University of Zurich[UZH]

CMS

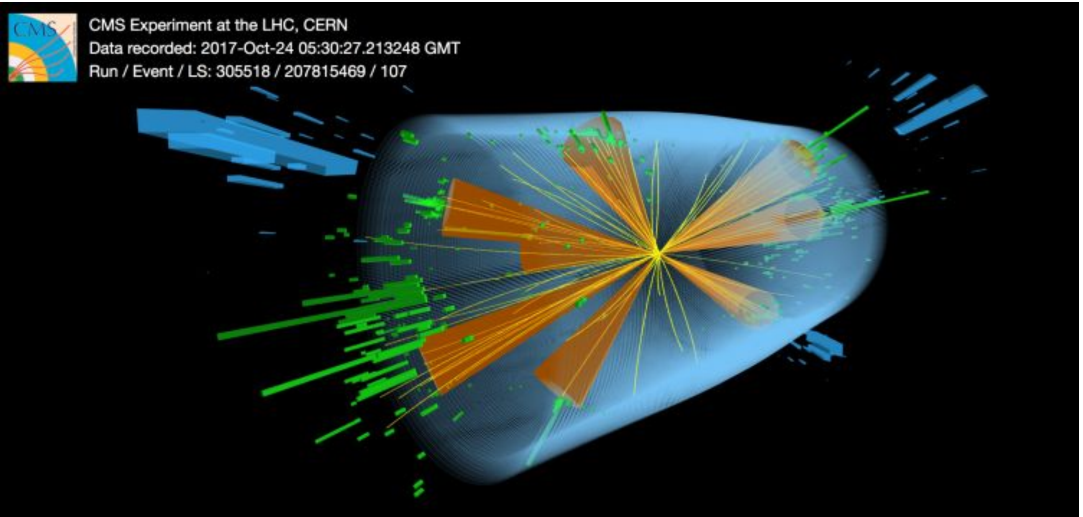
Compact Muon Solenoid

Measurement of the ttbb cross section in the all-jet final state with the CMS detector

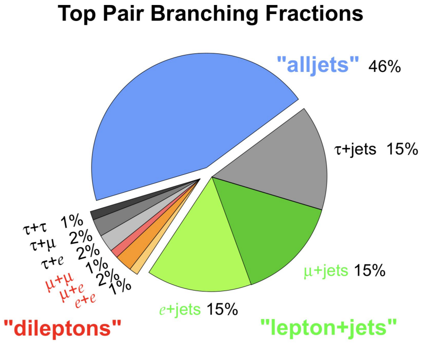
V. Mikuni

Introduction

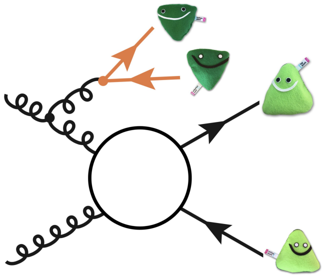
- A precise description of the production of **tt + b jets** is challenging due to difficult final state
- ttbb is one of the main backgrounds for **ttH(bb)** and **4 tops** analyses



- **All hadronic:**
  - **Largest branching fraction** from ttbar decays
  - **Lowest S/B**

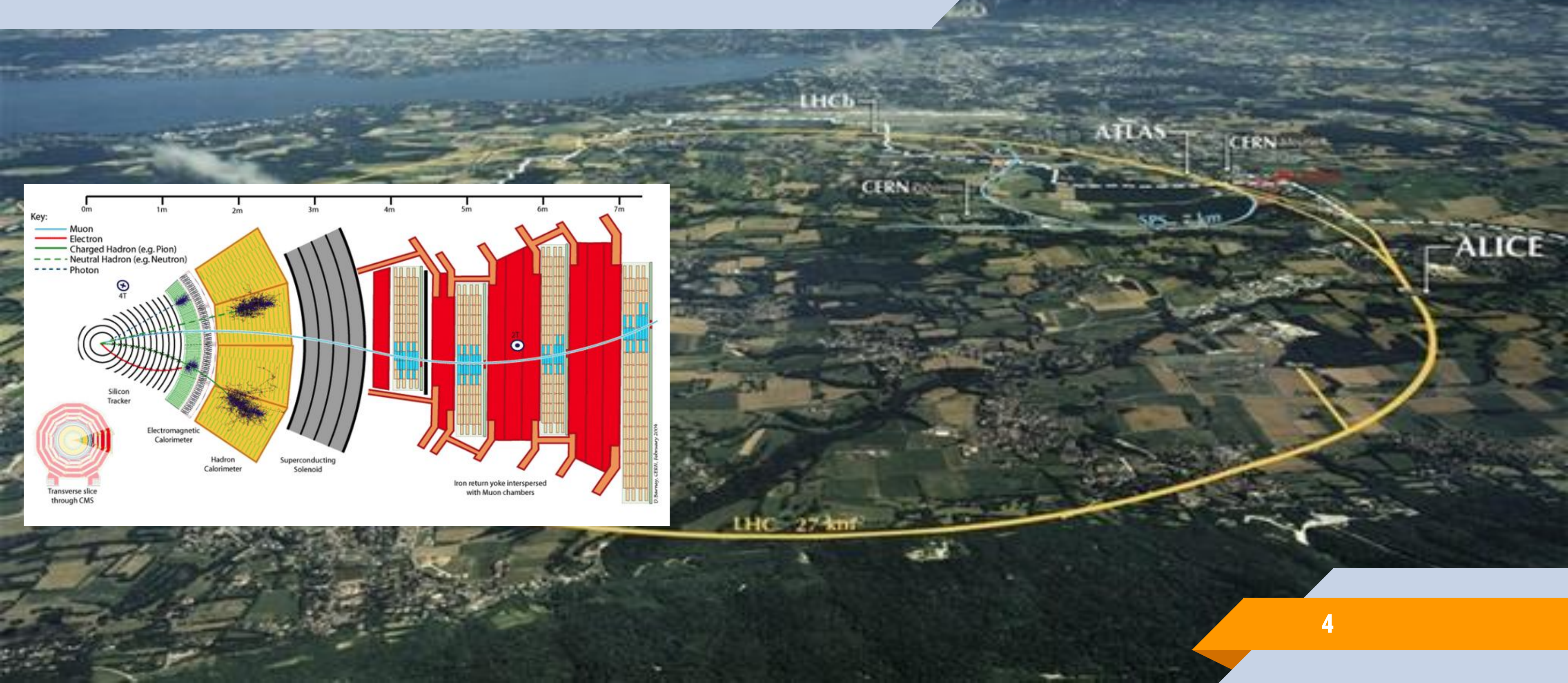


- Signature:
  - **8 jets** from which **4** should be **b-jets**



The CMS detector

The CMS detector



Region definition

- Measure the **inclusive cross section**
- Which one? Fiducial, total?
  - Answer: **Both!**

Total phase space

- ➔ At least **2 b-jets** matched to **2 B-hadrons not stemming from the top decays**
- ➔ Additional b-jets **p_T > 20 GeV** and **$|\eta|$ < 2.4**

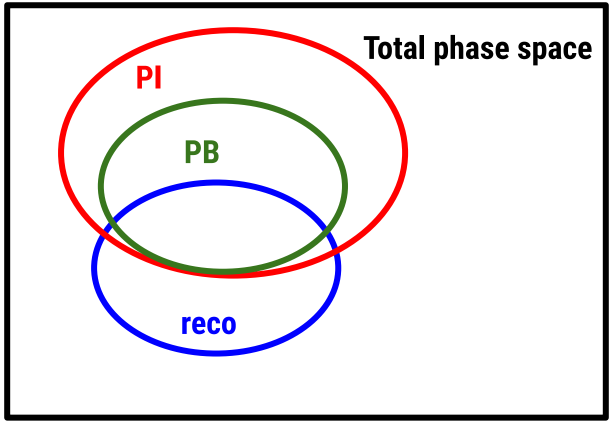
Fiducial phase space Parton independent (PI)

- ➔ At least **4 b-jets** at particle level
- ➔ At least **6 jets** with **p_T > 30 GeV**
- ➔ At least **8 jets** with **p_T > 20 GeV**
- ➔ All jets **$|\eta|$ < 2.4**

Fiducial phase space Parton based (PB)

- ➔ At least **4 b-jets** at particle level
- ➔ At least **6 jets** with **p_T > 30 GeV**
- ➔ At least **8 jets** with **p_T > 20 GeV**
- ➔ All jets **$|\eta|$ < 2.4**
- ➔ **At least 2 b-jets not stemming from the top decays**

Events



Event selection and MC samples

2016 data (35.92 fb-1)

Main signal sample:

- Inclusive **Powheg + Pythia 8** NLO ttbar

Main backgrounds:

- Multijet production (data driven) (**90%** of all collected events)
- ttbar + other jets (**8%** of all collected events)

Signal contributions:

ttbb + tt2b + ttb (split using the GenHFHadronMatcher)

ttbb

ttb

tt2b

Initial selection:

- At least **6 jets** with p_T > 40 GeV and $|\eta|<2.4$
- **2 or more** b-tagged jets
- HT > 500 GeV
- Additional jets with p_T > 30 GeV
- Lepton veto

BDT for permutations

BDT designed to identify the **top pair system**

Take into account all possible **distinguishable combinations**:

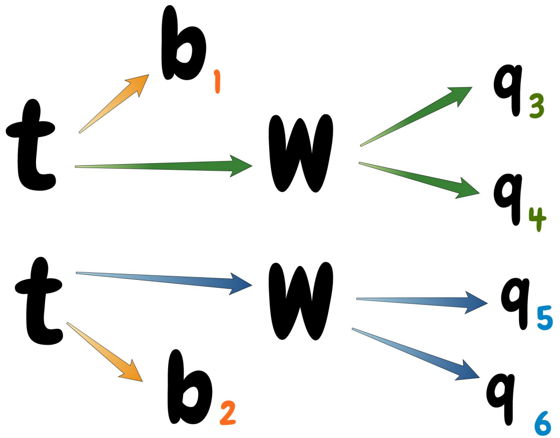
8 jets in the event = **2520** combinations

Only combinations with prob(χ^2) > 1e-6 are accepted to reduce the combinatorial background

About **60% efficiency** for ttbb identification

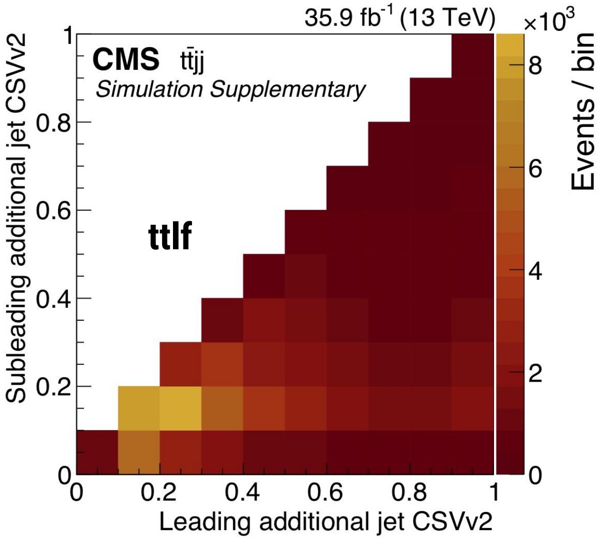
Variables: **26**:

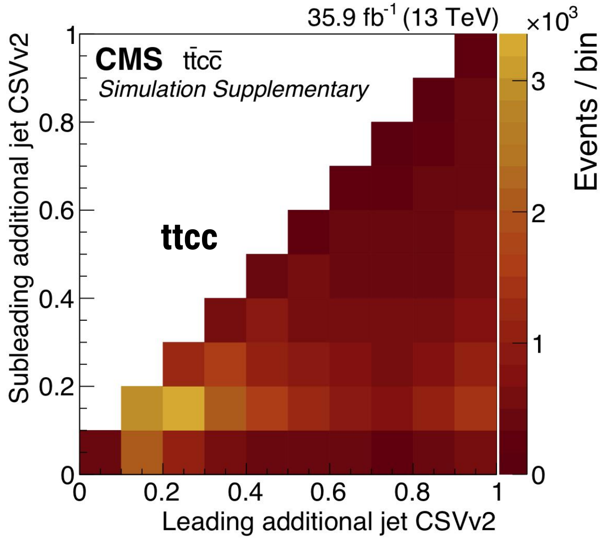
- Invariant masses
- ΔR between jets
- $\Delta\phi$ between jets
- $\Delta\eta$ between jets
- B-tagging discriminant for all selected jets
- χ^2 for the combination

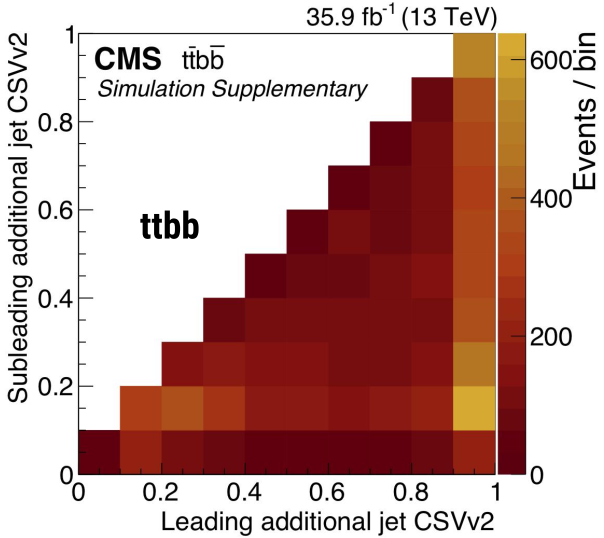


$$\chi^2 = \left[\frac{(1+3+4).M() - top_M}{\Gamma_t}\right]^2 + \left[\frac{(2+5+6).M() - top_M}{\Gamma_t}\right]^2 + \left[\frac{(3+4).M() - W_M}{\Gamma_W}\right]^2 + \left[\frac{(5+6).M() - W_M}{\Gamma_W}\right]^2$$

Main discriminant

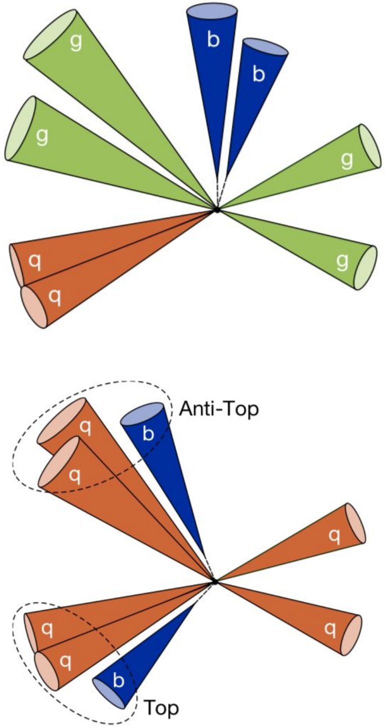


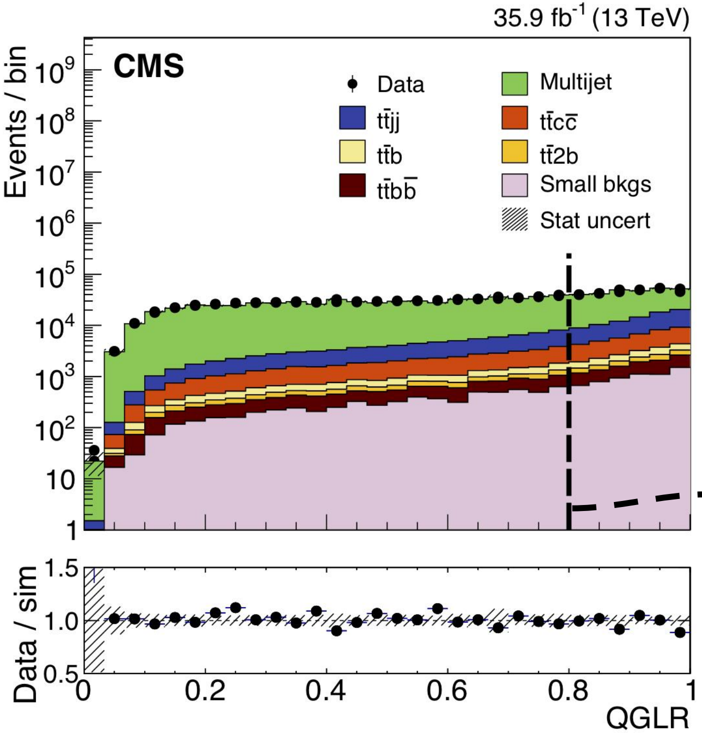




- **b-tagging** discriminant of the jets not selected by the **permutation BDT**
- Additional jets ordered by **b-tagging** discriminant value

Multijet rejection: QGLR





- In average, multijet events have **more** gluon initiated jets than tt+jets
- Build a **likelihood ratio** variable to compare how likely the event is to have **4 quark-initiated jets**

Cut used for the Control regions definition

Multijet rejection: Classification Without Labels (CWoLa*)

- Training a classifier using **multijet simulations** is complicated:
  - Difficult to simulate **enough representative events** in the PS of interest
  - Any data/simulation discrepancies decrease the performance
- What if we instead used **data directly**?

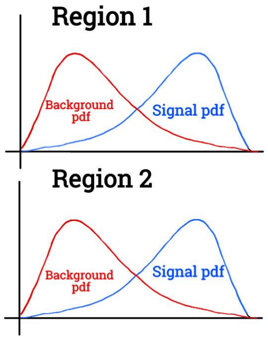
*[EMM, B. Nachman, and J. Thaler, arXiv: 1708.02949]
[T. Cohen, M.Freytsis, and B.Ostdiek, arXiv: 1706.09451]

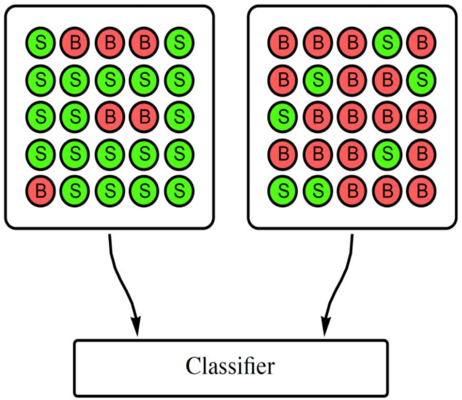
Data has no labels, so we can only define regions with different fractions of **tt+jets** and **QCD**

- **Hypotheses:**
  - Separate only 1 signal versus 1 background : **tt+jets vs. QCD**
  - The pdf distribution for signal and background events in region 1 has to be the same as region 2: **can only use distributions that are blind to the separation method**

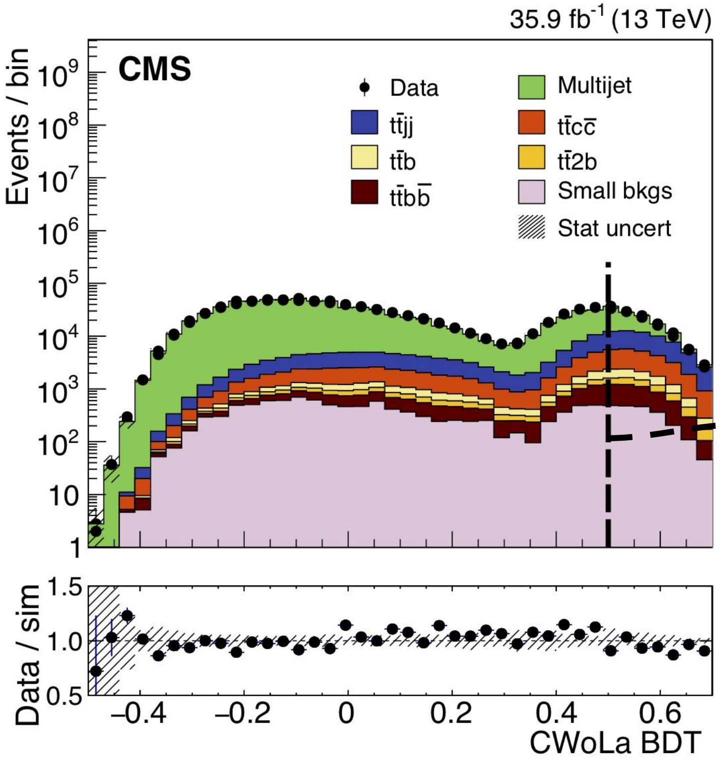
8 jets or more — **Signal region**

7 jets — **Region 1:** QGLR > 0.95 | **Region 2:** QGLR < 0.95

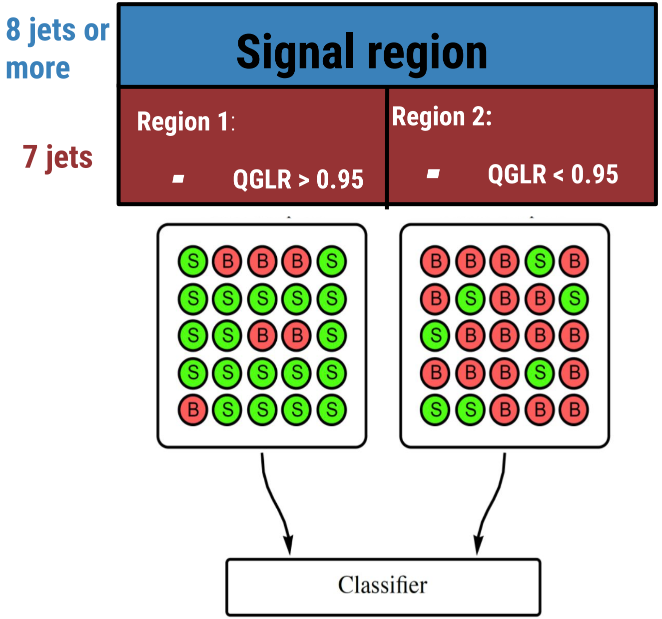




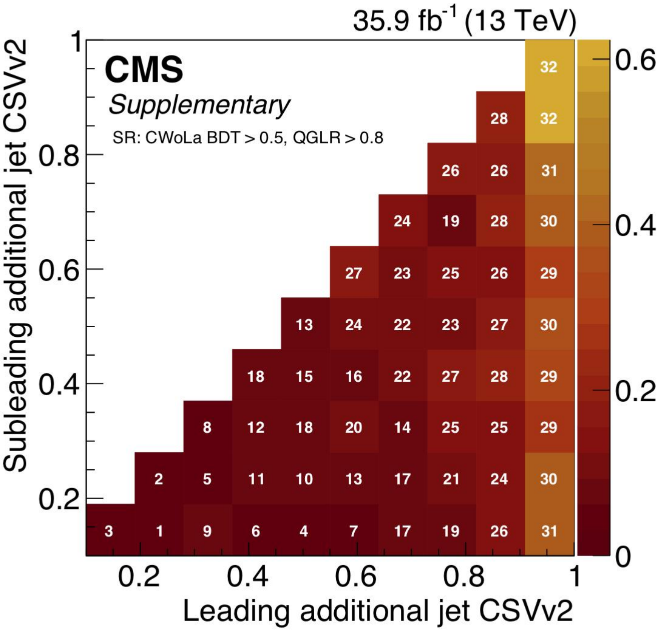
Multijet rejection: Classification Without Labels (CWoLa*)

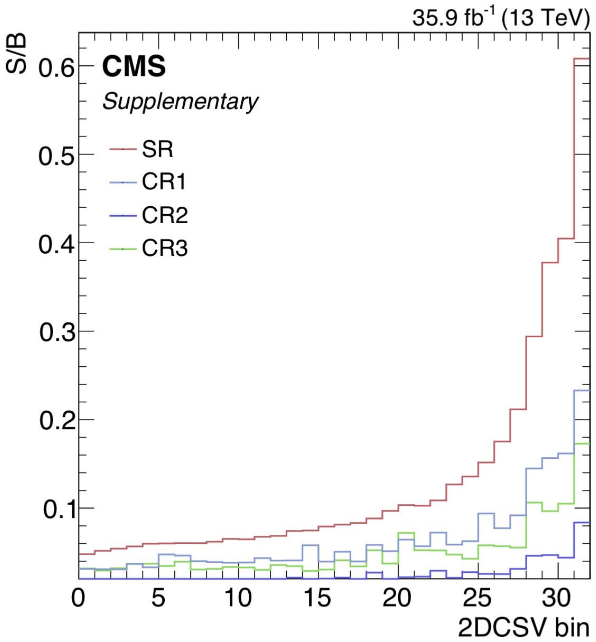


Cut used for the Control regions definition



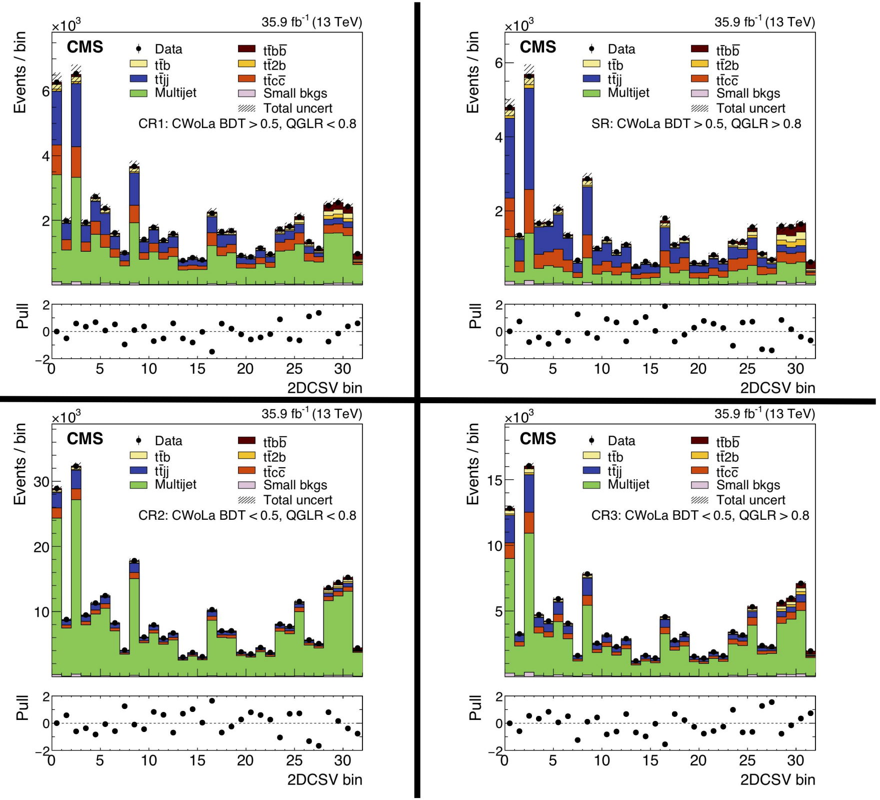
Data driven QCD estimation



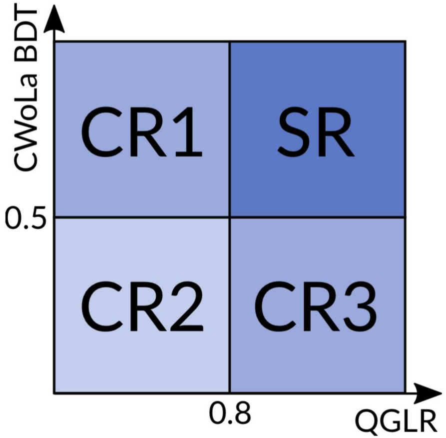


- Bins are **merged** to ensure enough events per region
- Distributions are "**unrolled**" to **1D**
- Bins are **ordered** by **increasing** value of the expected **S/B**
- Define 4 different regions using cuts on the **QGLR** and **CWoLa BDT discriminators**
- **60%** signal on the most significant bin

Data driven QCD estimation



- **ABCD** method used for each bin **simultaneously fitting** the **4 regions** at the same time
- Maximum likelihood estimation
- Sources of **systematic uncertainties** are **profiled** (more in the next slide)
- Assumption: For each bin i,
  - $N^{SR}_i/N^{CR1}_i = N^{CR3}_i/N^{CR2}_i \Rightarrow N^{SR}_i = N^{CR1}_i \times N^{CR3}_i/N^{CR2}_i$

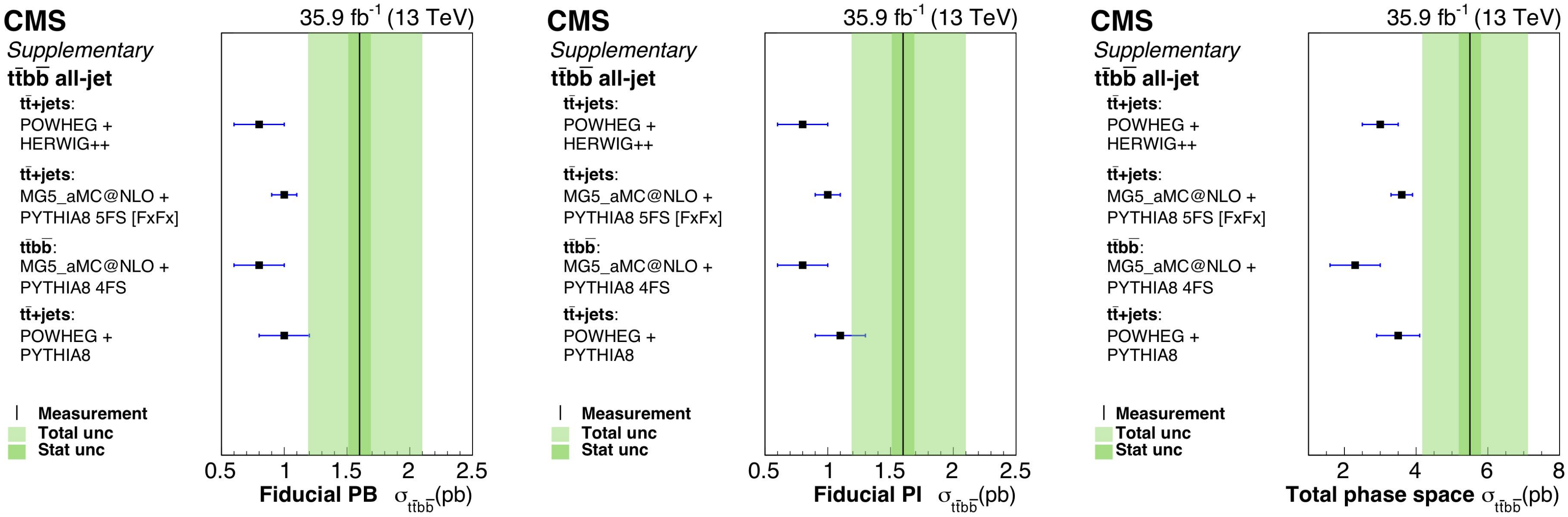


Systematic uncertainties

- **Theory uncertainties:**
  - **PDF, FSR, ISR, UE tune, hdamp, renormalization and factorization scales, Colour reconnection models**
  - Normalization for smaller backgrounds
  - **50%** normalization uncertainty for ttcc
- **Jet energy corrections**
- **Corrections:**
  - **Ecal correction**
  - **QGLR**
  - **Pile up**
  - **B-tagging**
  - top-p_T uncertainty
  - Trigger
- **Luminosity** (2.5% for all processes), **Jet energy resolution**
- **Multijet contribution and uncertainty** determined by the fit
- Limited MC statistics

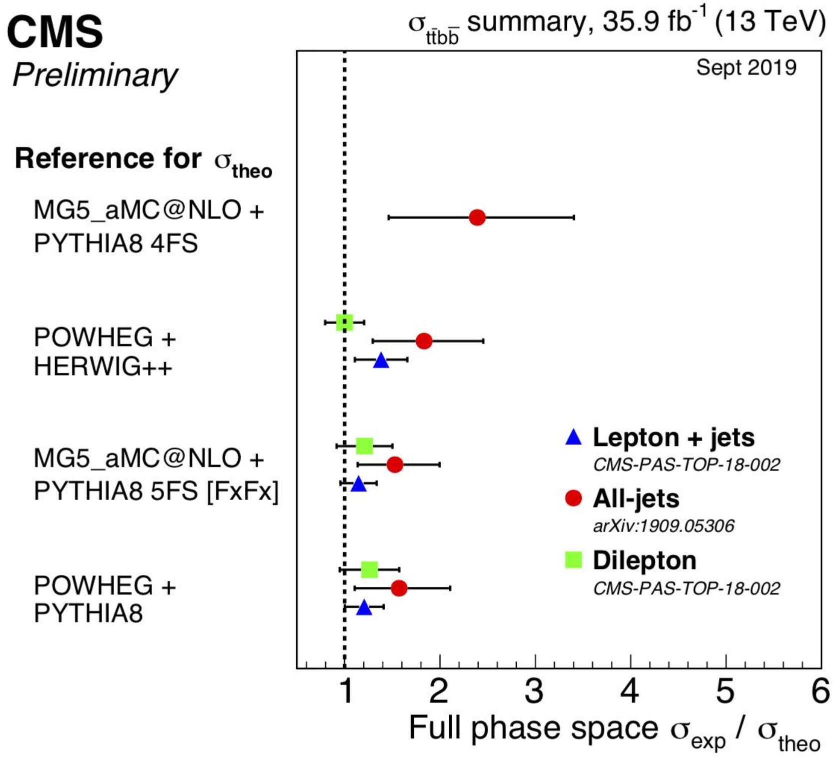
Source	FPS PI (%)	FPS PB (%)
Simulated sample size	$^{+15}_{-11}$	$^{+15}_{-11}$
Quark-gluon likelihood	$^{+13}_{-8}$	$^{+13}_{-8}$
b tagging of b quark	± 10	± 10
JES and JER	$^{+5.1}_{-5.2}$	$^{+5.0}_{-5.4}$
Integrated luminosity	$^{+2.8}_{-2.2}$	$^{+2.4}_{-2.2}$
Trigger efficiency	$^{+2.6}_{-2.1}$	$^{+2.5}_{-2.2}$
Pileup	$^{+2.3}_{-2.0}$	$^{+2.2}_{-1.9}$
μ_R and μ_F scales	$^{+13}_{-9}$	$^{+13}_{-9}$
Parton shower scale	$^{+11}_{-8}$	$^{+11}_{-8}$
UE tune	$^{+9.0}_{-5.3}$	$^{+9.0}_{-5.2}$
Colour reconnection	± 7.2	± 7.1
Shower matching (h_{damp})	$^{+4.3}_{-2.8}$	$^{+3.8}_{-2.7}$
$t\bar{t}c\bar{c}$ normalization	$^{+3.2}_{-4.4}$	$^{+2.9}_{-4.5}$
Modelling of p_T of top quark	± 2.5	± 2.4
PDFs	$^{+2.2}_{-2.0}$	$^{+2.2}_{-2.0}$
Total	$^{+28}_{-23}$	$^{+28}_{-23}$

Results



- Observe a **larger cross section** value compared to predictions
- **Theory uncertainties:** PDF, renormalization and factorization scales

Conclusions



- Novel methods for **QCD** and **Combinatorial** backgrounds rejection developed
- **First time** the measurement of the ttbb cross section in the all-jet channel was performed
- Sensitivity of: **26%** for **fiducial definitions** and **27%** for the **full phase-space**
- **Main uncertainties:**
  - **Experimental:** MC sample size, b-tagging and QGL reweighting
  - **Theory:** Parton shower, renormalization and factorization scales

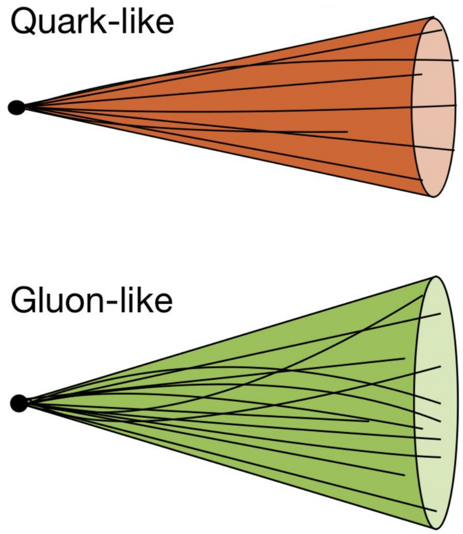
THANKS!

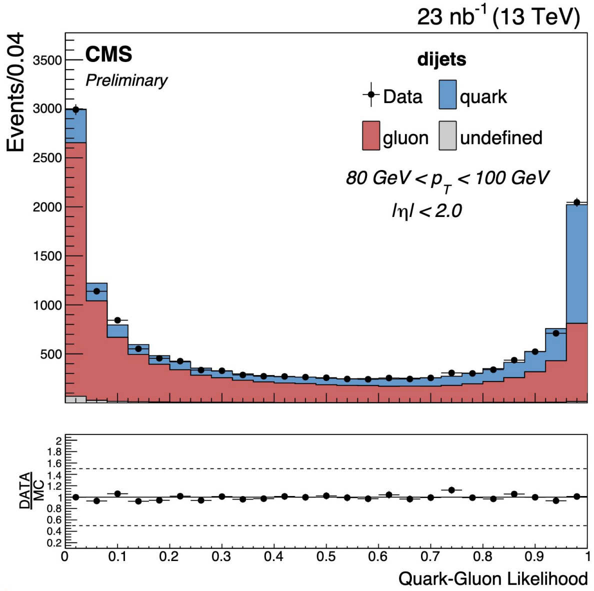
Any questions?

BACKUP

QGL

- Discrimination between quark and gluon initiated jets: **Quark-Gluon Likelihood**
- Uses:
  - ▻ **Number of constituents**
  - ▻ **Spatial collimation**
- Optimized to distinguish **light flavour quark jets** from **gluon jets**





$$q_{\mathrm{LR}} = \frac{\mathcal{L}_{qqqqq}}{\mathcal{L}_{qqqqq} + \mathcal{L}_{ggggg}}$$

$$\mathcal{L}_{qqqqq} \rightarrow f_q(\zeta_{j_1}) \cdot f_q(\zeta_{j_2}) \cdot f_q(\zeta_{j_3}) \cdot f_q(\zeta_{j_4}) \cdot f_q(\zeta_{j_5})$$

$$\mathcal{L}_{ggggg} \rightarrow f_g(\zeta_{j_1}) \cdot f_g(\zeta_{j_2}) \cdot f_g(\zeta_{j_3}) \cdot f_g(\zeta_{j_4}) \cdot f_g(\zeta_{j_5})$$

ζ = QGL value

$f_{q/g}(\zeta)$ = QGL distribution of quarks/gluons

Yields

	SR	CR1	VR	CR2
Multijet	12548.6±95.1	37073.3±280.9	89119.8±204.0	250370.2±573.0
ttlf	18952.0±126.0	13885.4±90.0	20440.2±137.2	21128.8±111.8
ttcc	8023.2±79.0	6680.6±61.5	9891.7±89.0	11494.3±80.6
ttV	325.9±8.1	174.0±5.6	412.4±9.1	289.2±7.3
Single top	460.9±28.4	366.5±20.0	1006.5±41.2	1110.7±36.0
VJ	249.7±53.3	288.9±43.7	988.7±92.5	1279.4±92.8
ttH	161.1±0.8	136.1±0.7	195.8±0.9	221.1±0.8
Diboson	1.5±0.7	3.3±1.2	27.2±6.4	17.2±5.3
ttb	2719.8±46.6	2207.6±35.8	3665.1±54.8	4118.6±48.7
ttbb	1641.1±36.4	1365.1±28.3	2426.5±43.2	2679.5±39.1
Data	45084	62181	128174	292709