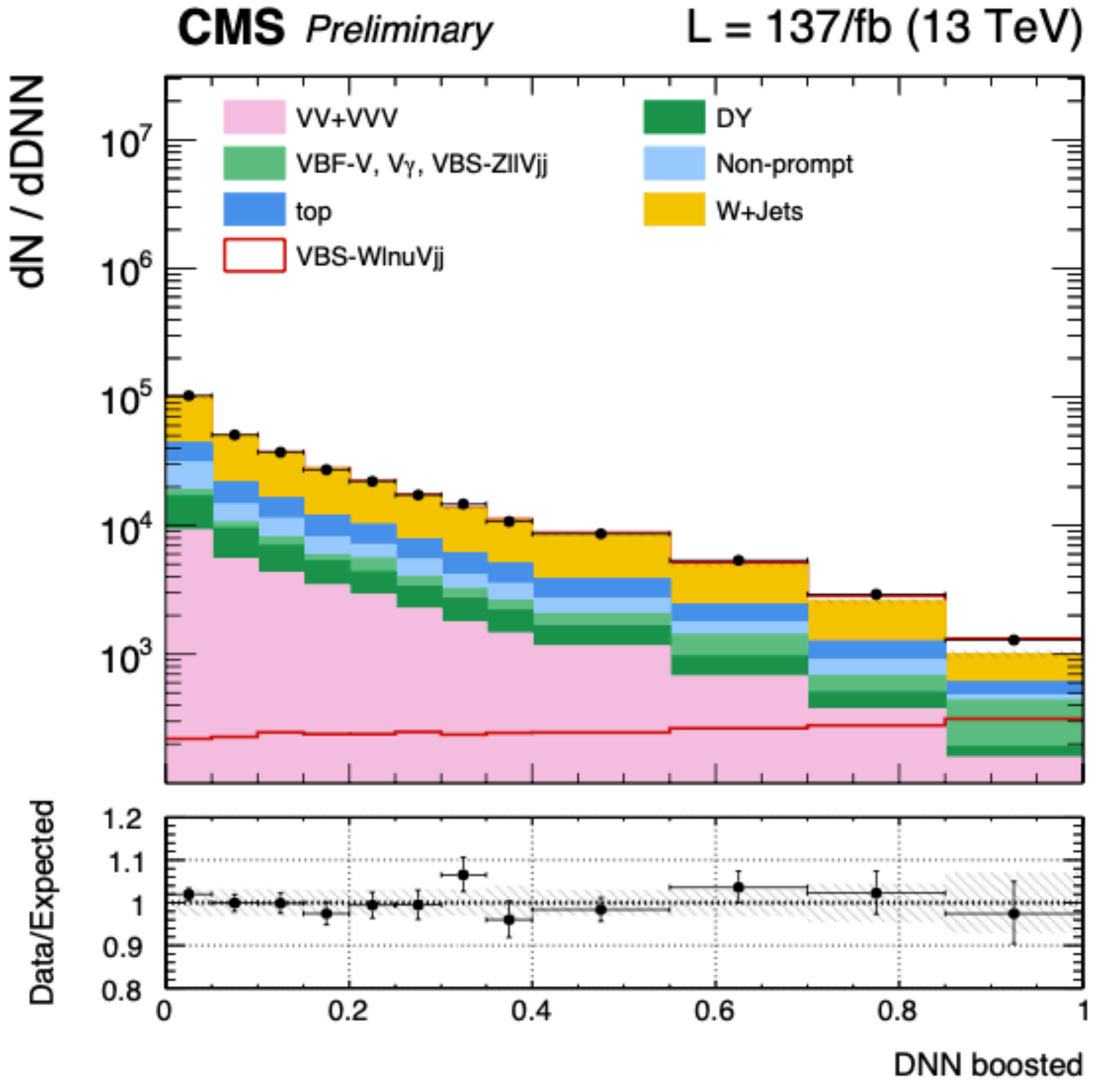
Resolved Category



Boosted Category



Resolved Category



Boosted Category

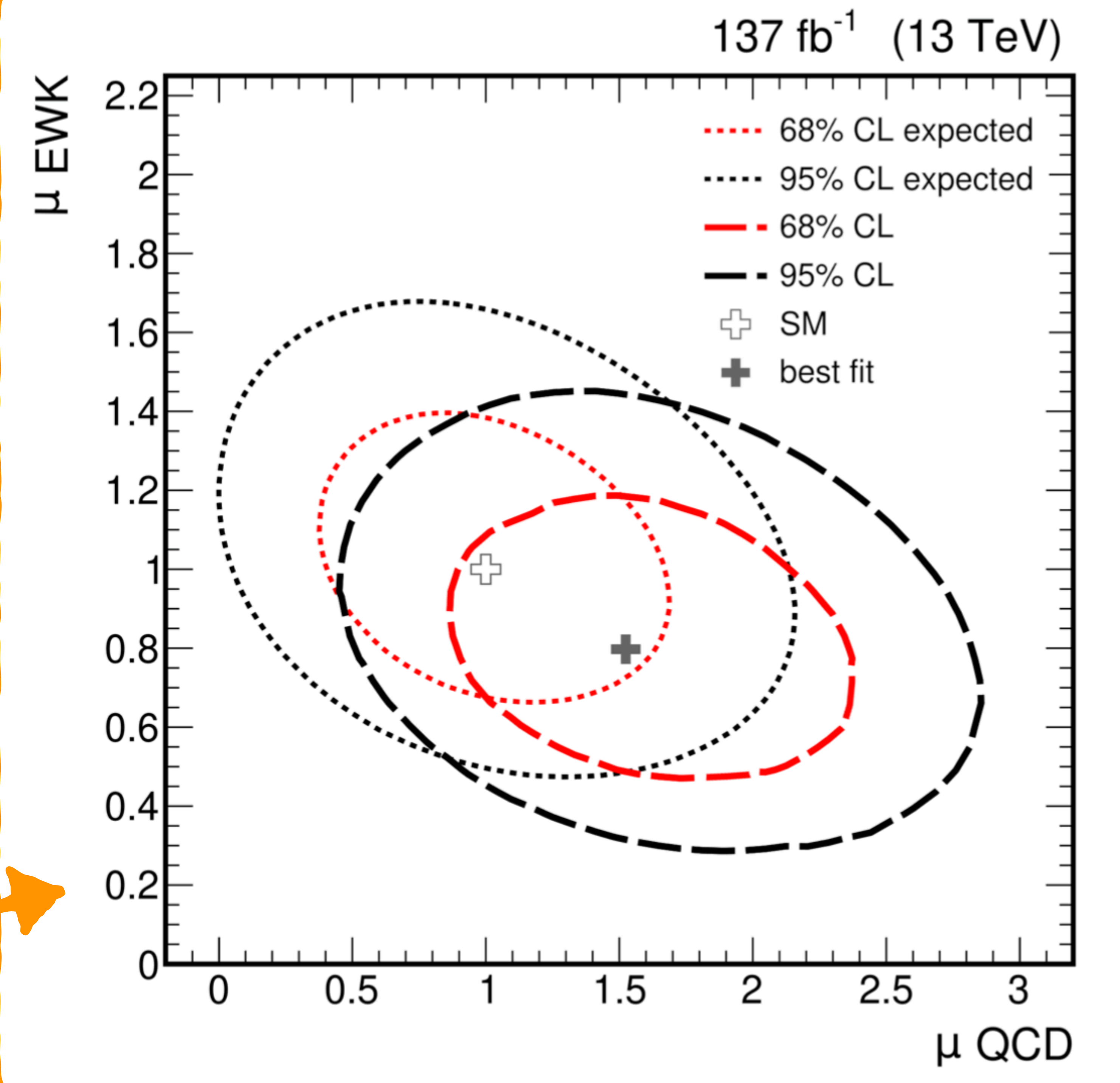
Systematic Uncertainty

- Systematic uncertainty can affect the shape and normalisation of the DNN distribution.
    - Largest impact is from statistics.
        - Expected as VBS signal is a rare process.
    - Experimental uncertainty is mainly dominated by b-tagging and jet energy scale/resolution

Uncertainty source	$\Delta\mu_{EW}$
Statistical	0.12
Limited sample size	0.10
Normalization of backgrounds	0.08
Experimental	0.06
b-tagging	0.05
Jet energy scale and resolution	0.04
Luminosity	0.01
Leptons identification	0.01
Boosted V boson identification	0.01
Theory	0.12
Signal modeling	0.09
Background modeling	0.08
Total	0.22

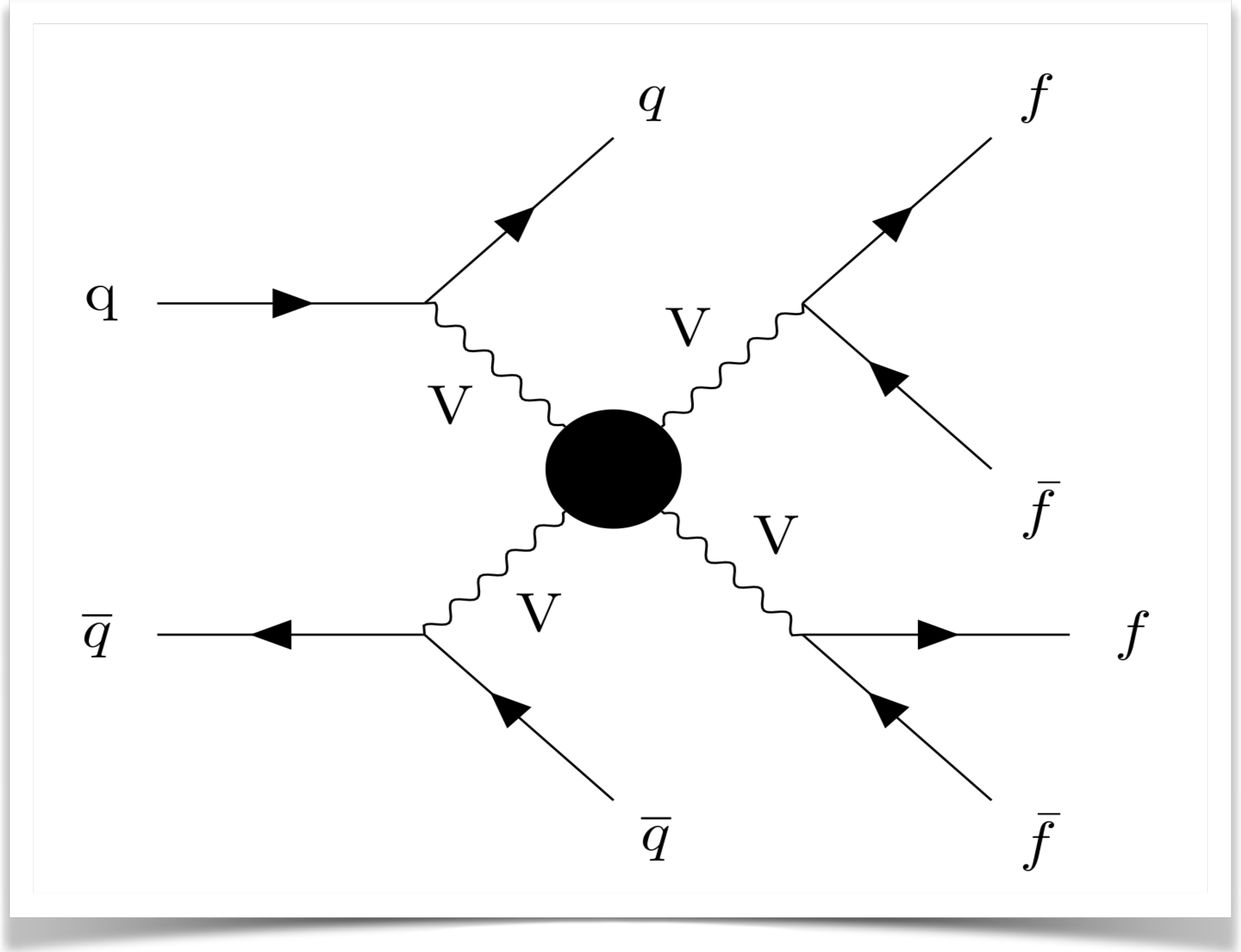
Summary & Results

- Fit DNN shape in the signal regions
- Fit W+jets subcategories normalizations in W+jets control regions
- Fit only normalization in top-quark control regions
- Results:
  - **SM EW signal strength**:
    $\mu_{EW} = \sigma^{obs}/\sigma^{SM} = 0.85^{+0.24}_{-0.20} = 0.85^{0.21}_{-0.17}\ (syst)^{+0.12}_{-0.12}\ (stat)$
    Signal significance of $4.4\ \sigma$ ($5.1\ \sigma$ expected)
    - Observed fiducial cross-section ($m_{qq} > 100\ GeV,\ p_t^q > 10\ GeV$) of 1.9 ± 0.5 pb
  - Considering **EW and QCD WV production** as signal, the **signal strength**:
    $\mu_{EW} = \sigma^{obs}/\sigma^{SM} = 0.98^{+0.20}_{-0.17} = 0.98^{0.19}_{-0.16}\ (syst)^{+0.07}_{-0.07}\ (stat)$
    - Measured cross-section: $16.6^{+3.4}_{-2.9}$ pb
  - **Simultaneous 2D fit of the EW and QCD WV signal strengths**



Backup

Introduction: VBS



VBS diagram