Background estimation for searches for supersymmetry in the single lepton final state in 13 TeV pp collisions

David Handl

INSTITUT FÜR HOCHENERGIEPHYSIK

IMPRS Workshop, Max-Planck-Institute for Physics

March 14th 2016

Supersymmetry (SUSY)

Symmetry between **bosons** and **fermions**
Supersymmetric particle is assigned to each particle in the SM
Quantum numbers remain unchanged, **spin differs** by $\pm 1/2$

Chiral supermultiplet

- leptons and quarks have spin-0 super partners (sleptons and squarks)
- 2 chiral Higgs doublets are necessary, partners are the higgsinos

Vector supermultiplet

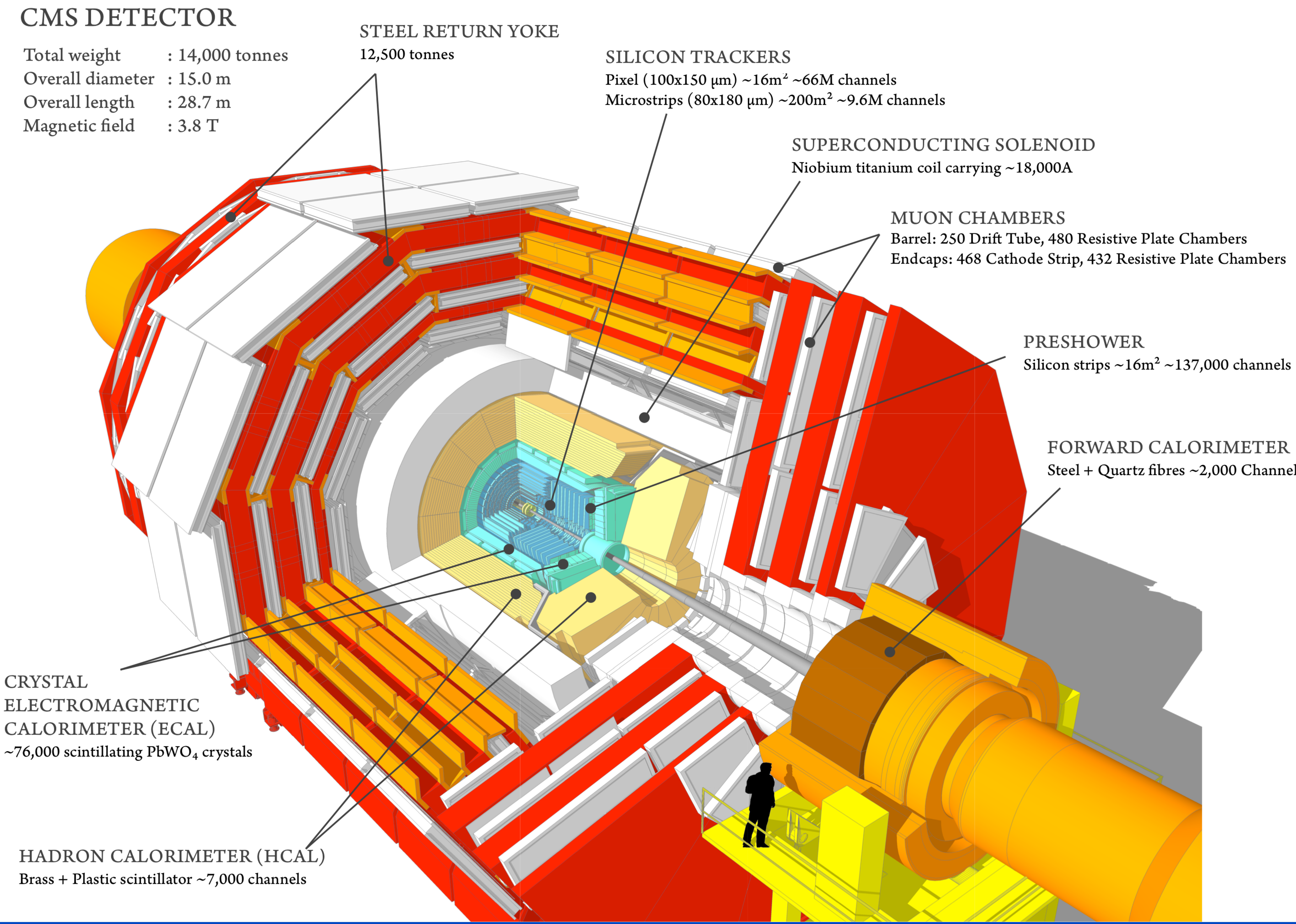
- SM gauge bosons have spin-$1/2$ superpartners, called gauginos
- charged (neutral) higgsinos, winos and binos can mix to charged (neutral) mass eigenstates, called charginos (neutralinos)

SUSY particle masses ≠ SM particle masses → **broken symmetry**

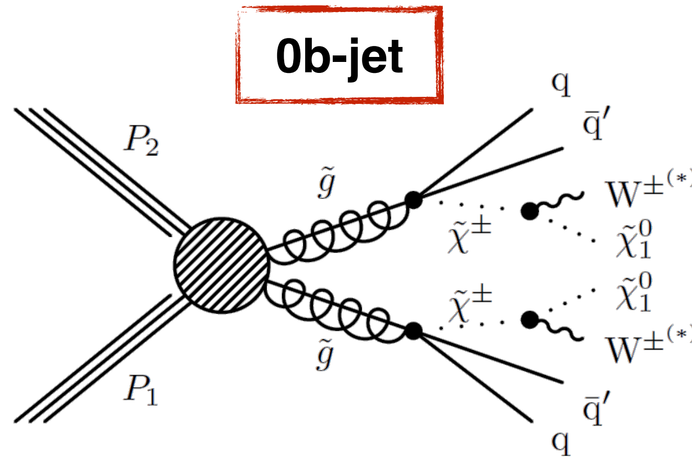
name	spin 0	spin 1/2
squarks, quarks (× 3 families)	$(\tilde{u}_L\ \tilde{d}_L)$ $\tilde{u}_R^*$ $\tilde{d}_R^*$	$(u_L\ d_L)$ $u_R^\dagger$ $d_R^\dagger$
sleptons, leptons (× 3 families)	$(\tilde{\nu}\ \tilde{e}_L)$ $\tilde{e}_R^*$	$(\nu\ e_L)$ $e_R^\dagger$
Higgs, higgsinos	$(H_u^+\ H_u^0)$ $(H_d^0\ H_d^-)$	$(\tilde{H}_u^+\ \tilde{H}_u^0)$ $(\tilde{H}_d^0\ \tilde{H}_d^-)$

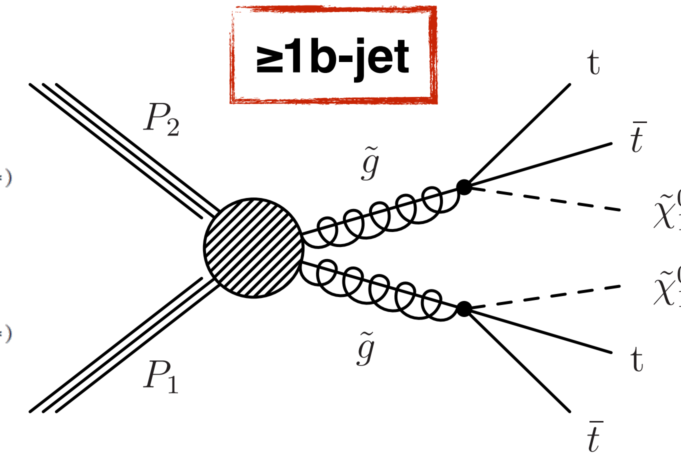
name	spin 1/2	spin 1
gluino, gluon	$\tilde{g}$	g
winos, W bosons	$\tilde{W}^\pm\ \tilde{W}^0$	$W^\pm\ W^0$
binos, B bosons	$\tilde{B}^0$	B^0

Compact Muon Solenoid (CMS)

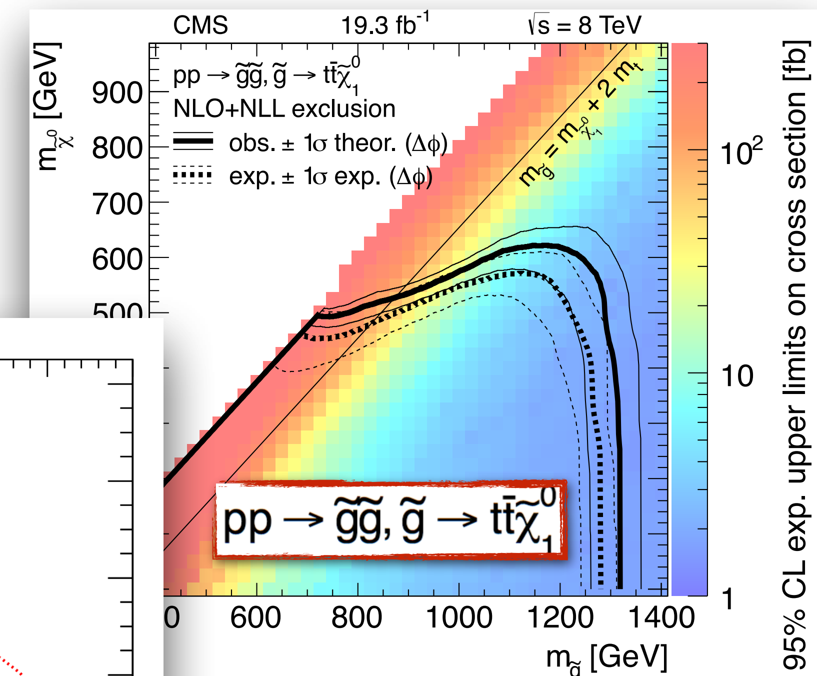


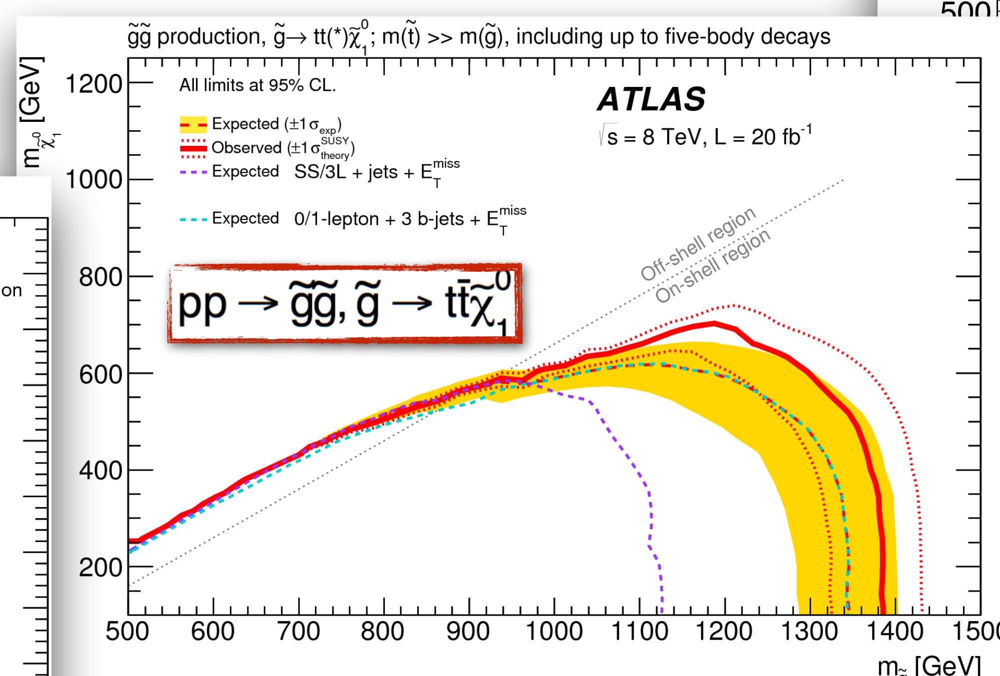
Search for SUSY in events with one electron or muon

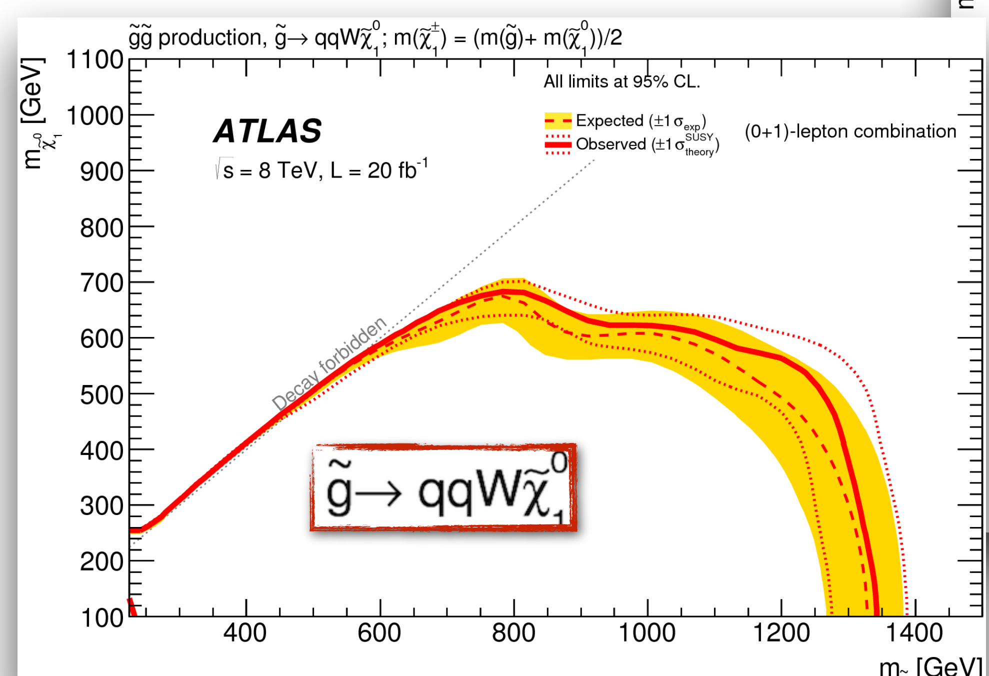




- joint effort of data analysis groups from Athens, CERN, DESY Hamburg and HEPHY Vienna
- consider **simplified models**
- **separated** by the number of **bottom quarks**
- search focuses on **gluino pair production**
- **First analysis** in **CMS** of $\tilde{g}\tilde{g} \rightarrow q\bar{q}q\bar{q}WW\tilde{\chi}^0\tilde{\chi}^0$ model
- CMS results from 8 TeV data exclude gluino masses < **1.26 TeV** [1]
- ATLAS performed searches for both models [2] on 8 TeV data
- masses up to **1.2 - 1.31 TeV** are excluded

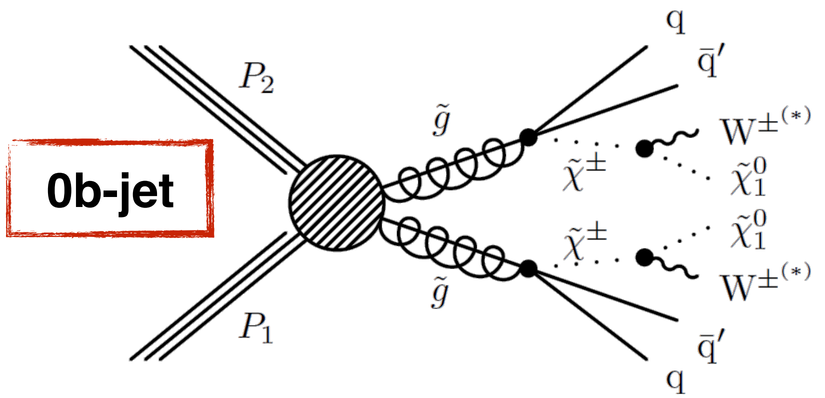




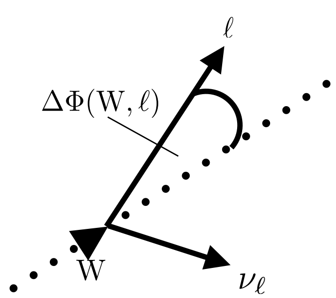


Analysis strategy

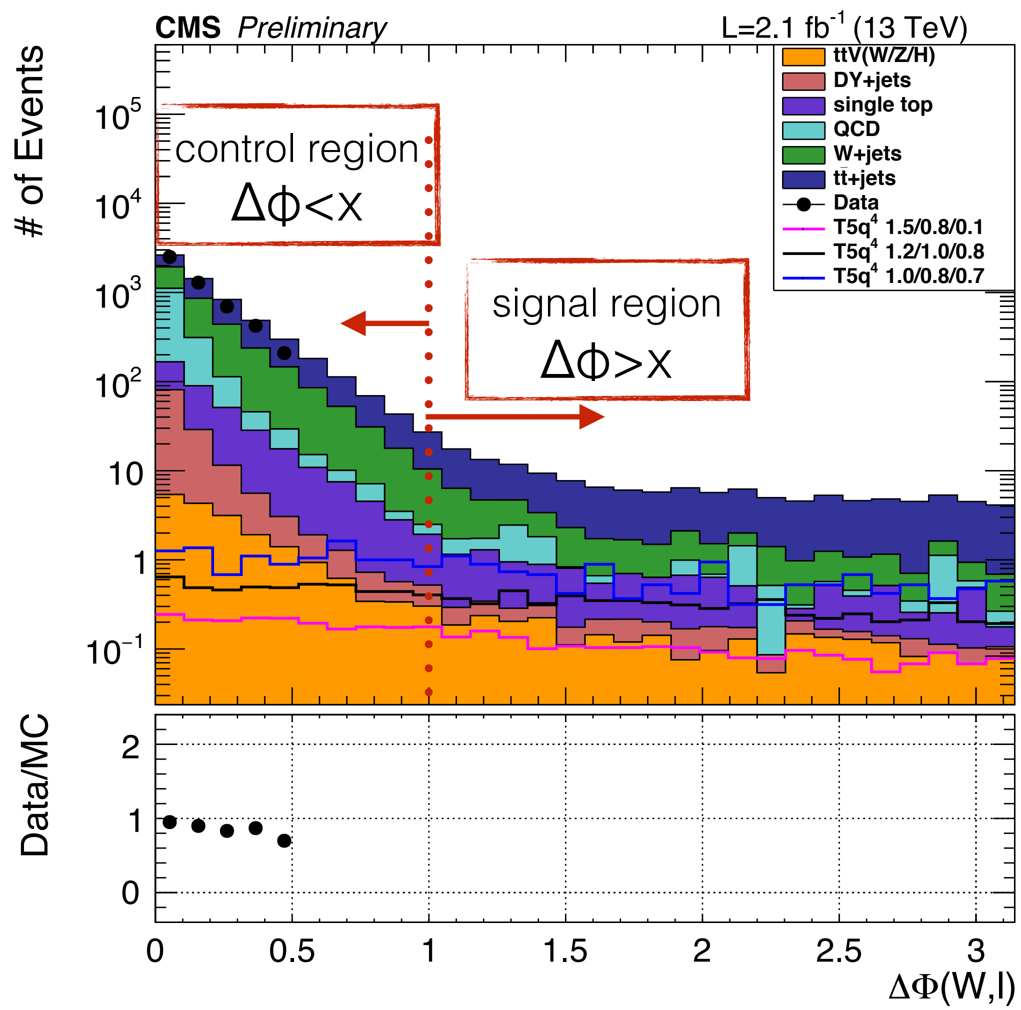
- **Vienna** performed the SUSY search in **0b-jet channel**
- search focuses on **gluino pair production** and **decay** through **chargino**
    - large number of jets, n_{jets}
    - high hadronic activity, $H_T = \Sigma p_T(\text{jets})$
    - high momentum imbalance, $L_T = p_T(\text{lepton}) + E_T^{miss}$



- search variable: **azimuthal angle between W and lepton**



- events with a genuine W→lν decay have a maximum
- SUSY signal has a uniform distribution because of two neutralinos



- expected signal yields after event selection:

	$T5q^4 1/0.8/0.7$	$T5q^4 1.2/1.0/0.8$	$T5q^4 1.5/0.8/0.1$	All BKG
$\Delta\Phi > x$	18.3	9.2	2.9	44

- search performed on **2.1fb^{-1} dataset** recorded in 2015
- data was blinded during the completion of my thesis

Data-driven background estimation

Dominant backgrounds are **tt+jets** and **W+jets**

- estimation is based on the **R_{CS} transfer factor:**

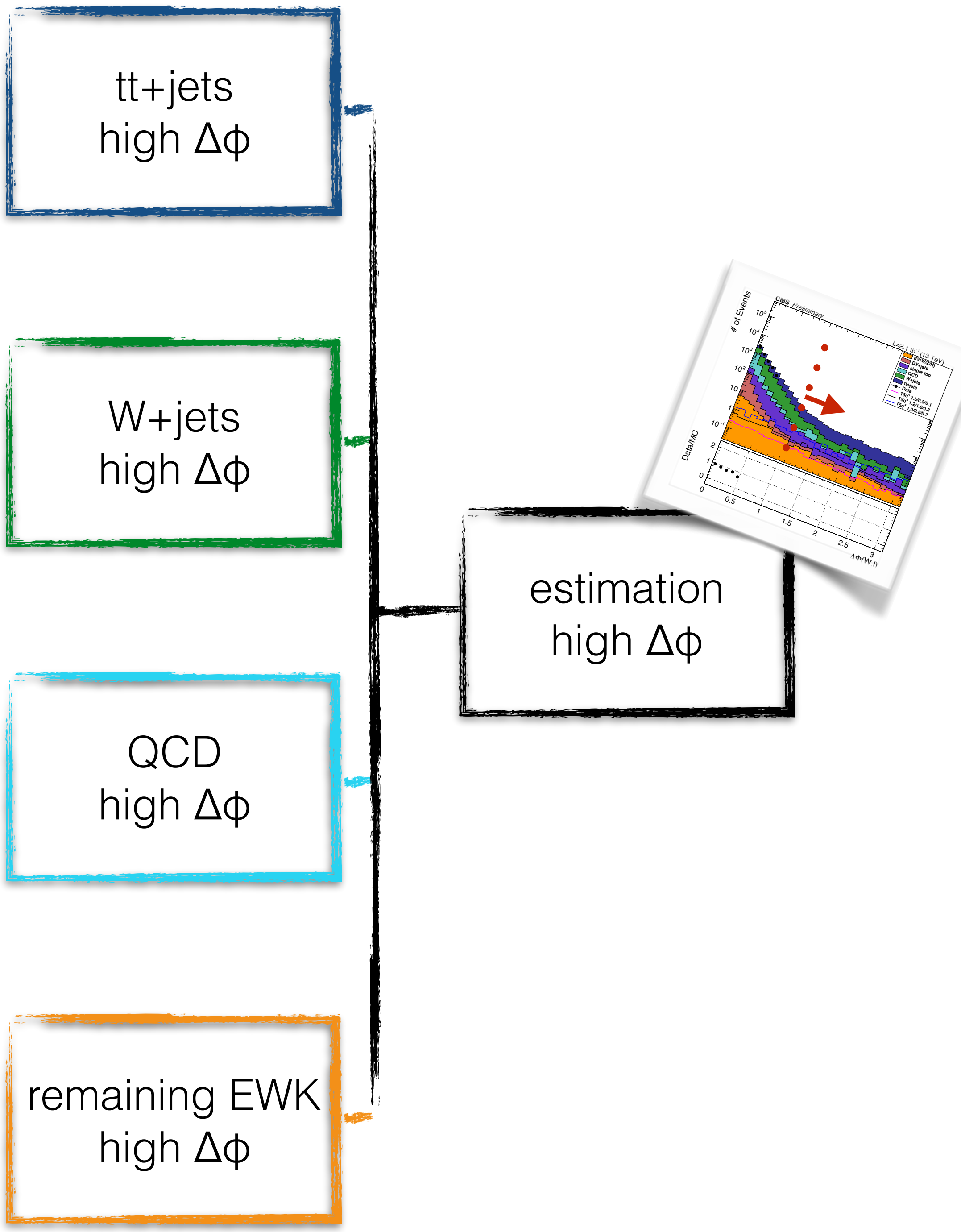
$$R_{\mathrm{CS}} = \frac{N_{\mathrm{SR}}}{N_{\mathrm{CR}}} = \frac{\text{Number of events with } \Delta\Phi(\mathrm{W},\ell)>\mathrm{x}}{\text{Number of events with } \Delta\Phi(\mathrm{W},\ell)<\mathrm{x}}$$

- determine **R_{CS}** in **data** in a **low jet sideband** region
- R_{CS} is almost **independent** of jet multiplicity
- **scale** the event yield from **low Δφ region** to **signal region**

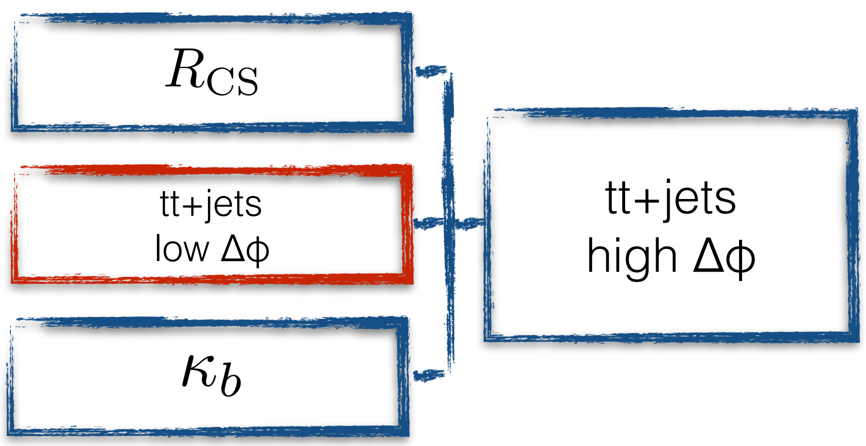
QCD multijet background is estimated with a different approach

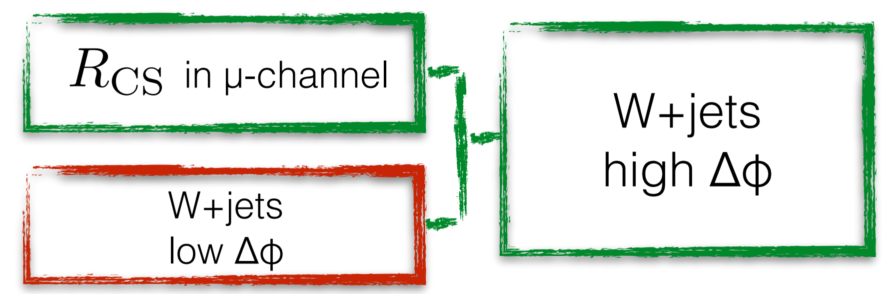
- separately estimated in electron and muon channel
- **subtraction** in low jet **sidebands**

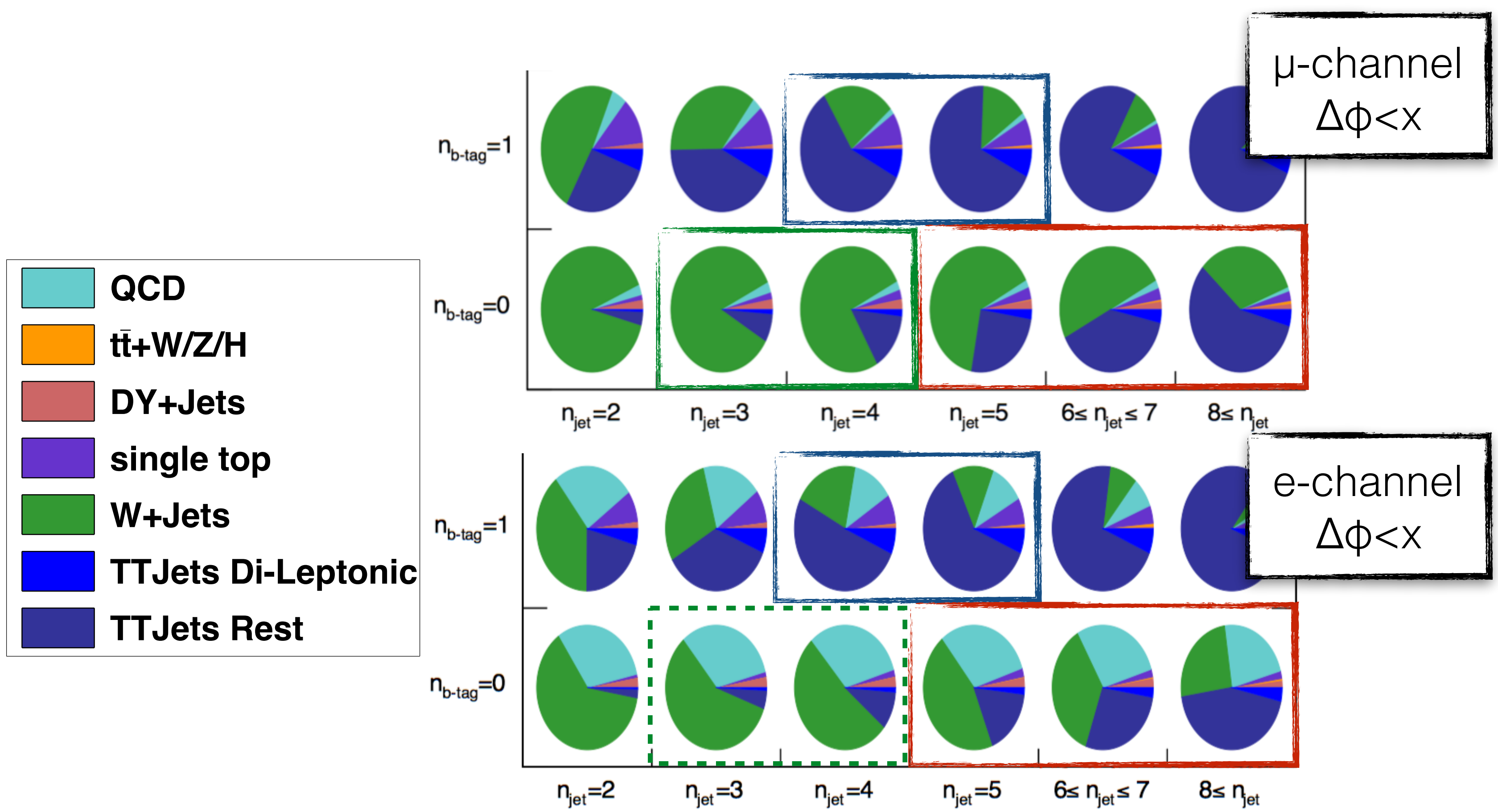
Remaining EWK backgrounds are taken from **MC** simulations



Prediction of the tt+jets and W+jets background



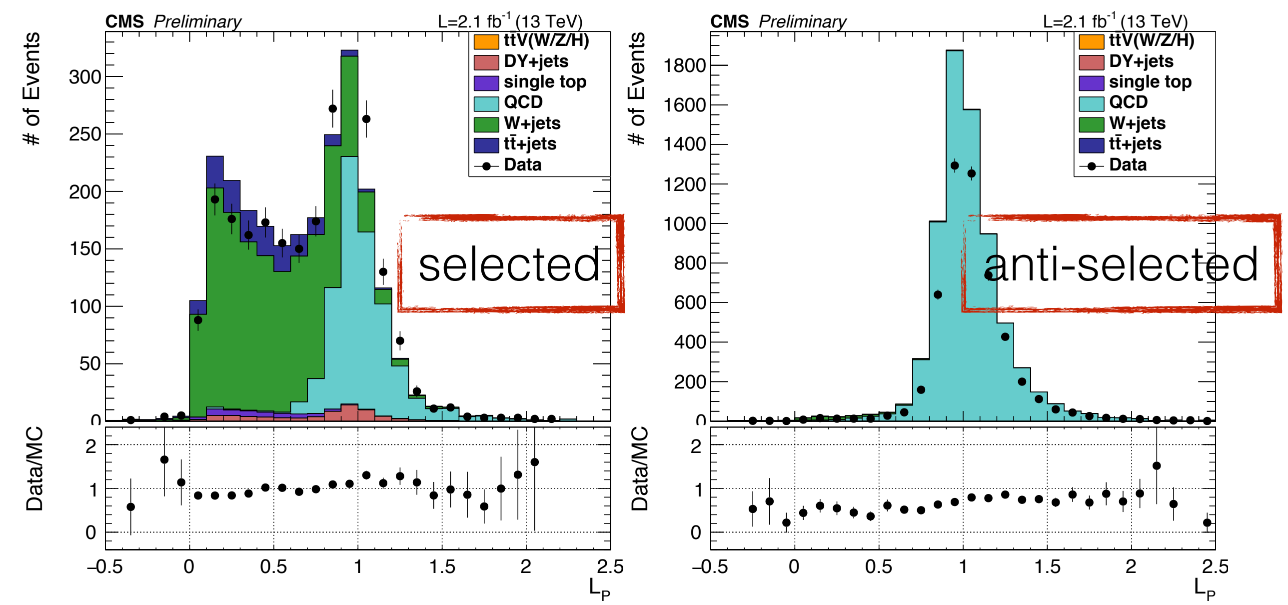


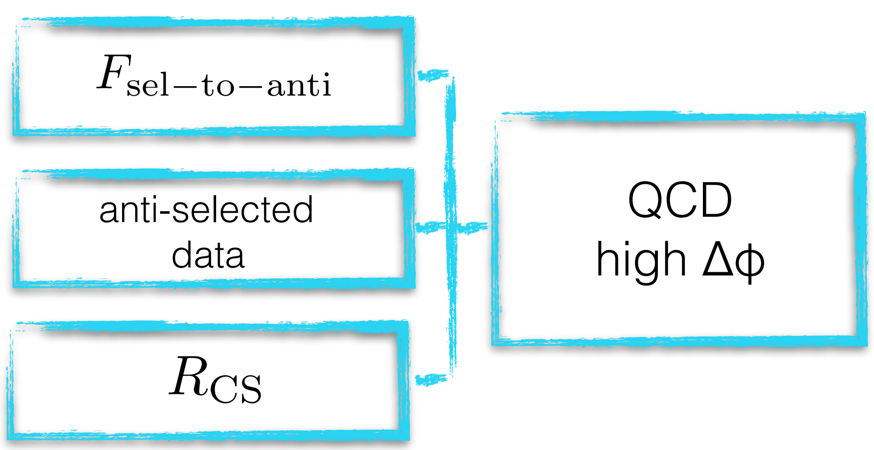


QCD multijet background estimation

- estimated in the **electron channel**
- data-driven method based on L_P variable
- L_P reflects the **W boson polarization**

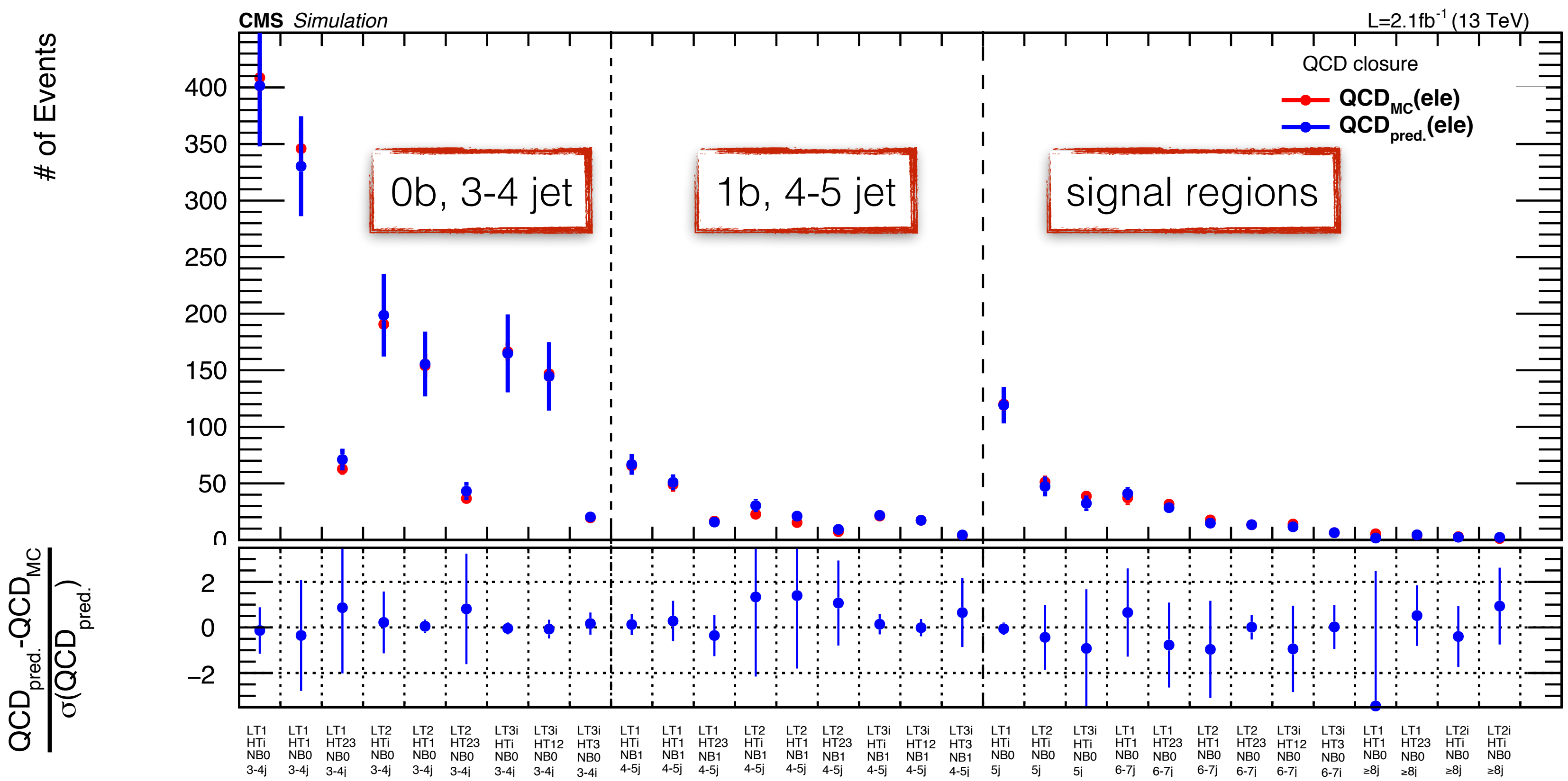
- **invert** the electron **identification** requirements
- **no impact** on L_P distribution

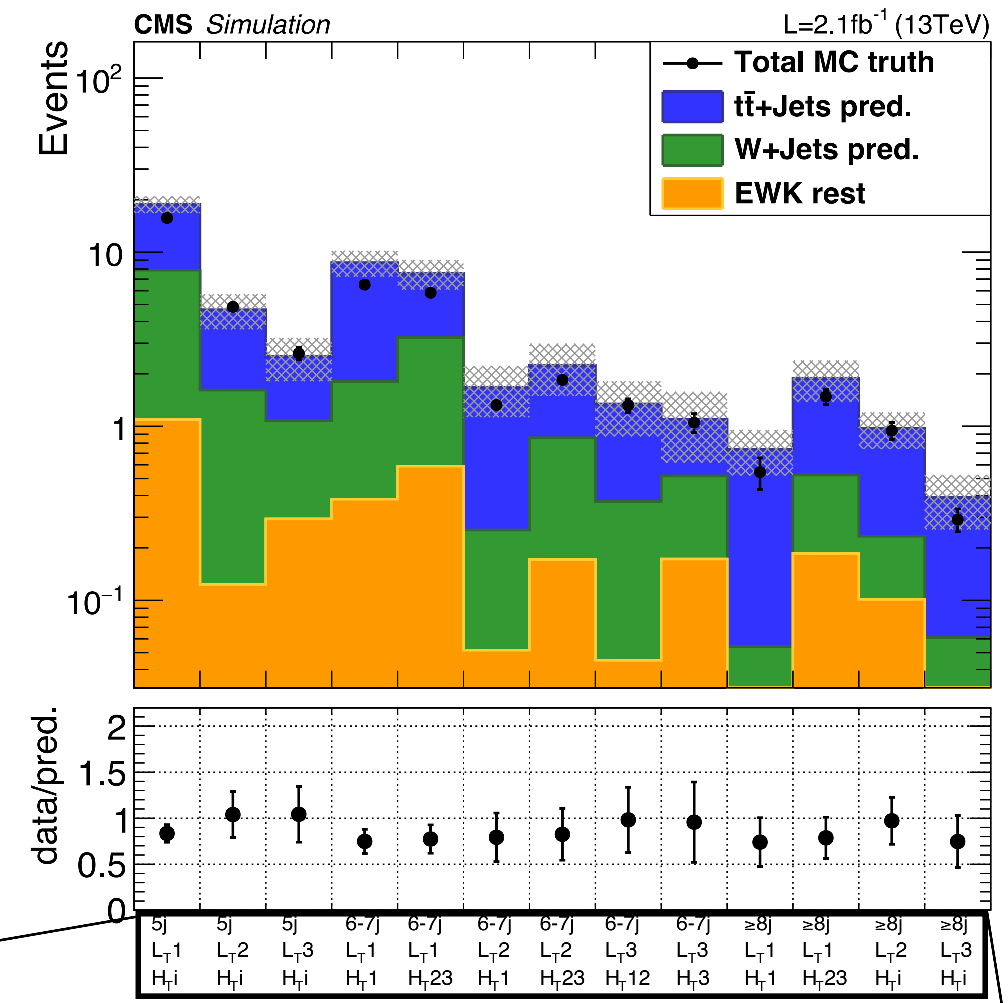




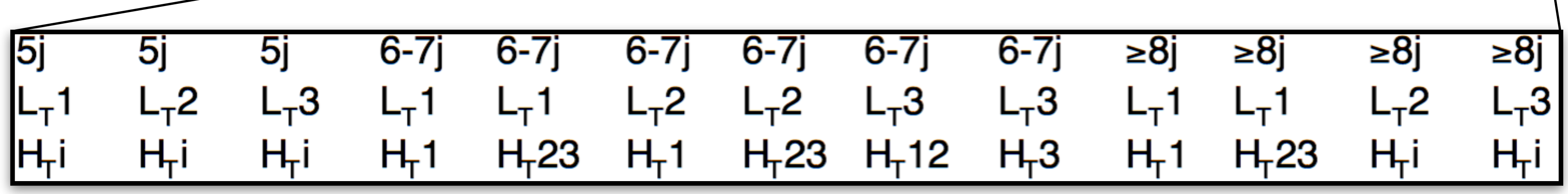
- ratio of **'selected'** to **'anti-selected'** events in 3-4 jet, 0b-jets sideband
- $\mathbf{F_{sel\text{-}to\text{-}antisel}}$ almost independent of the jet multiplicity
- R_{CS} **well below 1%**
- QCD mulitjet background **negligible** in **signal regions**

Closure test



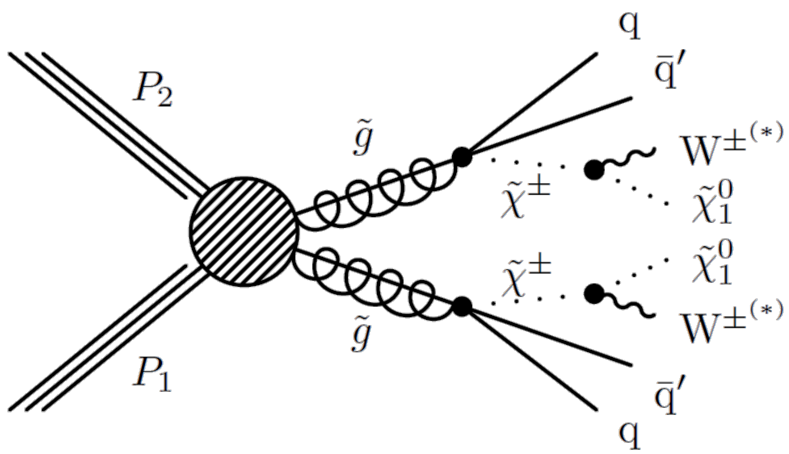


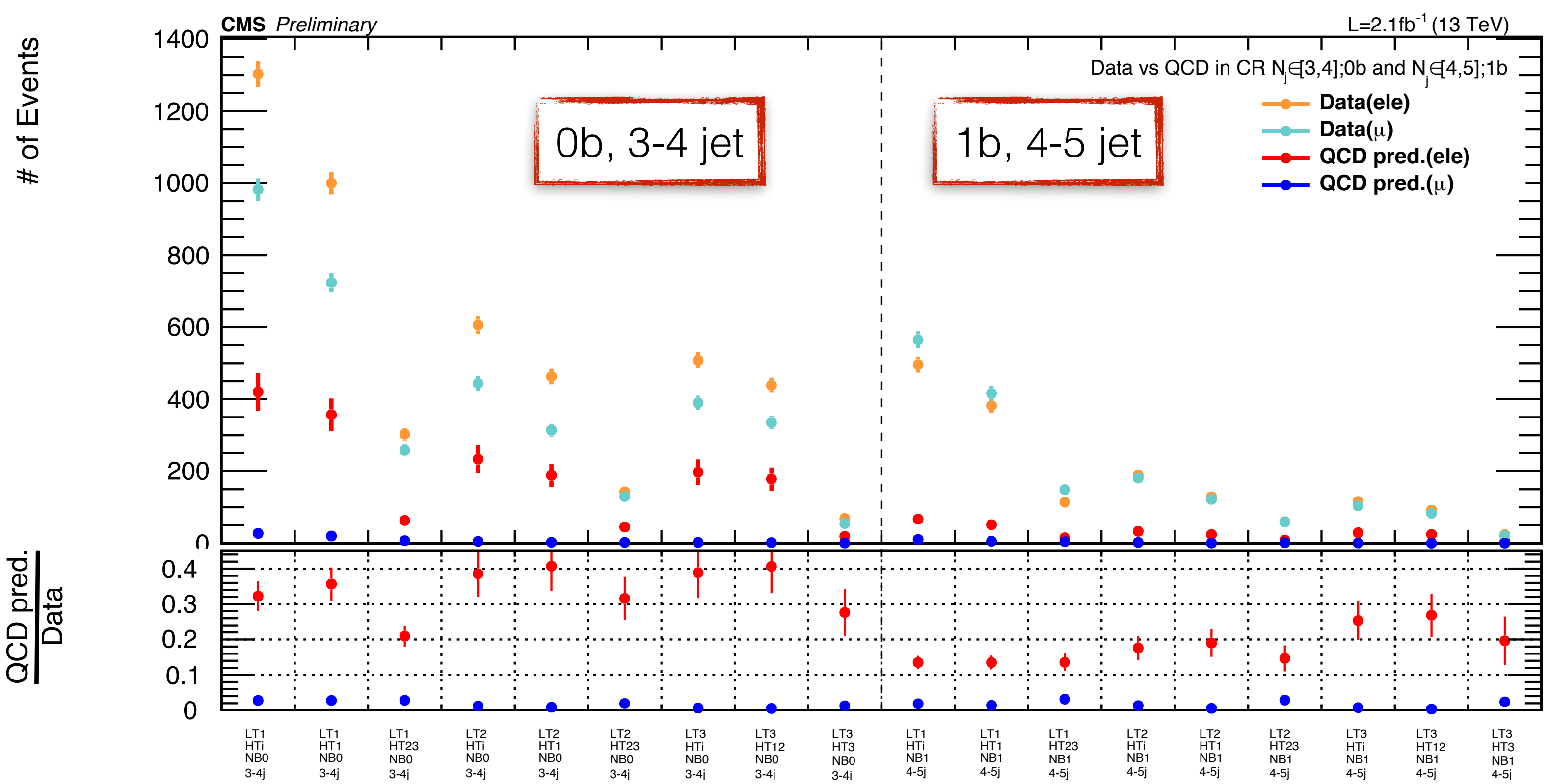
- combined EWK and QCD MC samples are used as a data template
- closure test performed in the two sidebands and the high jet multiplicity search regions
- good agreement between prediction and expectation
- derived results are reliable

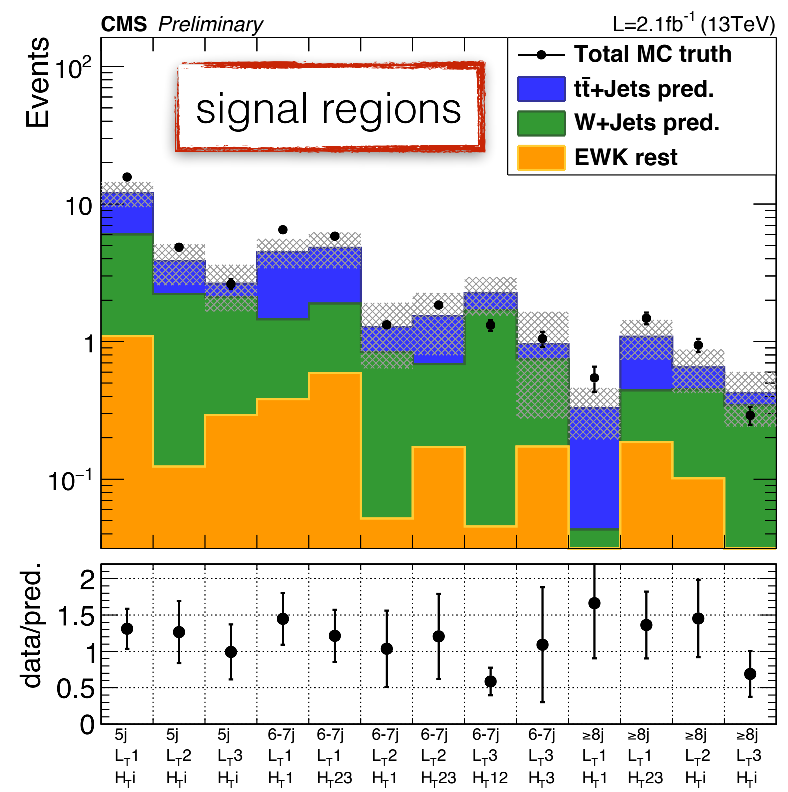


Summary and conclusion

- First analysis in CMS, focussing on gluino pair production and decay through chargino
- data-driven methods to estimate tt+jets, W+jets and multijet backgrounds
- reliable results are obtained
- analysis was still under study
  - data in signal regions is concealed
  - comparison of the background prediction and data is omitted







Thank you for your attention!

Acknowledgements
Prof. Jochen Schieck, Dr. Robert Schöfbeck (Supervisors)
Dr. Wolfgang Adam (Group leader)
& all members of the Institute

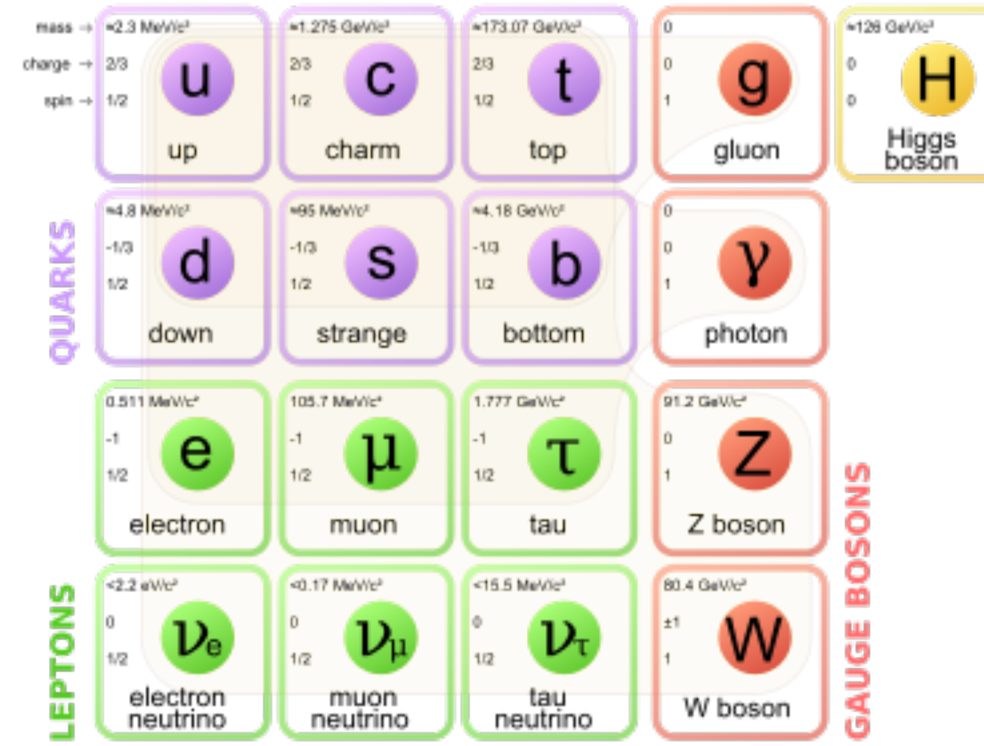
Backup

Motivation

Standard Model of particle physics (SM)

- **matter** is made of elementary particles, known as the **fermions**
- fermions have **spin ½**, separated into **leptons** and **quarks**
- three fundamental forces: electro-weak and strong nuclear force
- **interactions** are mediated by the **exchange of bosons**
- the **Higgs mechanism** describes how **particles acquire mass**
- the **discovery** of the **Higgs boson** in **2012** completed the SM
- theoretical **predictions** are **verified** by various **experiments**

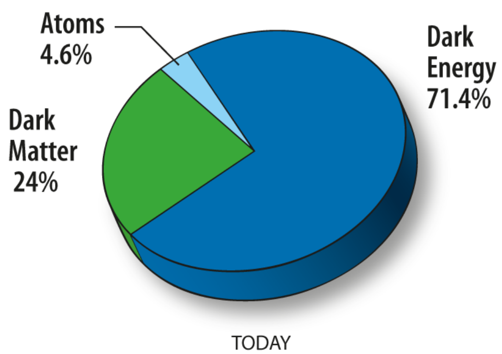
particle content of the SM



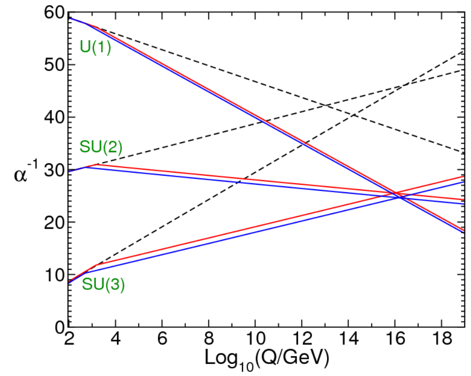
Open questions?

Gravity

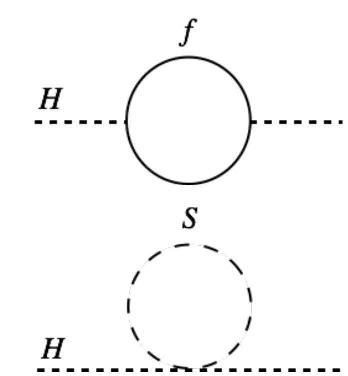
Dark matter



Grand unification



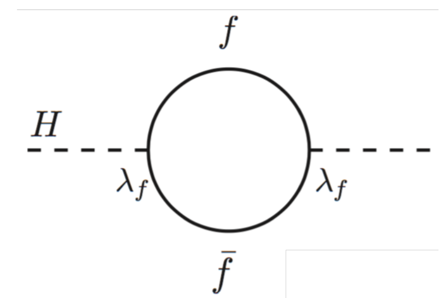
Hierarchy problem



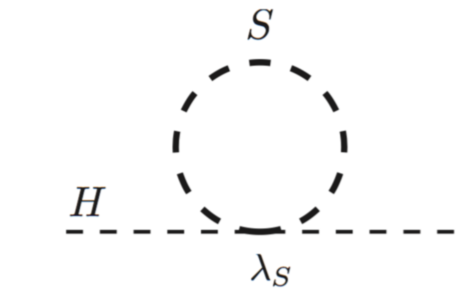
Hierarchy problem

Hierarchy problem

- Particles that couple to the Higgs field induce large quantum corrections to the Higgs boson mass
- Quantum corrections are some orders larger than the mass of the Higgs boson itself



$$\Delta m_{\mathrm{H}}^2 = -\frac{\lambda_f^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \dots$$



$$\Delta m_{\mathrm{H}}^2 = \frac{\lambda_s}{16\pi^2}\Lambda_{\mathrm{UV}}^2 - \dots$$

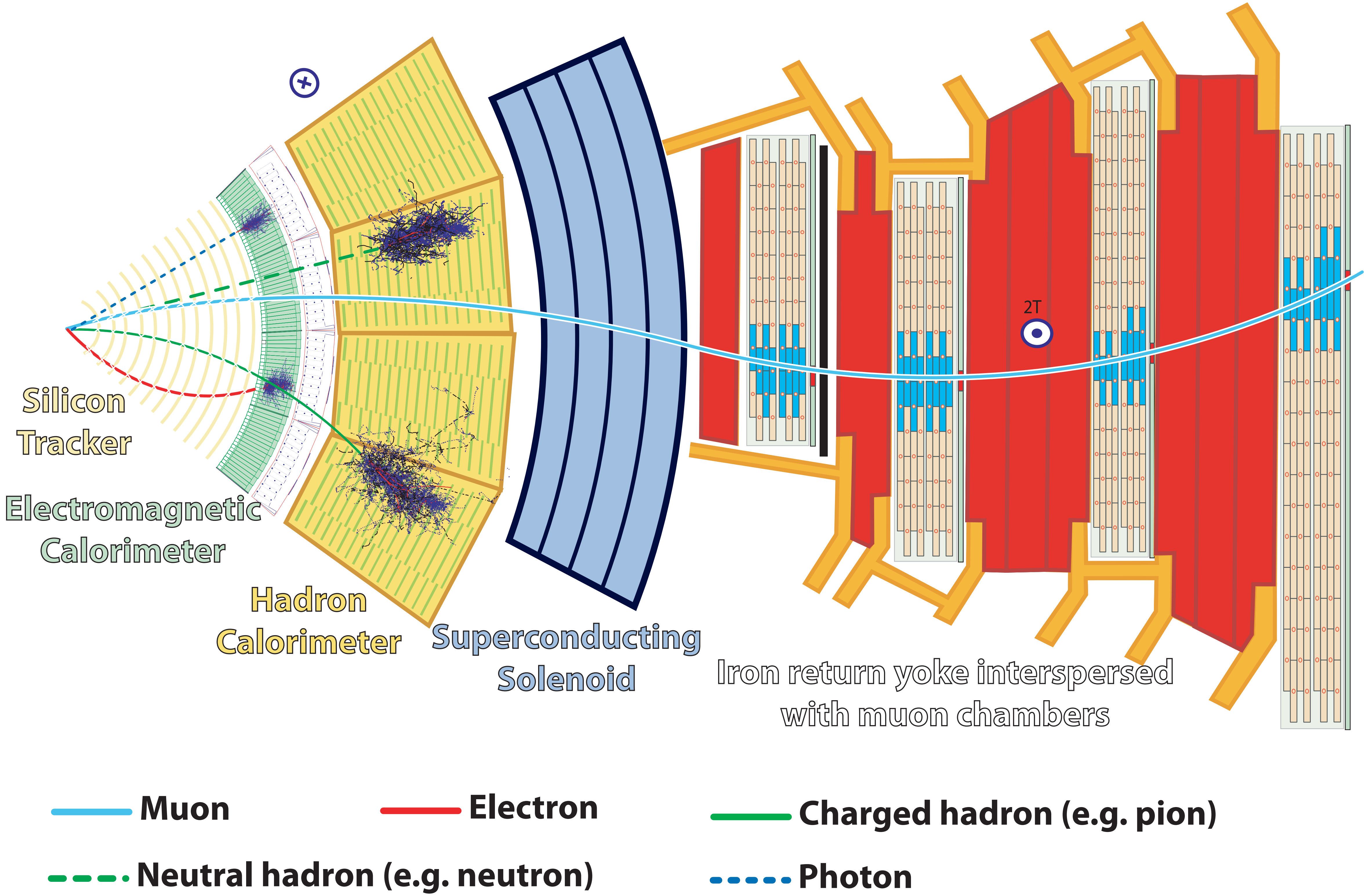
Supersymmetry can **solve** the **hierarchy problem** in an elegant way

$$Q\,|\mathrm{Boson}\rangle = |\mathrm{Fermion}\rangle \qquad Q\,|\mathrm{Fermion}\rangle = |\mathrm{Boson}\rangle$$

Fermionic operator Q must satisfy **anti-commutation** relations

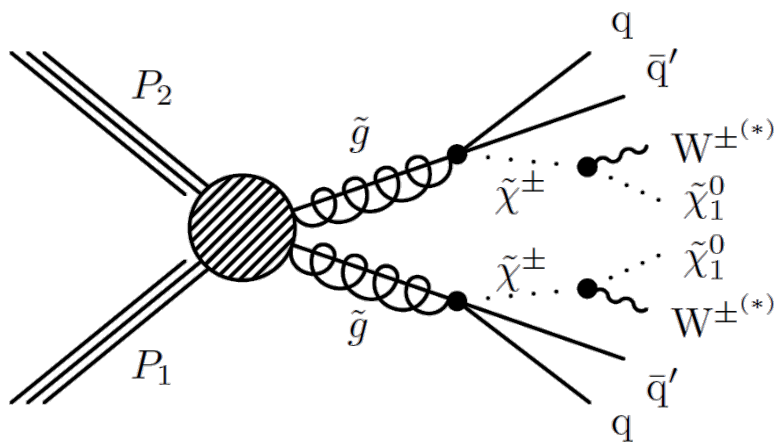
SUSY algebra **extend** usual **space-time** to **super-space**

Cross section of the CMS detector

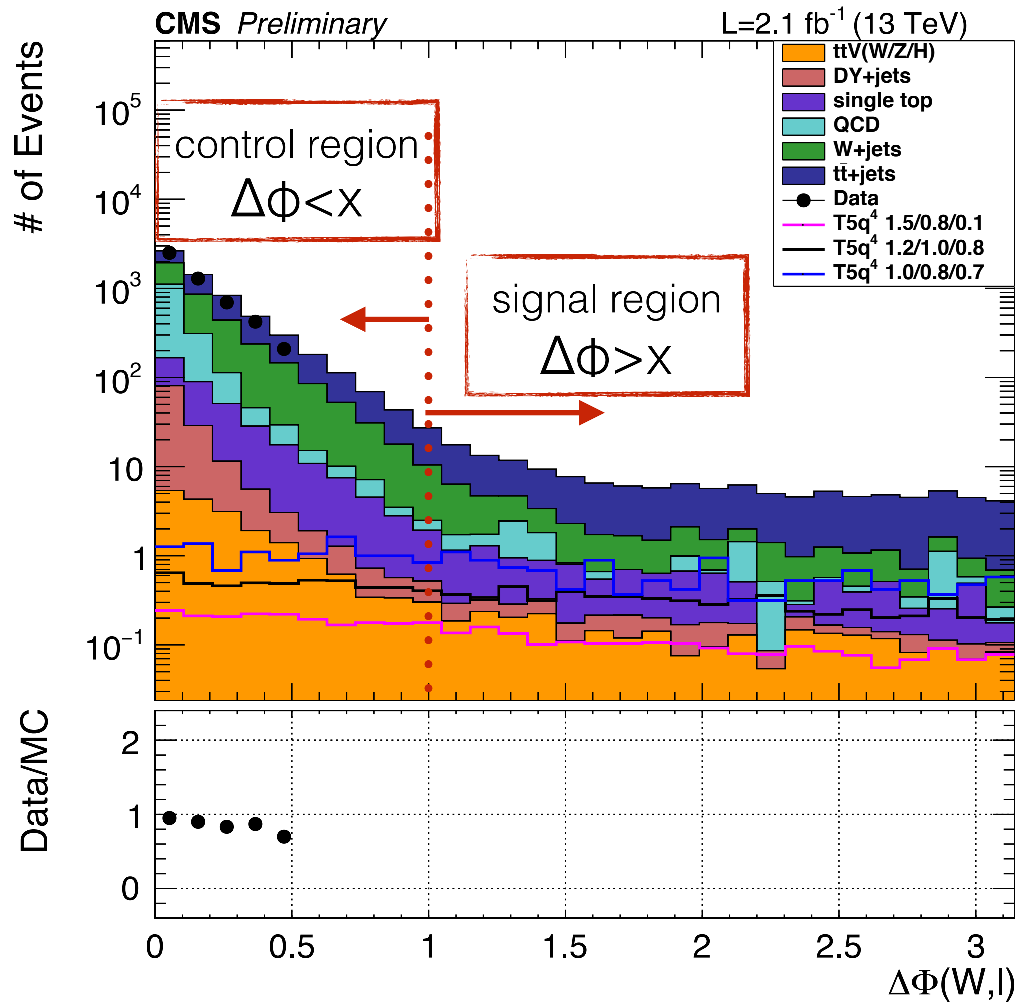


Event selection

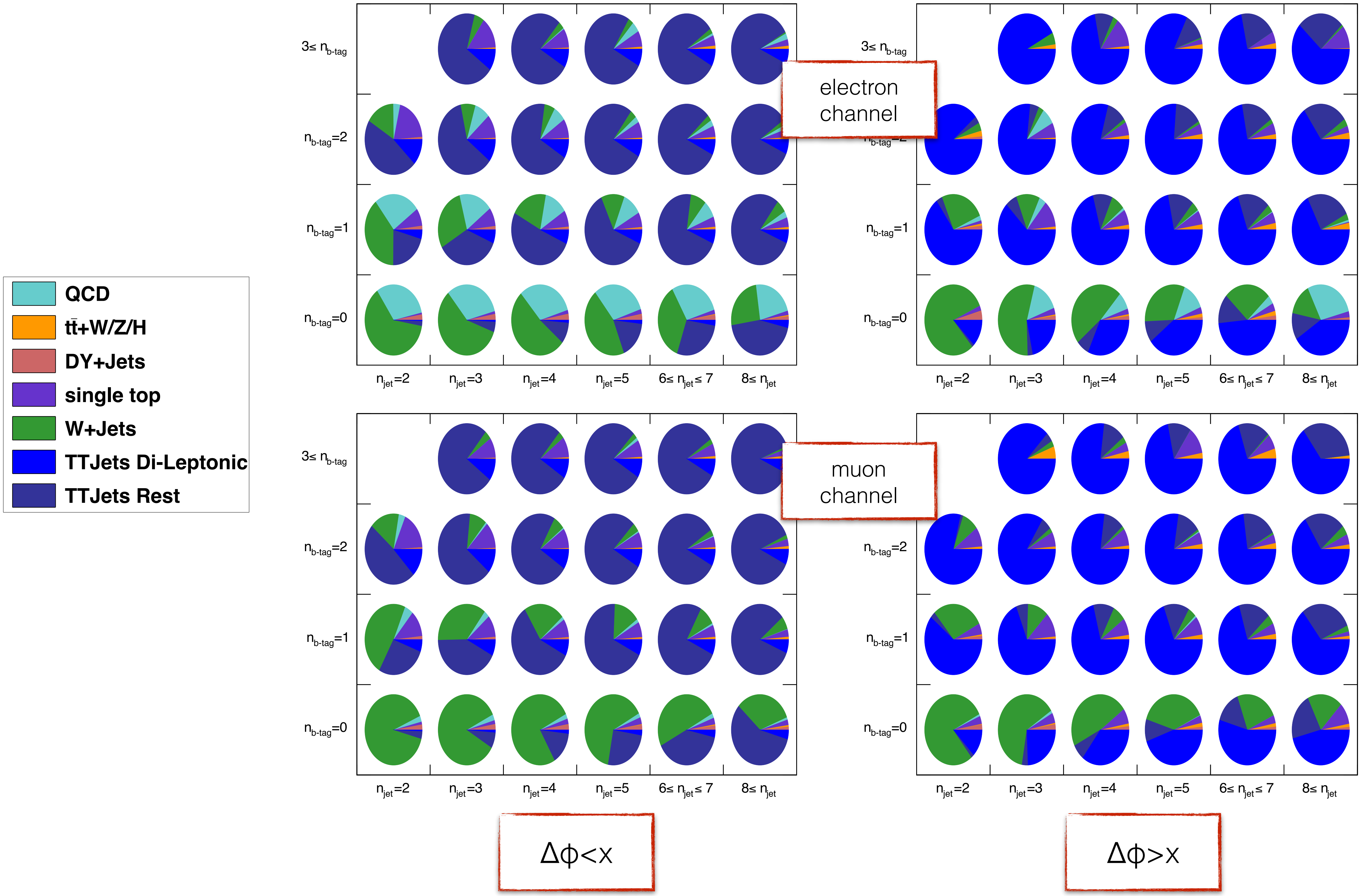
- exactly one electron or muon, p_T > 25 GeV
- veto events with additional soft leptons (p_T > 10 GeV)
- at least five jets (p_T > 30 GeV)
- two jets should have at least p_T > 80 GeV
- no jet tagged as a bottom quark



n_{jet}	L_T [GeV]	H_T [GeV]	$\Delta\Phi(W,\ell)$
5	[250, 350]	≥ 500	1.0
	[350, 450]	≥ 500	1.0
	≥ 450	≥ 500	1.0
[6, 7]	[250, 350]	[500, 750]	1.0
		≥ 750	1.0
	[350, 450]	[500, 750]	1.0
		≥ 750	1.0
	≥ 450	[500, 1000]	0.75
		≥ 1000	0.75
≥ 8	[250, 350]	[500, 750]	1.0
		≥ 750	1.0
	[350, 450]	≥ 500	0.75
	≥ 450	≥ 500	0.75



Background composition



Prediction of the tt+jets background (contd.)

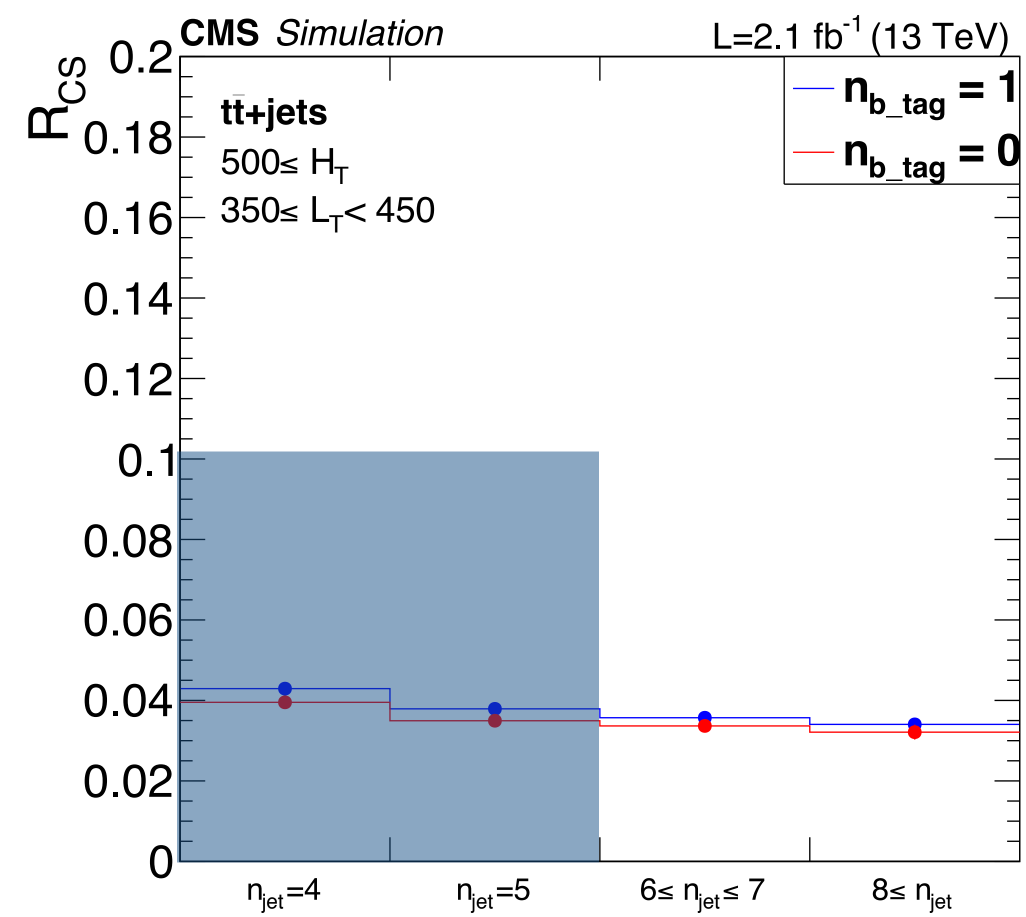
	$n_{\text{b-jet}} = 0$	$n_{\text{b-jet}} = 1$
$n_{\text{jet}} = 3$	Bkg. estimation of W+jets and QCD multijets	
$n_{\text{jet}} = 4$		Bkg. estimation of $t\bar{t}+$jets
$n_{\text{jet}} = 5$	Search regions	
$6 \geq n_{\text{jet}} \geq 7$		
$n_{\text{jet}} \geq 8$		

- R_{CS} obtained in **1b-jet, 4-5 jet region**
- small **differences** between **0b-jet and 1b-jet** regions and possible **contamination** from other **EWK** sources are corrected by:

$$\kappa_{\rm b}^{\rm MC} = \frac{R_{\rm CS}^{\rm MC}(n_{\rm jet} \in [4,5], 0{\rm b}, {\rm t\bar{t}+jets})}{R_{\rm CS}^{\rm MC}(n_{\rm jet} \in [4,5], 1{\rm b}, {\rm EWK})}$$

- expected number of tt+jets events in the signal region

$$N^{\rm pred.}_{t\bar{t}+{\rm jets}}(n^{\rm SR}_{\rm jet}, 0{\rm b}, \Delta\Phi > {\rm x}) = \kappa_{\rm b}^{\rm MC} \cdot {\rm R}^{\rm data}_{\rm CS}({\rm n_{jet}} \in [4,5], 1{\rm b}) \cdot {\rm y}^{\rm fit}_{\rm t\bar{t}+jets}({\rm n}^{\rm SR}_{\rm jet}, 0{\rm b}, \Delta\Phi < {\rm x})$$



- the fraction of tt+jets in the low Δϕ control region, y^{fit} is derived by a data/MC fit
- subtraction of QCD multijet background is considered in the fit method
- dependency of R_{CS} as a function of jet multiplicity is covered by a systematic uncertainty

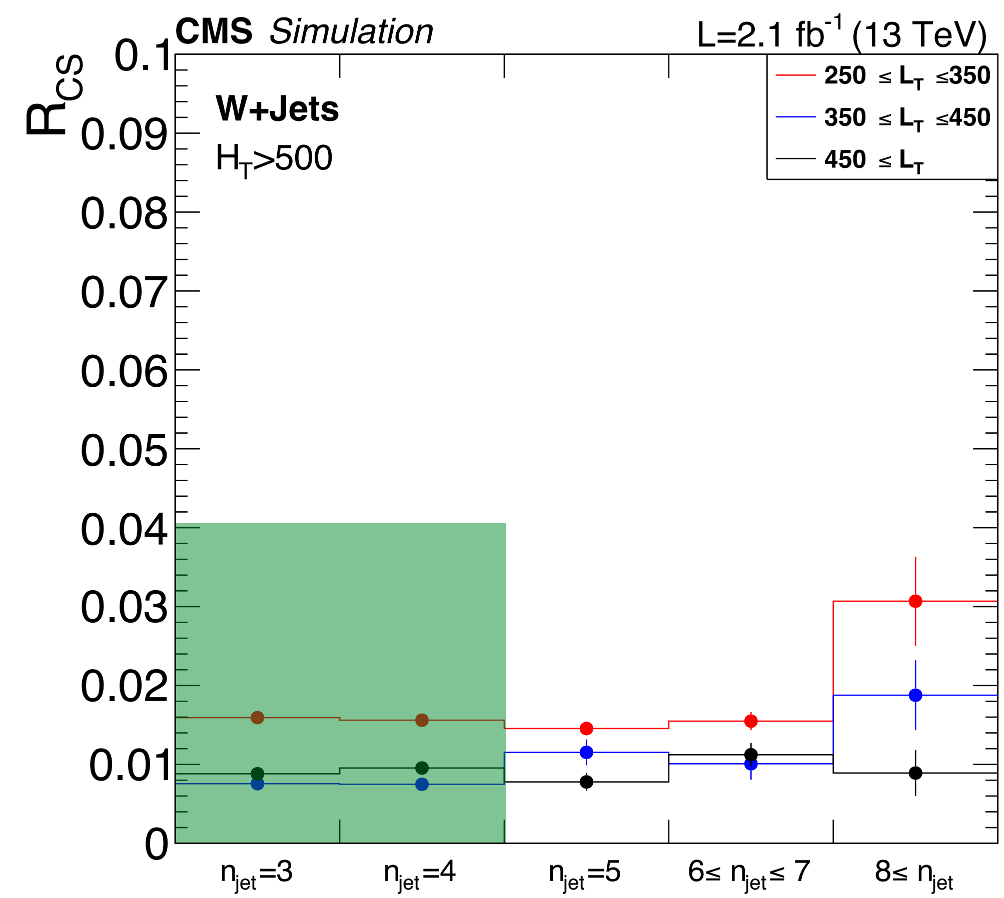
Prediction of the W+jets background (contd.)

	$n_{\text{b-jet}} = 0$	$n_{\text{b-jet}} = 1$
$n_{\text{jet}} = 3$	Bkg. estimation of W+jets and QCD multijets	
$n_{\text{jet}} = 4$		Bkg. estimation of $t\bar{t}$ + jets
$n_{\text{jet}} = 5$		
$6 \geq n_{\text{jet}} \geq 7$	Search regions	
$n_{\text{jet}} \geq 8$		

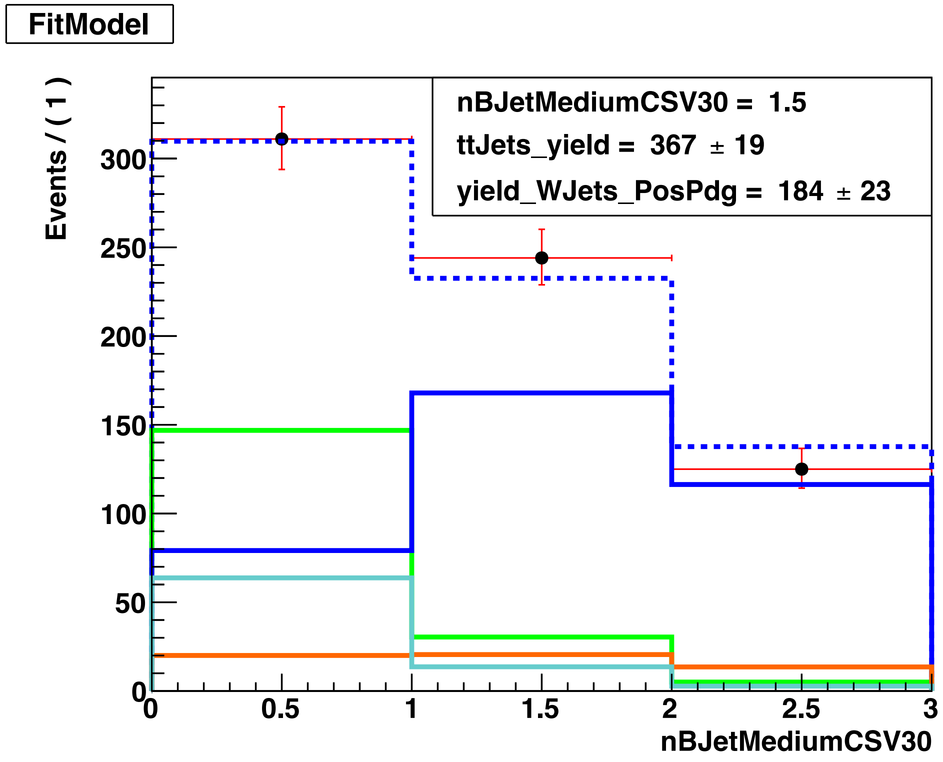
- R_{CS} obtained in **0b-jet, 3-4 jet region**
- R_{CS} is measured in the **μ-channel** to avoid significant **QCD contribution**
- ratio $R_{CS}(\mu)/R_{CS}(e+\mu)$ is assigned as an uncertainty
- expected number of W+jets events in the signal region

$$N^{\text{pred.}}_{W+jets}(n^{\text{SR}}_{\text{jet}}, 0\text{b}, \Delta\Phi > \text{x}) = \text{R}^{\text{corr.}}_{\text{CS}}(\text{n}_{\text{jet}} \in [3,4], 0\text{b}, \mu) \cdot \text{y}^{\text{fit}}_{\text{W+jets}}(\text{n}^{\text{SR}}_{\text{jet}}, 0\text{b}, \Delta\Phi < \text{x})$$

- possible contribution from tt+jets is considered in R^{corr}_{CS}
- dependency of R_{CS} as a function of jet multiplicity is covered by a systematic uncertainty

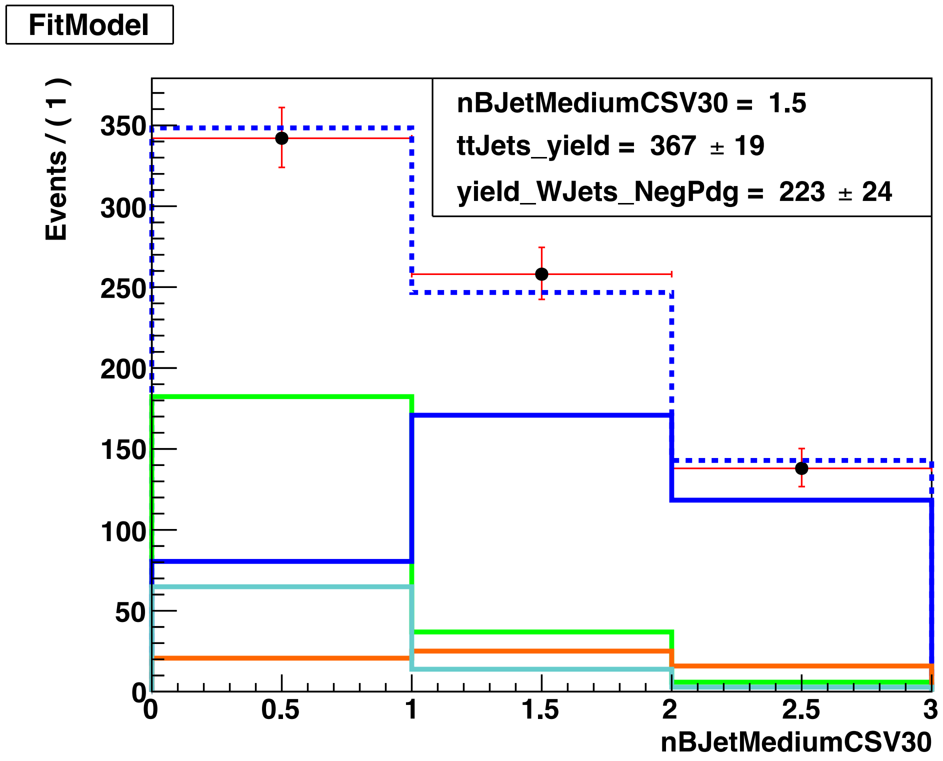


b-tag multiplicity fit

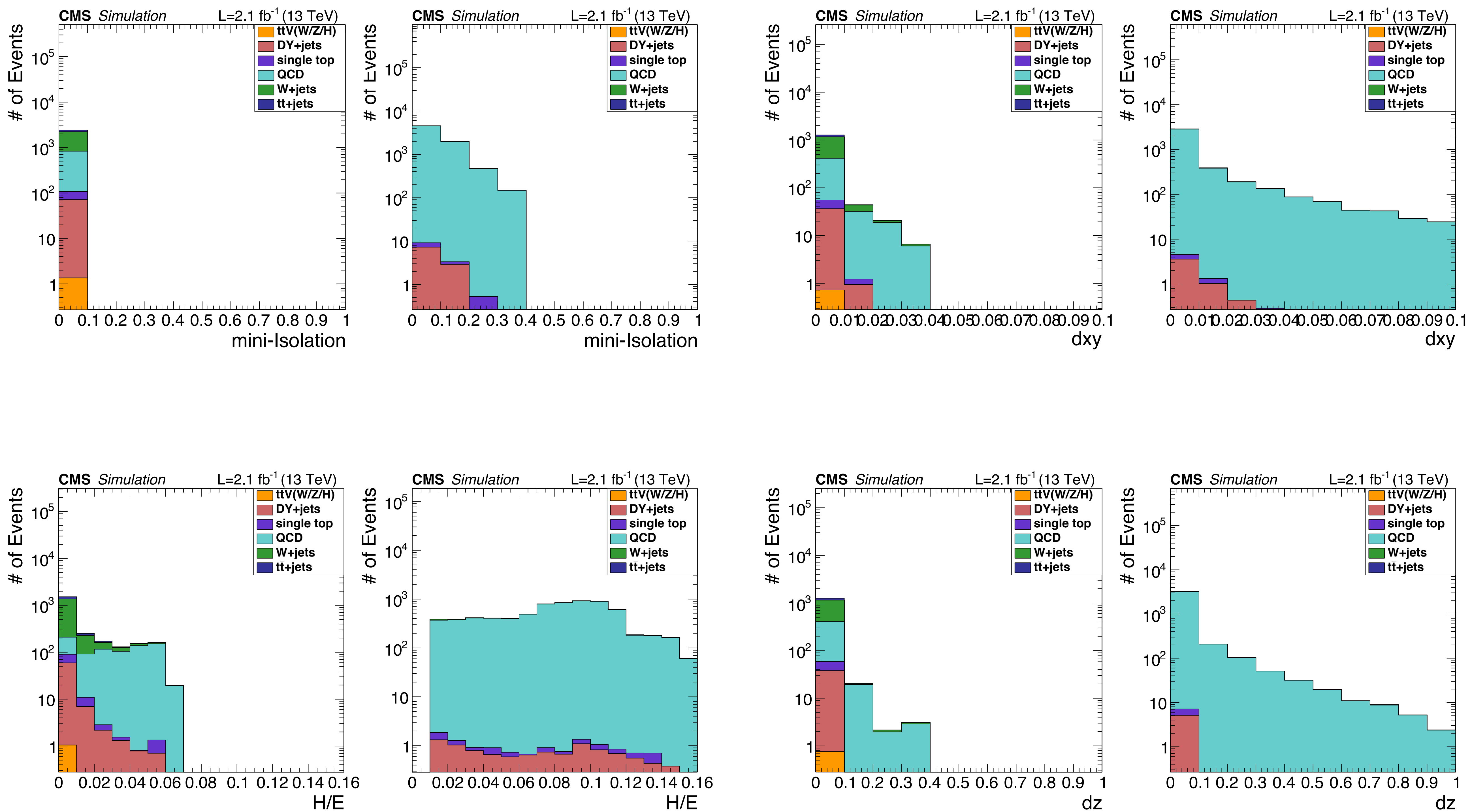


determine fractions tt+jets/W+jets in low ΔΦ control region

- binned likelihood fit in low ΔΦ control region
- make $W^{\pm}$+jets /tt+jets templates in b-jet multiplicity
- create constant QCD template from L_P-method
- set remaining EWK background as constant
- fit $W^{\pm}$+jets /tt+jets templates to data



Electron identification inversion for the QCD background estimation



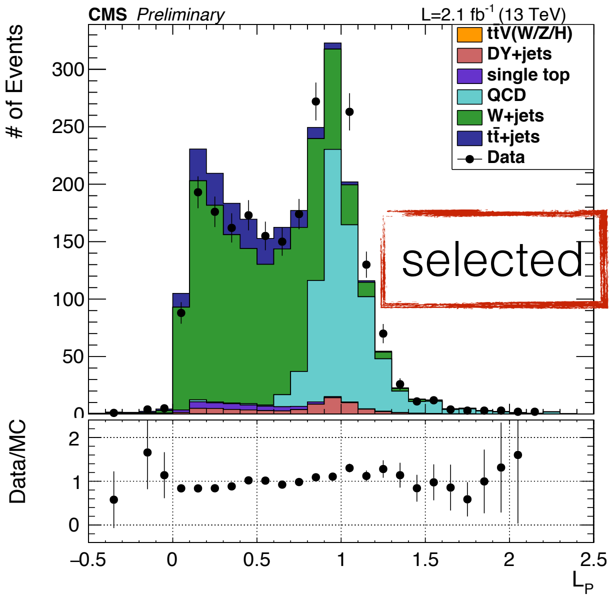
QCD multijet background estimation

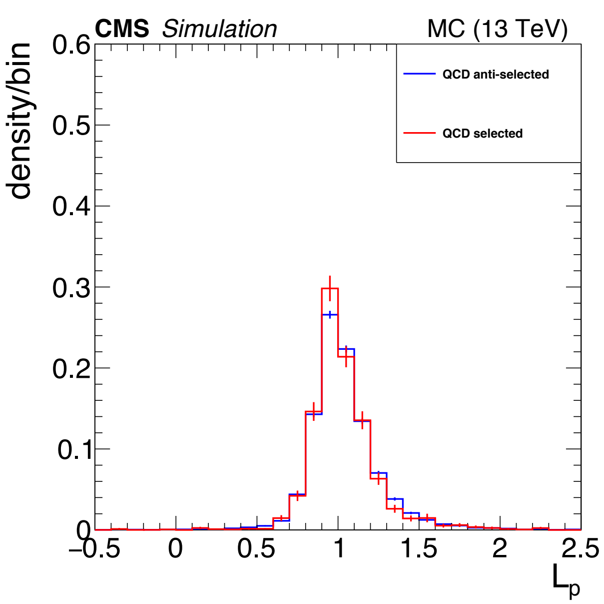
electron channel

- **majority** of electrons are **misidentified jets** and **converted photons**
- method is based on the L_P variable:

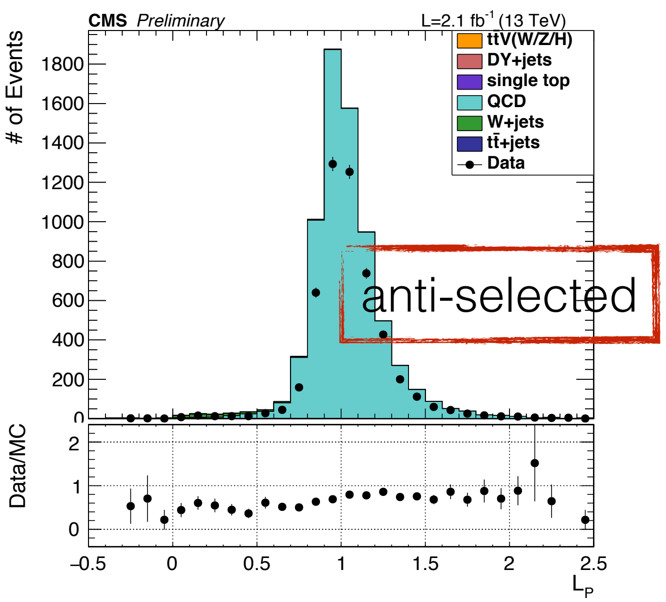
$$L_P = \frac{p_T(\ell)}{p_T(W)} \cdot \cos(\Delta\Phi(W,\ell))$$

- reflects the **W boson polarization**
- well understood for events with a genuine W decay
- different L_P distribution for lepton candidates from QCD sample





- **invert** electron identification requirements
  - → dominated by fake electrons
  - → **QCD enriched** and **EWK suppressed** sample
- 2 orthogonal lepton criteria: **'selected' — 'anti-selected'**
- **no impact** on the L_P distribution



QCD multijet background estimation (contd.)

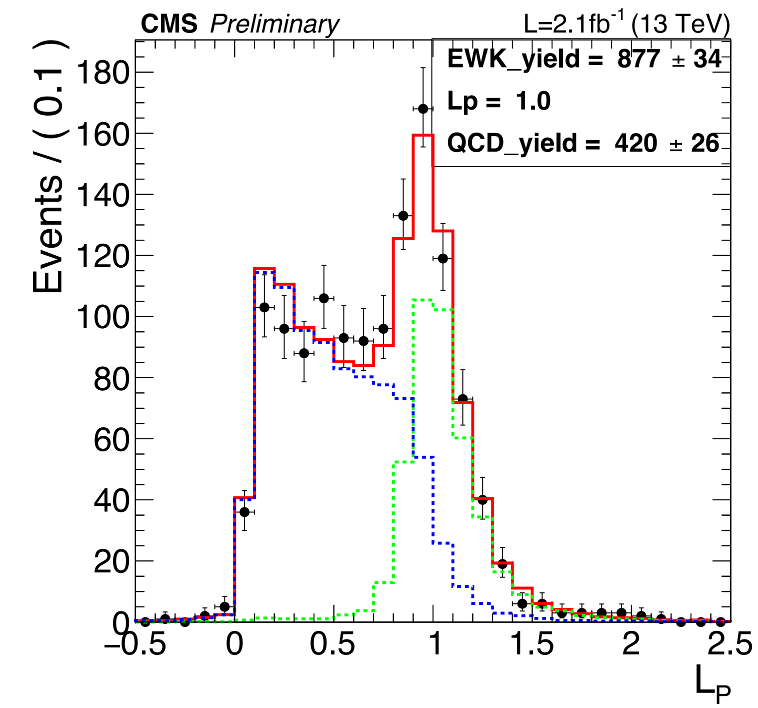
electron channel

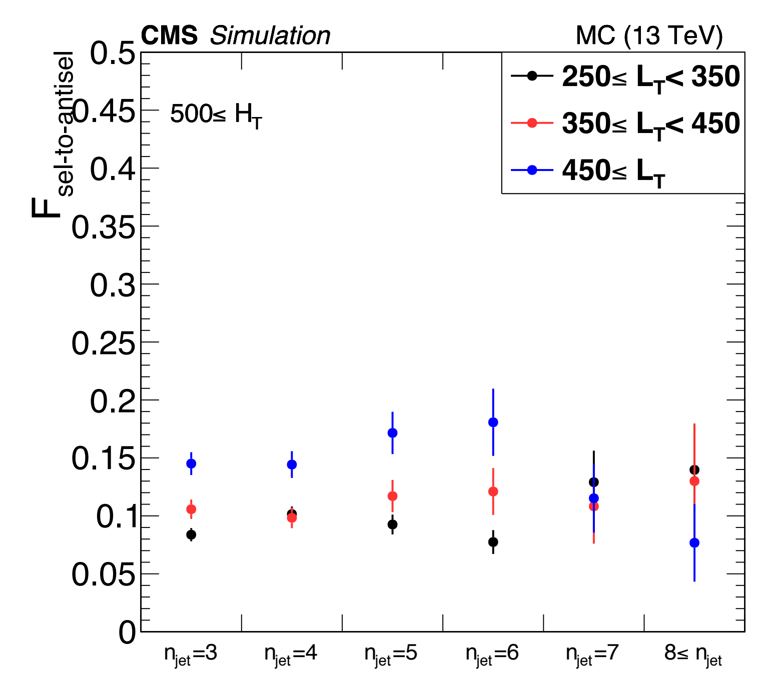
- perform a **binned likelihood fit** in the **0b-jet, 3-4 n_{jet} region**
  - **'selected'-EWK** template from simulation
  - **'anti-selected' data** template
- **fit result**: expected number of multijet events in the low n_{jet} region
- derive a ratio of 'selected' to 'anti-selected' events

$$F_{\text{sel-to-anti}}(L_T, n_{\text{jet}} \in [3,4] n_{\text{b-jet}} = 0) = \frac{N^{\text{fit}}_{\text{QCD selected}}(L_T, n_{\text{jet}} \in [3,4] n_{\text{b-jet}} = 0)}{N^{\text{data}}_{\text{QCD anti-selected}}(L_T, n_{\text{jet}} \in [3,4] n_{\text{b-jet}} = 0)}$$

- $F_{\text{sel-to-antisel}}$ is almost **independent** of n_{jet} multiplicity and H_T
- expected number of QCD events can be predicted:

$$N^{\text{pred.}}_{\text{QCD selected}}(L_T, H_T, n_{\text{jet}}) = F_{\text{sel-to-anti}} \cdot N^{\text{data}}_{\text{anti-selected}}(L_T, H_T, n_{\text{jet}})$$

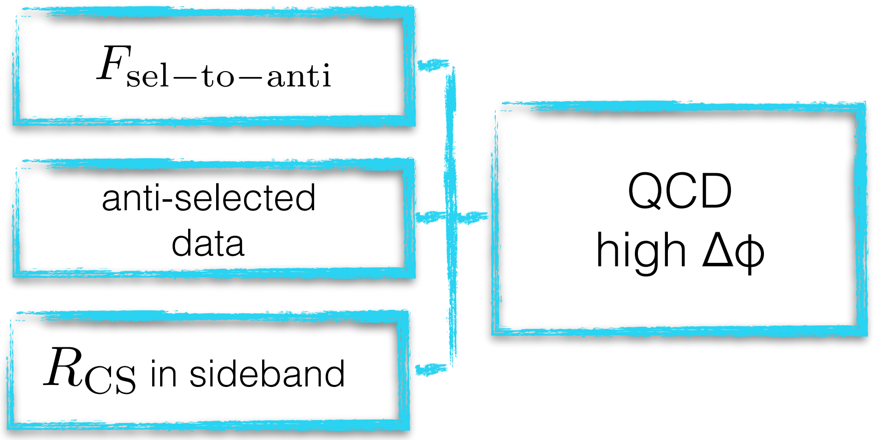


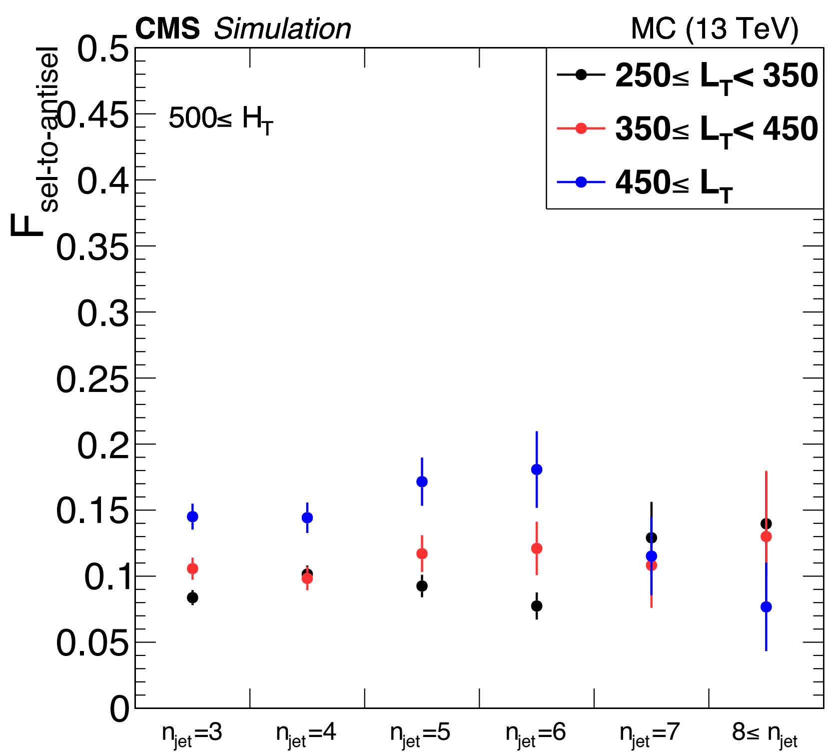


QCD multijet background estimation

<u>electron channel</u>

- perform a **data/MC fit** in the **0b-jet, 3-4 jet region**
    - **’selected’-EWK** template from simulation
    - **’anti-selected’ data** template
- **fit result**: expected number of multijet events in the sideband
- derive a **ratio** of **’selected’** to **’anti-selected’** events in sideband
- $F_{sel\text{-}to\text{-}antisel}$ is almost **independent** of n_{jet} multiplicity and H_T
- estimate the number of QCD events by scaling ’anti-selected’ data with $F_{sel\text{-}to\text{-}antisel}$



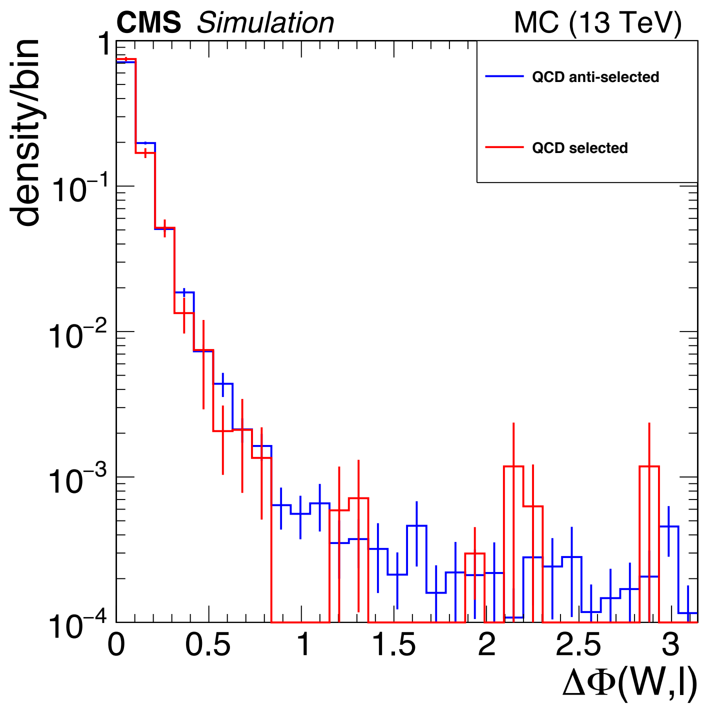


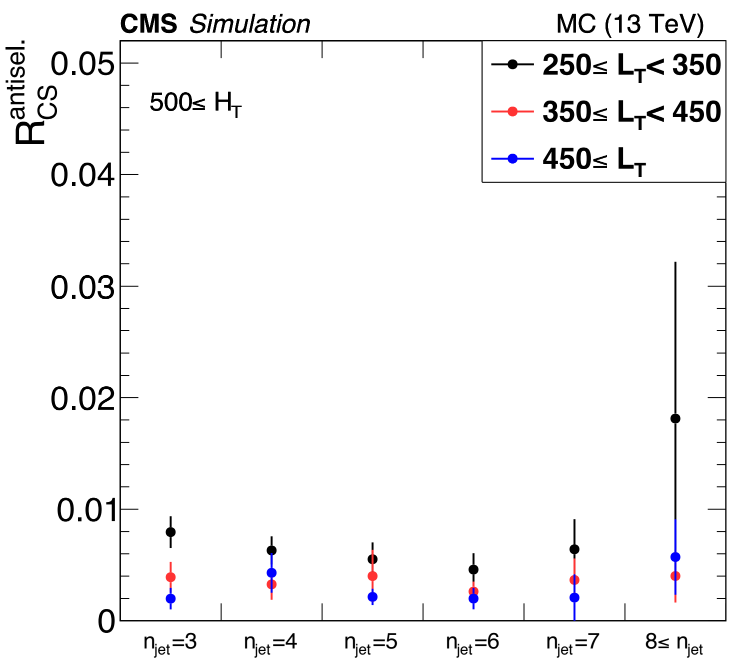
QCD multijet background estimation

electron channel

- **estimation** of multijet background over the **full Δϕ range**
- subtraction of QCD background in **’R_{CS}-method’** requires **separation** into **low Δϕ** and **high Δϕ** fractions

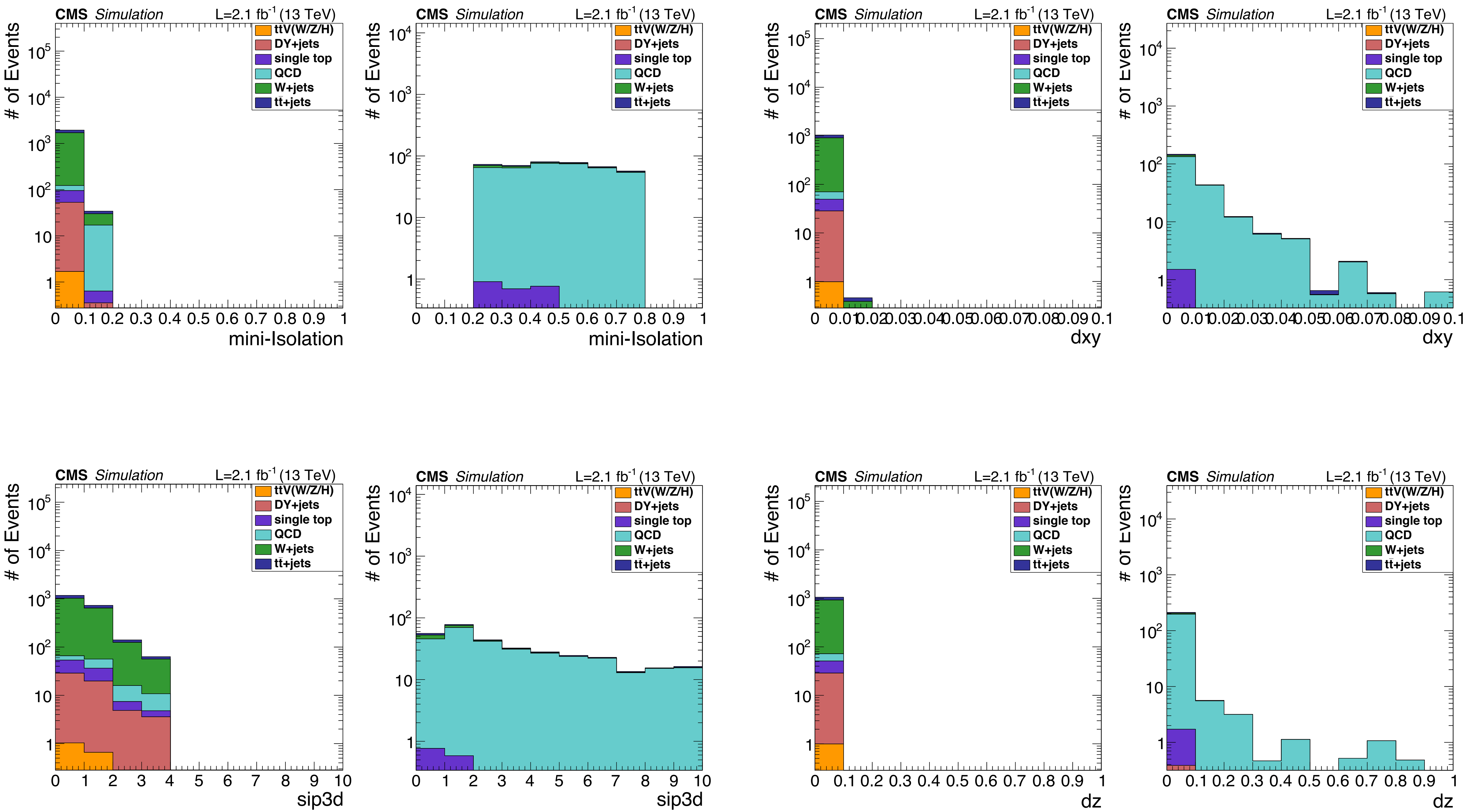
- **Δϕ distributions** of ’selected’ and ’anti-selected’ samples **agree**
- **better statistics** in the **’anti-selected’** sample





- R_{CS} transfer factors measured in **’anti-selected’** data
- **R_{CS}** transfer factors are **stable** and always **below 1%**
- **contribution** of multijet events in **signal regions** is **negligible**
- QCD background needs to be **subtracted** in the **control regions**

Muon identification inversion for the QCD background estimation



QCD multijet background estimation

muon channel

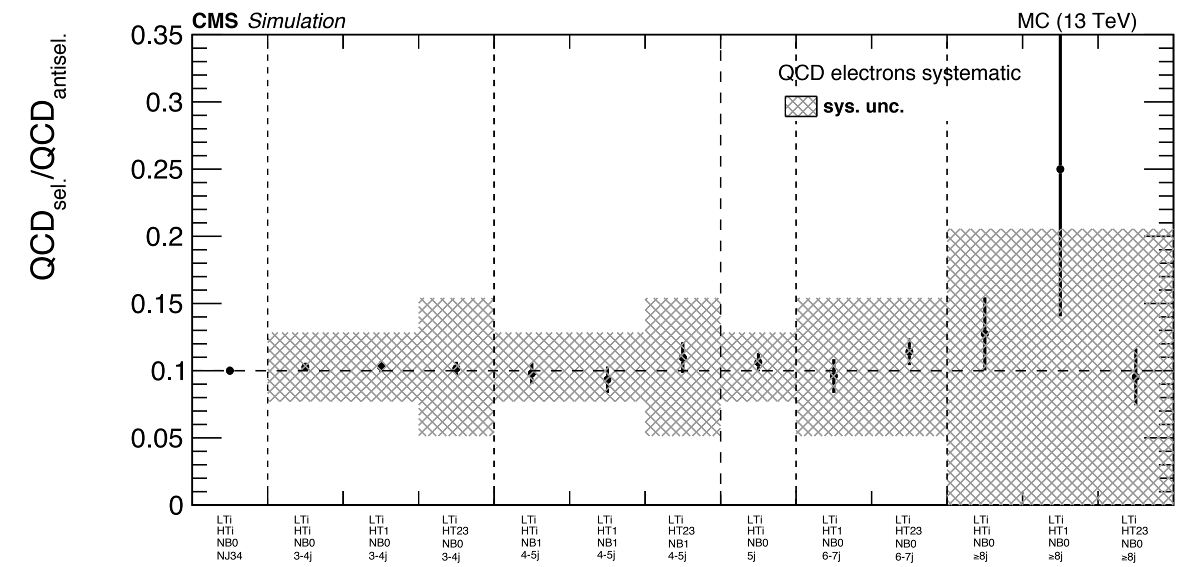
- very **small contribution** of multijet background in the **μ-channel**
- derive an **upper limit**
- method is similar to the electron channel, but relies more on properties from simulations
- **2 orthogonal lepton criteria** are defined to derive $\mathbf{F_{sel\text{-}to\text{-}antisel}}$
- very **small** $\mathbf{R_{CS}}$ transfer factors in 'anti-selected' sample
- **low contamination** of QCD background in full $\Delta\phi$ range

→ contribution of multijet events in **control** and **signal regions** is **negligible**

Systematic uncertainties

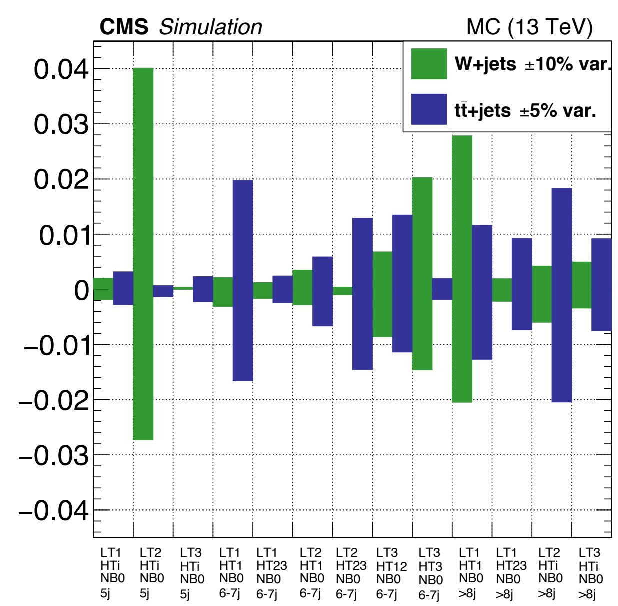
Systematic effects in the **multijet background**:

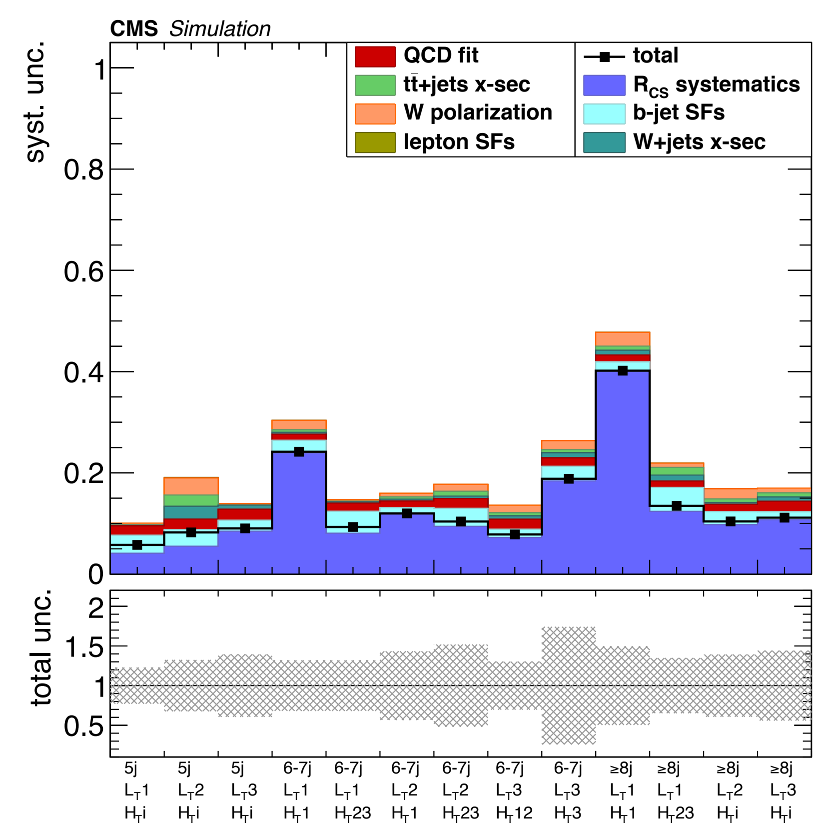
- $F_{sel\text{-}to\text{-}antisel}$ is independent of H_T and L_T
- conservative uncertainty of 25%-100% is assigned



Systematic uncertainties of the background estimation:

- b-jet scale factor uncertainty
- Jet energy and E_T^{miss} uncertainty
- Lepton identification and reconstruction efficiency
- Fraction of dilepton events
- Pileup
- Uncertainty on the QCD multijet prediction
- W+jet cross-section uncertainty
- **W polarization**
  - W+jets varied ±10%, tt+jets varied ±5%
  - $w = (1 \pm (0.1 \text{ or } 0.05) \cdot (1 - \cos(\theta^*))^2) \cdot \mathrm{x}^{\pm 1}$
- tt+jets cross-section uncertainty
- Top-quark p_T reweighting





Simulated samples

Sample name	cross section [pb]
TTJets_LO_25ns	831.76
TTJets_SingleLeptonFromT	269.49
TTJets_SingleLeptonFromTbar	269.49
TTJets_DiLepton	87.315
TTJets_HT600to800	2.665
TTJets_HT800to1200	1.097
TTJets_HT1200to2500	0.199
TTJets_HT2500toInf	0.002
TToLeptons_tch	70.314
TToLeptons_sch	3.681
TBar_tWch	35.6
T_tWch	35.6
DYJetsToLL_M50_HT100to200	177.038
DYJetsToLL_M50_HT200to400	54.293
DYJetsToLL_M50_HT400to600	6.981
DYJetsToLL_M50_HT600toInf	2.807
TTZToLLNuNu_25ns	0.2529
TTZToQQ_25ns	0.5297
TTWJetsToLNu_25ns	0.2043
TTWJetsToQQ_25ns	0.4062
WJetsToLNu_HT100to200	1656.81
WJetsToLNu_HT200to400	442.8
WJetsToLNu_HT400to600	60.147
WJetsToLNu_HT600toInf	23.087
WJetsToLNu_HT600to800	15.744
WJetsToLNu_HT800to1200	6.47
WJetsToLNu_HT1200to2500	1.636
WJetsToLNu_HT2500toInf	0.038
QCD_HT100to200	27540000
QCD_HT200to300	1735000
QCD_HT300to500	366800
QCD_HT500to700	29370
QCD_HT700to1000	6524
QCD_HT1000to1500	1064
QCD_HT1500to2000	121.5
QCD_HT2000toInf	25.42

simulated SM background samples

Sample name	cross section [pb]
$T5q^4 1/0.8/0.7$	0.325
$T5q^4 1.2/1/0.8$	0.086
$T5q^4 1.5/0.8/0.1$	0.014

simulated SUSY signal samples