Imperial College London

Searches for BSM+jets at CMS

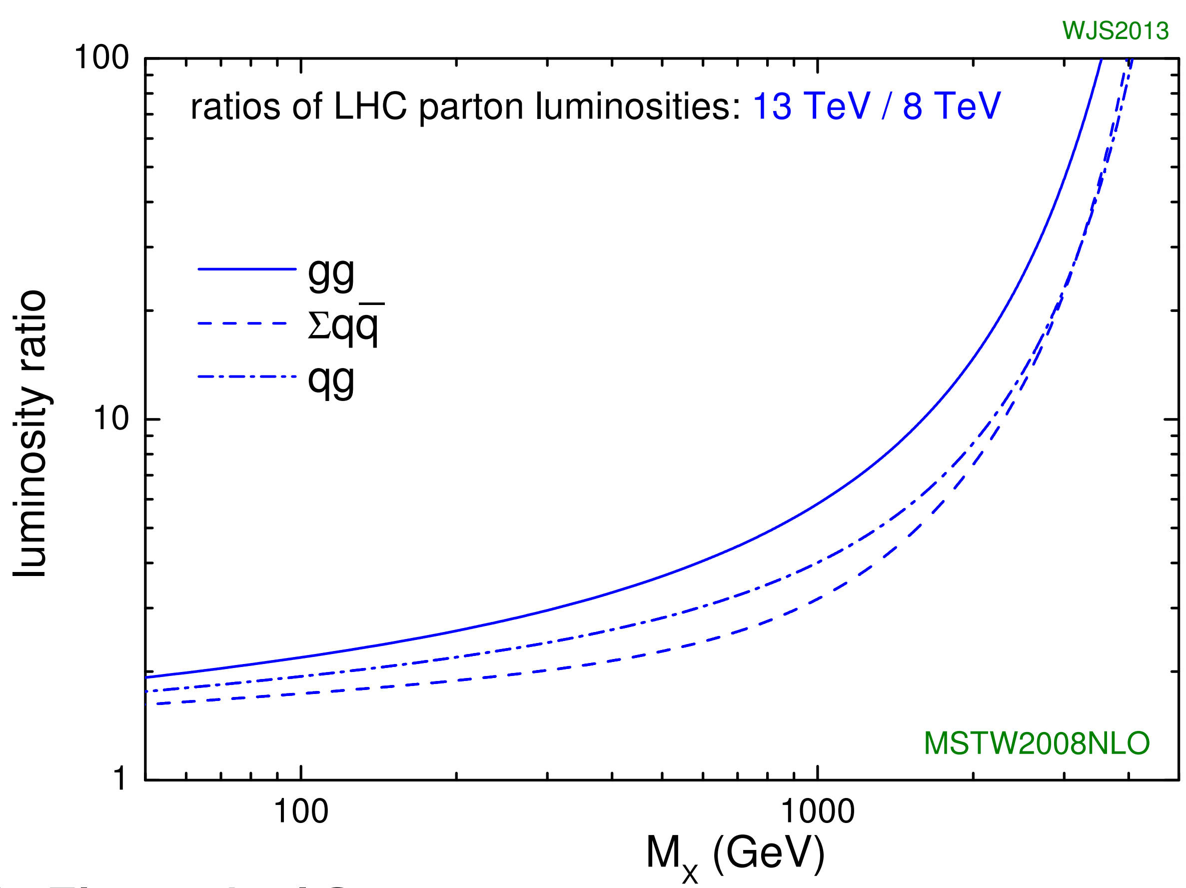
Adam Elwood
on behalf of the CMS Collaboration

Jet vetoes and multiplicity observables workshop
IPPP, Durham, 19-21 September 2016

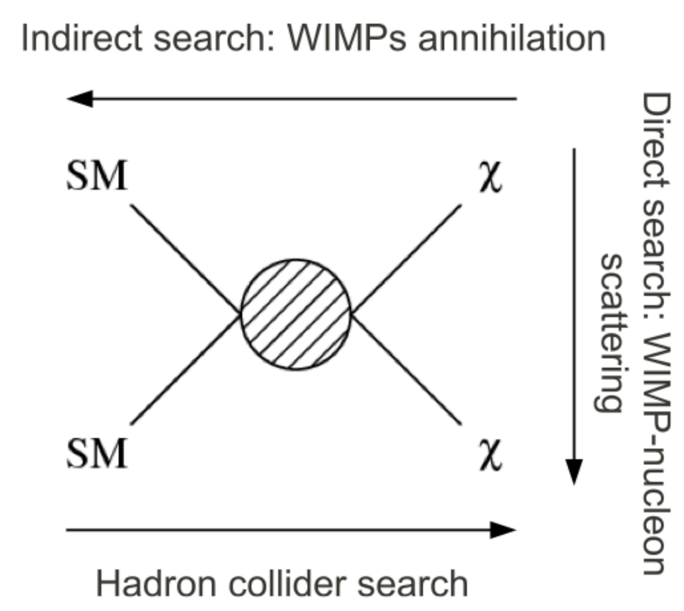
ace09@ic.ac.uk

Motivation for BSM searches with jets at the LHC

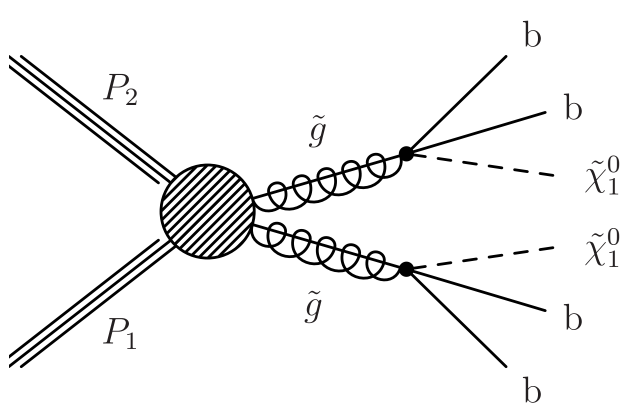
- A Higgs boson has been discovered but a lot of questions about the standard model remain
  - Hierarchy problem, dark matter, gravity, dark energy
- As the LHC is a p-p collider the highest cross section BSM production processes occur hadronically
- In this talk focus on searches of BSM with jets and missing energy



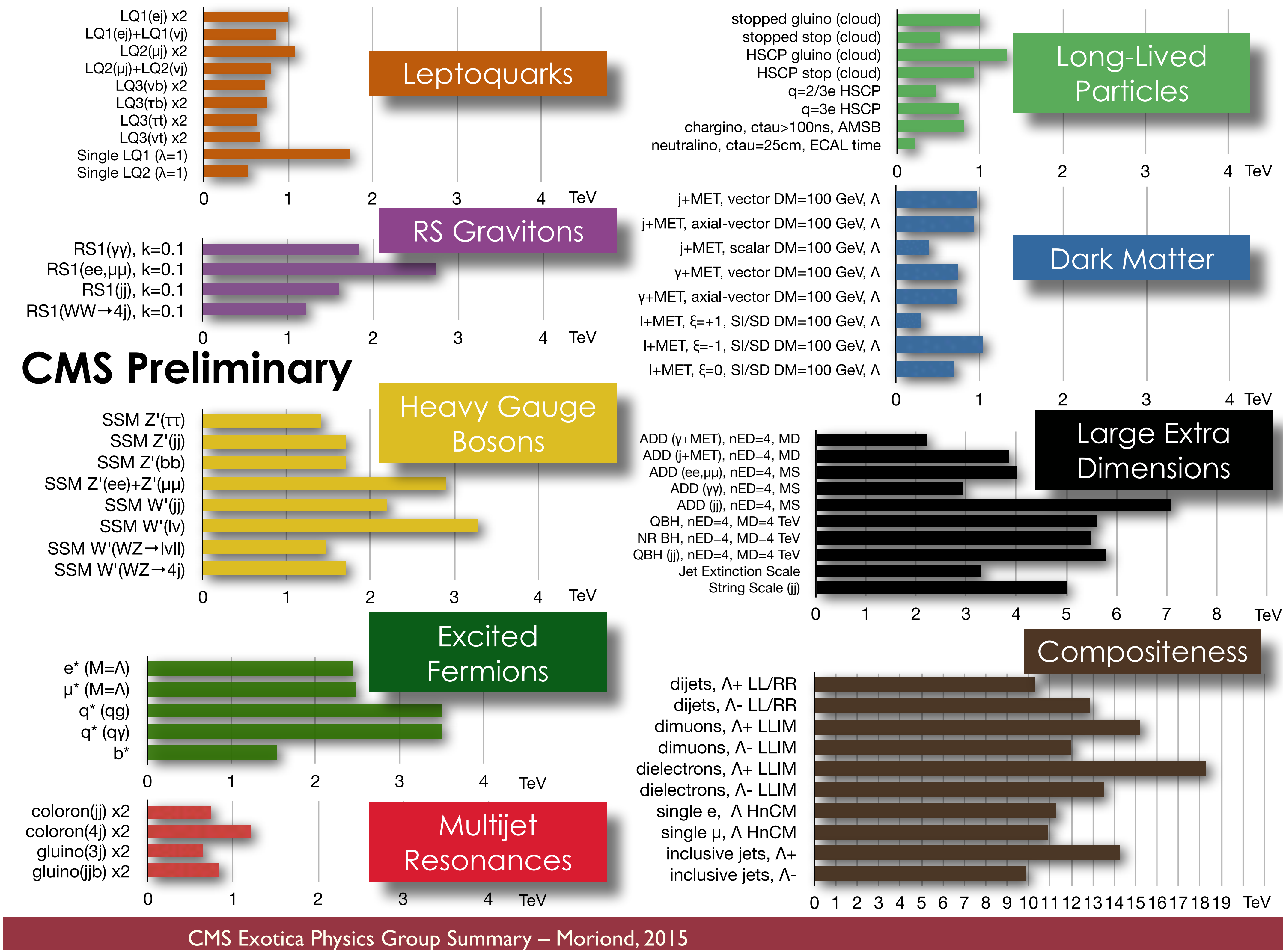
Dark matter



Supersymmetry

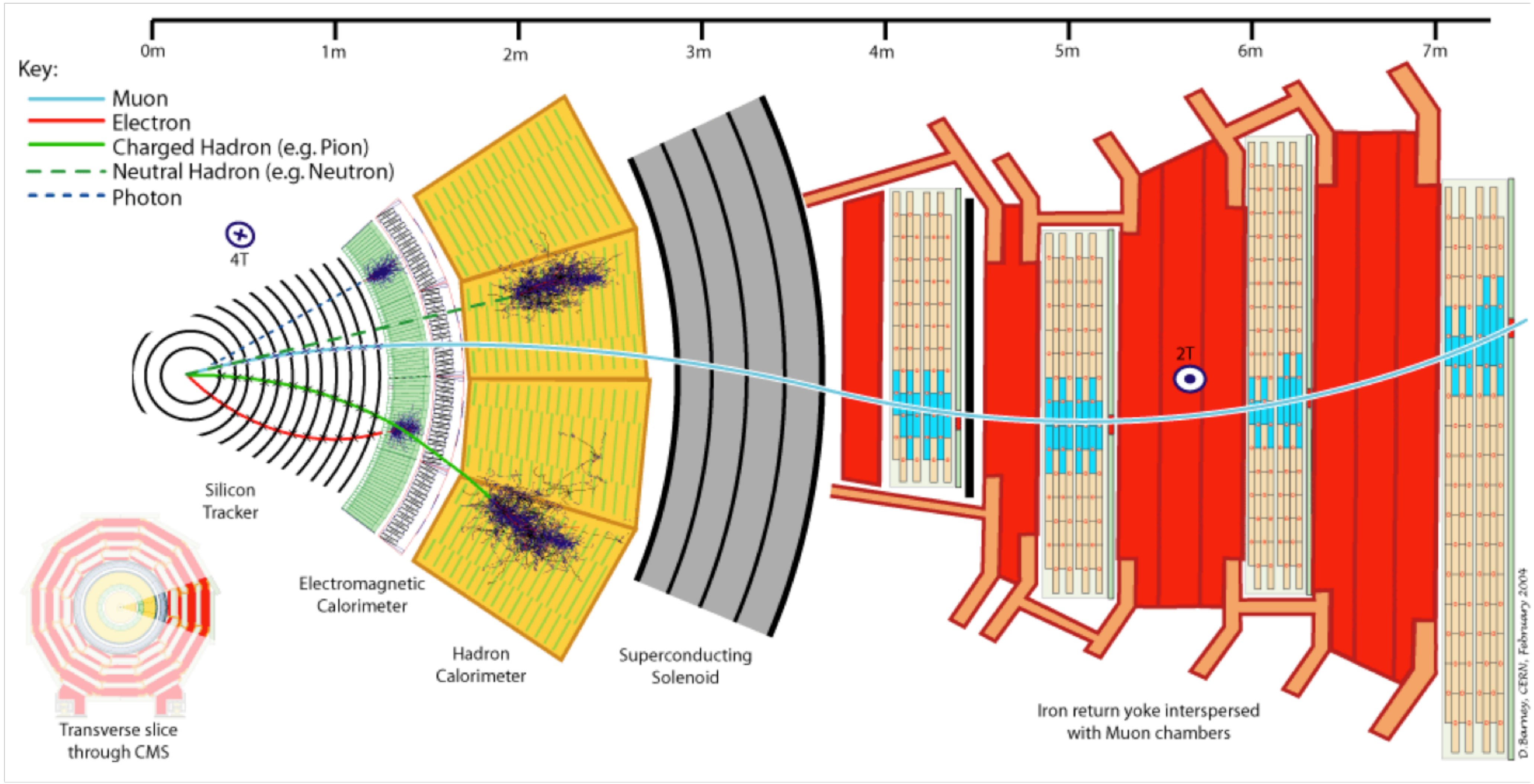


Much more beyond SUSY and DM..



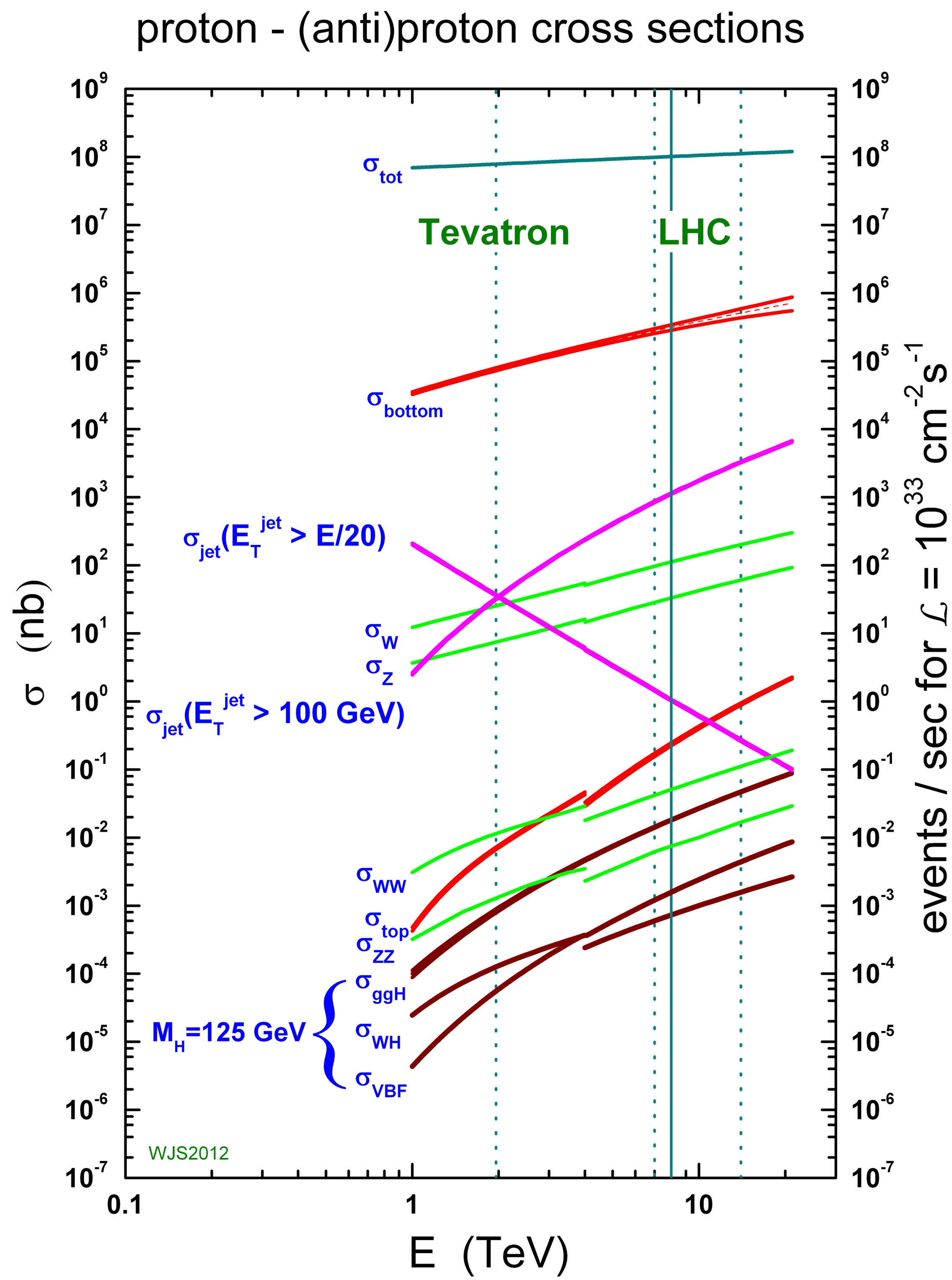
The Compact Muon Solenoid (CMS)

- CMS is a general purpose detector designed to search for new particles produced in proton proton collisions at the LHC
- It consists of a series of sub detectors built around a 4 Tesla solenoid that are exploited to reconstruct the particles produced in each collision event
- Due to its comprehensive coverage of all angles the missing momentum in each collision event can be accurately determined



Challenges of hadronic BSM searches

- Very high cross-section for QCD events at the LHC that mainly produce non-genuine E_T^{miss} through mismeasurement
  - High production rates present significant challenge for the trigger
  - Difficult to quantify due to challenges involved in simulating QCD events
- Electroweak processes that produce genuine E_T^{miss}
  - W decays that are suppressed with a lepton veto
  - An irreducible Z to neutrino background that must be accurately predicted
- The types of events that are most sensitive to new physics typically sit in the tails of SM distributions, e.g. E_T^{miss} , N_{jet}
  - MC does not model these tails very reliably



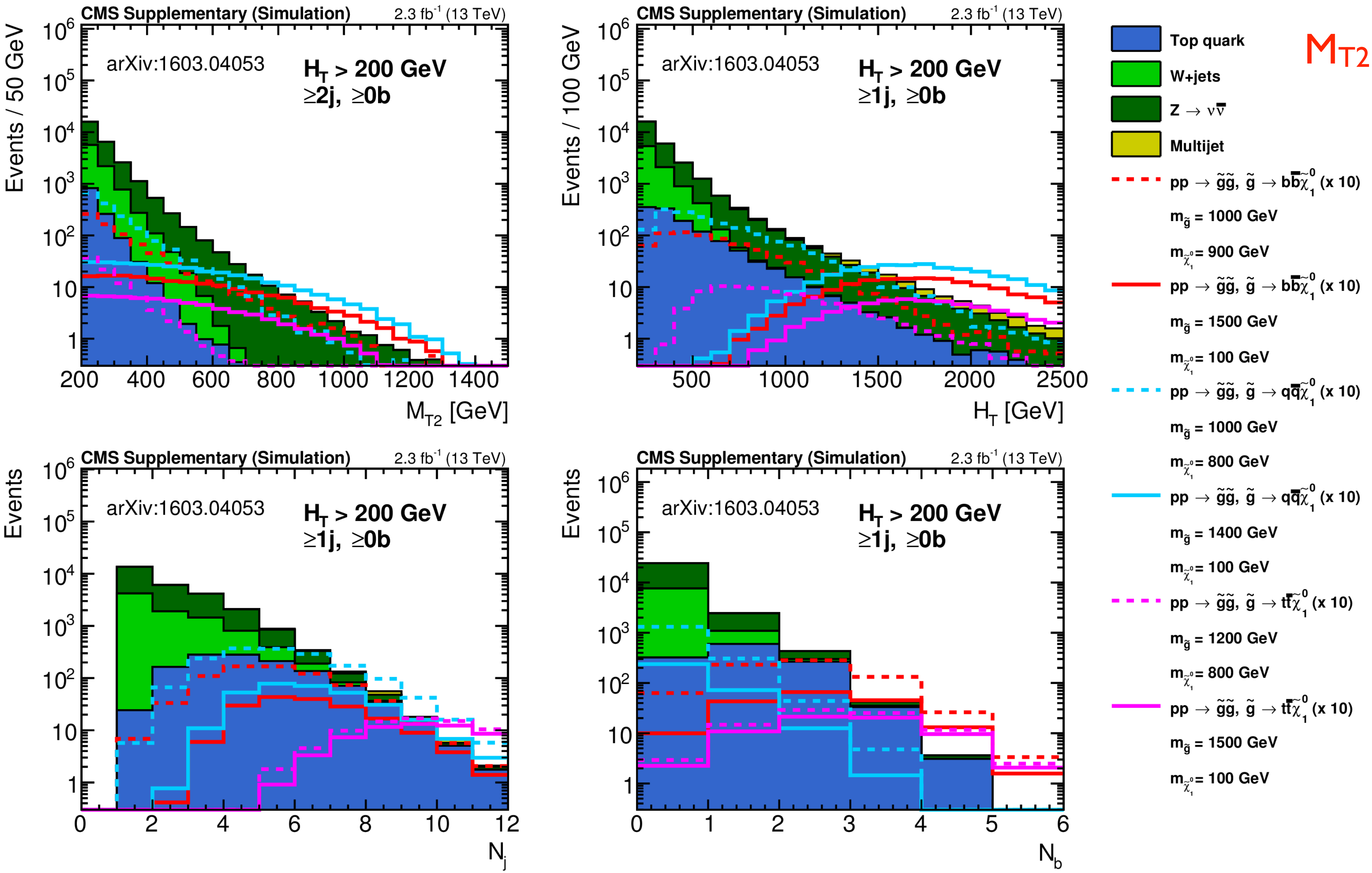
Outline of typical search

- The use of a missing energy based discriminating variable
  - e.g. E_T^{miss} , M_{T2}, α_T
  - They act as 'QCD-killers' (e.g. α_T)
  - Can also separate signal from electroweak backgrounds (e.g. M_{T2})
- A trigger strategy is then defined based on the discriminating variables and energy scale of the events
- Additional QCD suppression through angular distribution of missing energy and jets
  - e.g. $\Delta\phi(\text{jet}, E_T^{miss}) > 0.3$
- Backgrounds are predicted with data driven methods based on control regions
  - Single lepton to predict top and W+jets backgrounds
  - Z to neutrinos from double lepton and photon control regions
  - Remaining QCD predicted with multi-jet enriched QCD control sample (e.g. inverted $\Delta\phi$ cut)
- Sensitivity obtained by binning in the discriminating variable, energy scale and jet multiplicity variables

Example SUSY analyses

- In the bulk of this talk will use plots and examples from three hadronic SUSY analyses made with the 2015 and 2016 data collected at 13 TeV
  - The "α_T" analysis
    - CMS-PAS-SUS-16-016 and CMS-PAS-SUS-15-005
  - The "M_{T2}" analysis
    - CMS-PAS-SUS-16-016 and CMS-SUS-15-003
  - The "H_T^{miss}" analysis
    - CMS-PAS-SUS-16-014 and CMS-SUS-15-002

Distributions of binning variables



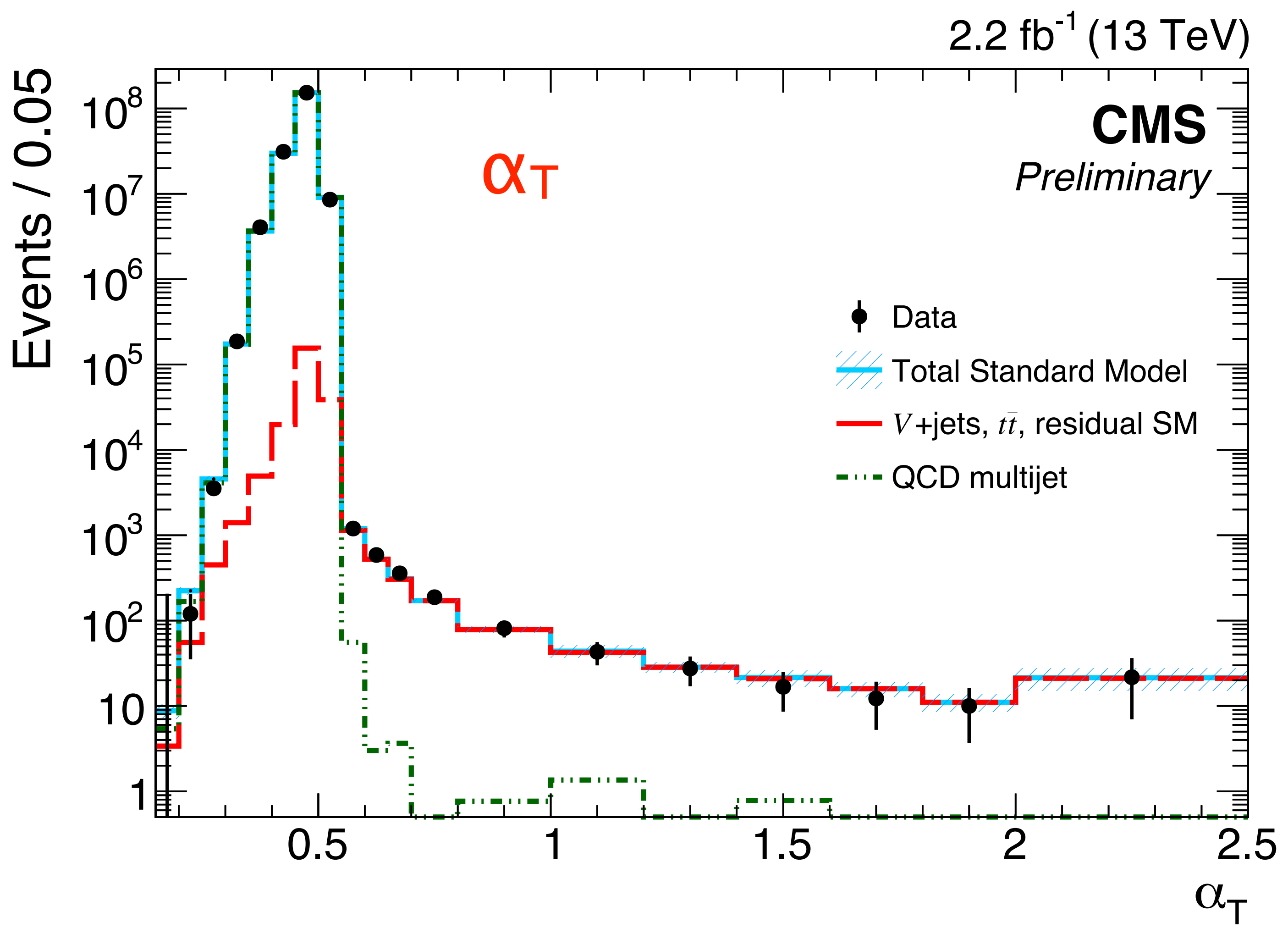
http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-003/index.html

Specific example - the α_T analysis

- **Inclusive search targeting SUSY and Dark Matter**
  - Vetoing leptons or photons
  - Consider all jet and b-tag multiplicities ($n_{jet} \geq 1$, $n_b \geq 0$)
  - Low threshold on jet activity to retain sensitivity to compressed SUSY scenarios
    - $H_T \equiv \Sigma\, p_T^{jet} > 200$ GeV, $H_T^{miss} \equiv -\Sigma\, \underline{p}_T^{\,jet} > 130$ GeV
  - Thresholds maintained through a suite of dedicated triggers
- **Robust analysis**
  - Suppress QCD multijet background with tight cuts on dedicated variables: α_T, $\Delta\phi^*$, H_T^{miss}/E_T^{miss}
  - Data-driven estimation of the other backgrounds
- **Sensitivity**
  - Binning in n_{jet}, n_b, H_T and H_T^{miss}
  - Further categorisation based on n_{jet} topologies

The α_T variable

- The analysis strategy revolves around reducing QCD to a negligible level with an appropriate cut on α_T
- It does this by separating processes with genuine missing energy from those with mismeasured jets that result in missing energy along the jet axis
- Originally designed for di-jet events then generalised to multi-jets



$$\alpha_T = \frac{E_T^{j_2}}{M_T}$$

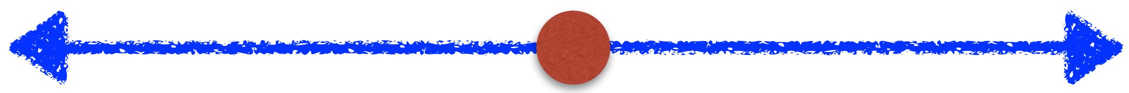
$$\alpha_T = \frac{1}{2} \times \frac{H_T - \Delta H_T}{\sqrt{H_T^2 - (H_T^{miss})^2}}$$

For a pseudo di-jet system with pseudo jet p_T difference: ΔH_T

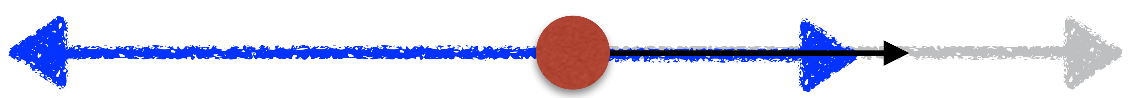
The pseudo jets are constructed from a sum of all jets in the system so as to minimise ΔH_T

What α_T is doing

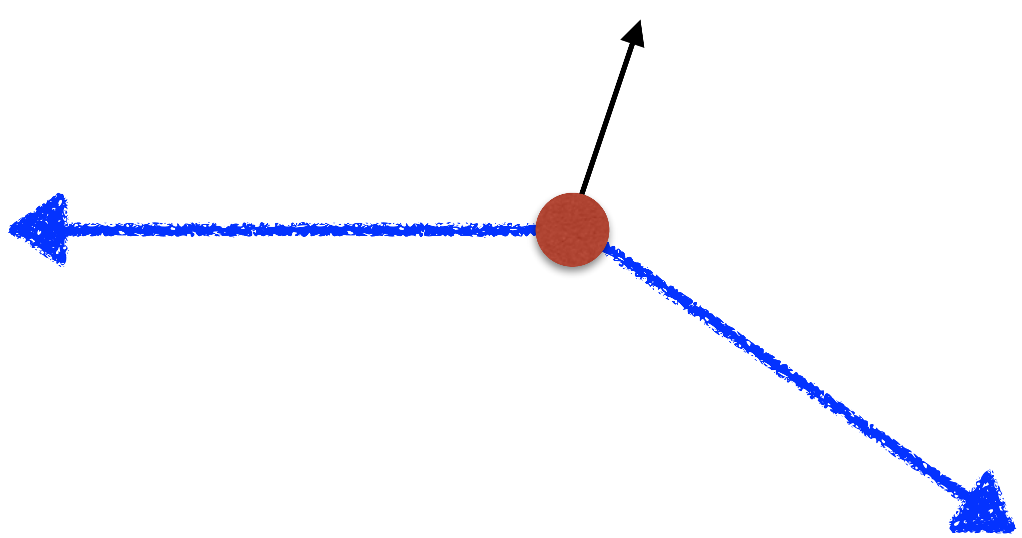
$$\alpha_T = \frac{1}{2} \times \frac{H_T - \Delta H_T}{\sqrt{H_T^2 - (H_T^{miss})^2}}$$

(1) 

Balanced multijet event:
$\alpha_T = 0.5$

(2) 

Mismeasured multijet event:
$\alpha_T < 0.5$

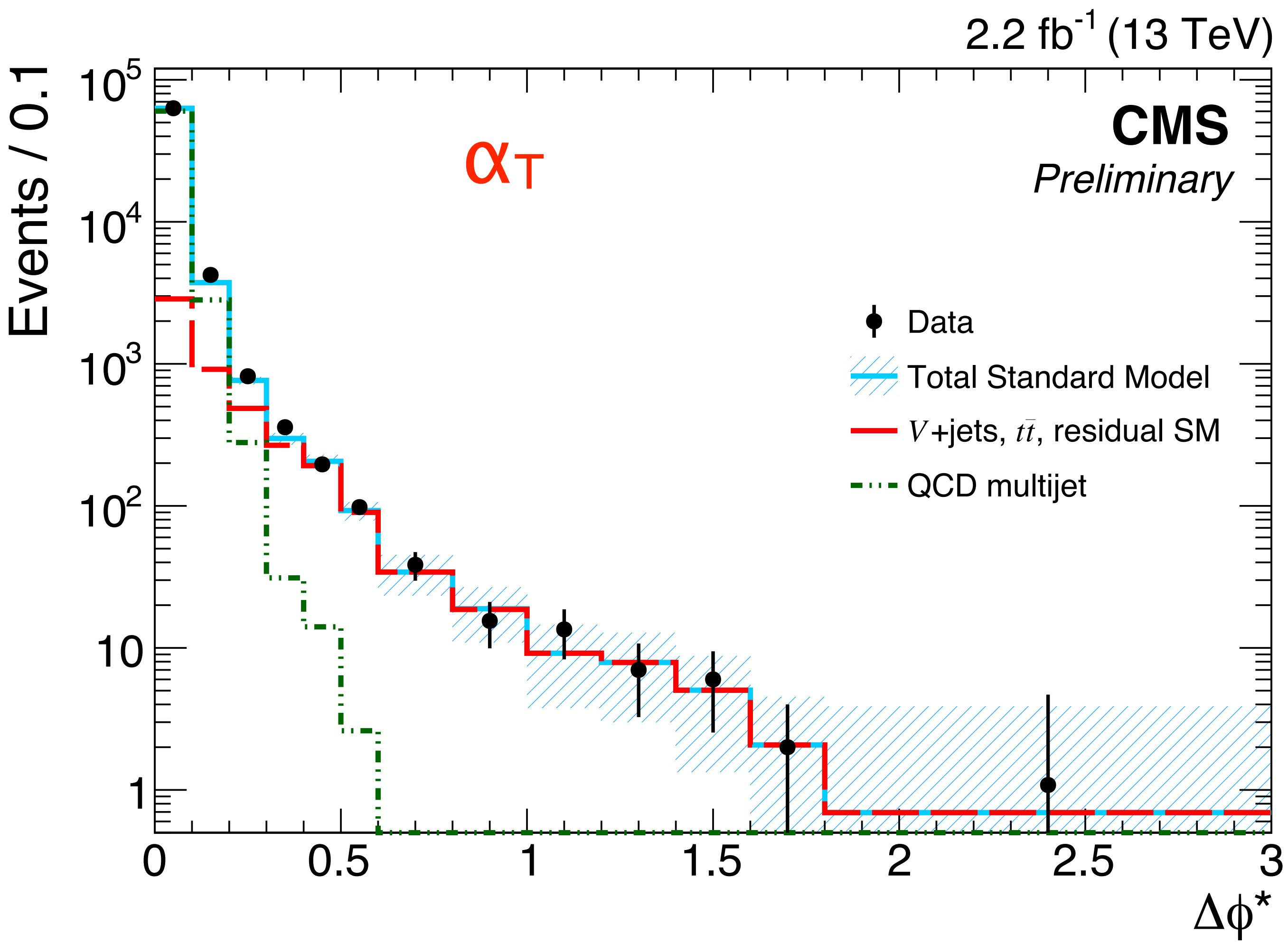
(3) 

Genuine missing energy multijet event:
$\alpha_T > 0.5$

= Jet = H_T^{miss}

The $\Delta\phi^*$ variable

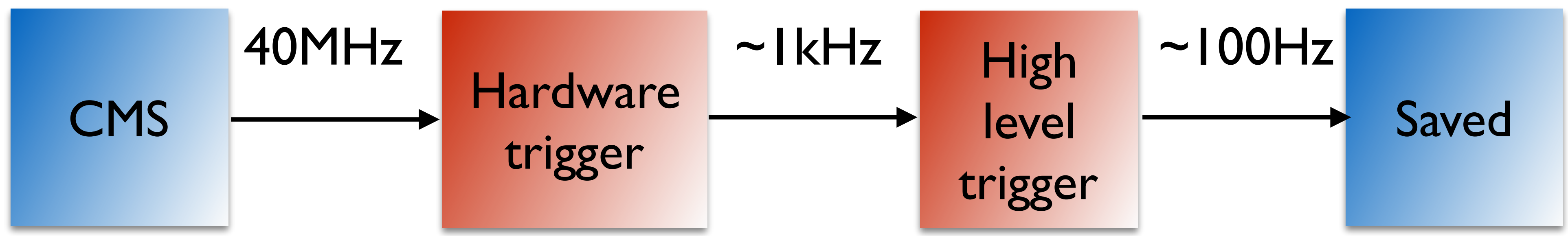
- In multi-jet systems QCD processes can still pass the α_T cut
  - e.g. heavy flavour QCD with neutrinos in the final state or extreme mismeasurement of multi jet systems
- Define a variable based on $\Delta\phi$ to further remove the remaining QCD events
- Constructed in such a way to be sensitive to under and over measurement



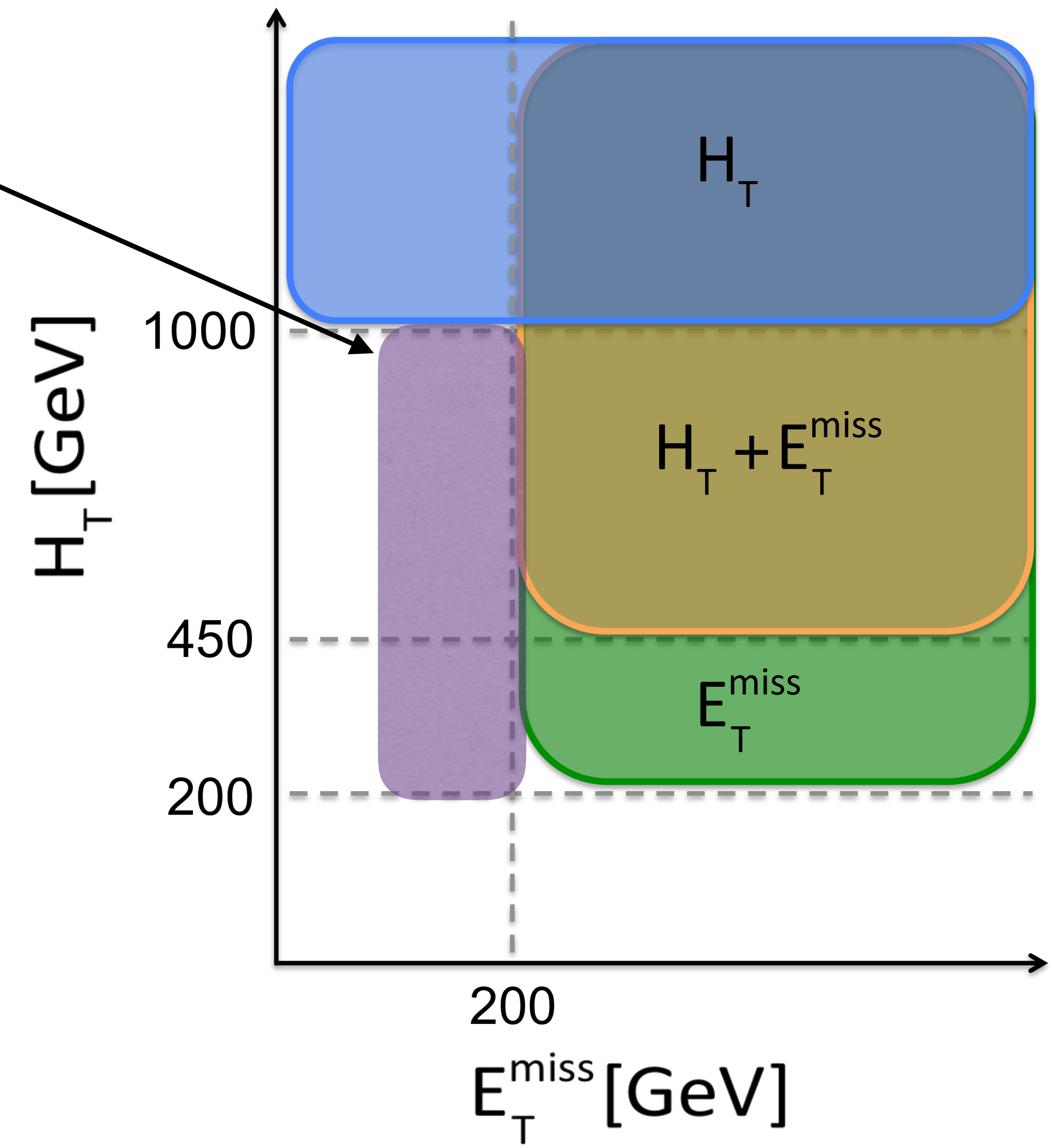
$\Delta\phi$ = minimum azimuthal angle between any jet and H_T^{miss}

$\Delta\phi^*$ = minimum azimuthal angle between any jet and H_T^{miss} computed without the jet (a stricter requirement than canonical $\Delta\phi$)

Trigger



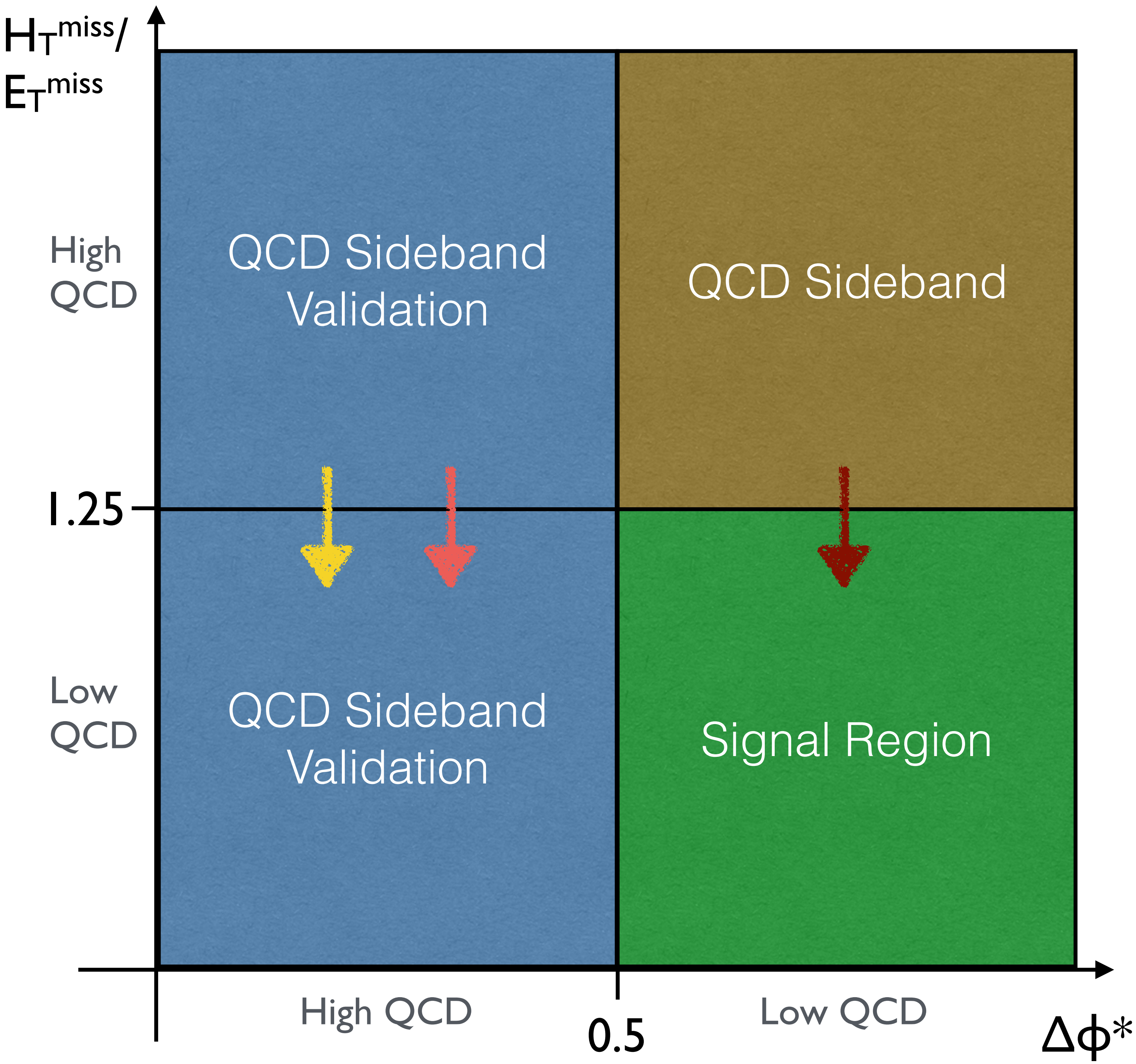
- **Signal triggers**
  - Seeded by E_T^{miss} and H_T at the hardware trigger level
  - 5 dedicated α_T X H_T triggers for a series of H_T thresholds below 800 GeV
    - Require a minimum value of α_T that depends on the H_T threshold (above 200 GeV)
  - Pure H_T trigger above 800 GeV
  - An additional E_T^{miss} based trigger to maximise acceptance
- **Control triggers**
  - Require a single lepton or photon with low as possible threshold for reasonable rate
- close to 100% efficiency in the full signal region phase space



http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-003/index.html

QCD prediction and validation

- Appropriate cuts reduce QCD background to the % level of the electroweak background but must be predicted



= MC based transfer factor, R

= data driven transfer factor, R^{Data}

A data driven estimate of QCD is carried out by subtracting the predicted electroweak backgrounds from the data counts

Require that ~=

Any difference is covered with a fully correlated systematic

Genuine E_T^{miss} background prediction

- After QCD reduction left with non-multijet Standard Model backgrounds
  - In association with lost leptons: $t\bar{t}$,W+jets
  - Irreducible: Z decaying to neutrinos
- Make use of the single muon, double muon and single photon control samples
- Emulate the missing energy by ignoring the leptons when calculating variables
- Extrapolate yields from one control sample to signal region via appropriate transfer factors in MC simulation

$$N_{\mathrm{pred}}^{\mathrm{signal}}(n_{\mathrm{jet}}, n_{\mathrm{b}}, H_{\mathrm{T}}) = \frac{N_{\mathrm{MC}}^{\mathrm{signal}}(n_{\mathrm{jet}}, n_{\mathrm{b}}, H_{\mathrm{T}})}{N_{\mathrm{MC}}^{\mathrm{control}}(n_{\mathrm{jet}}, n_{\mathrm{b}}, H_{\mathrm{T}})} \times N_{\mathrm{obs}}^{\mathrm{control}}(n_{\mathrm{jet}}, n_{\mathrm{b}}, H_{\mathrm{T}})$$

Control regions used for predictions

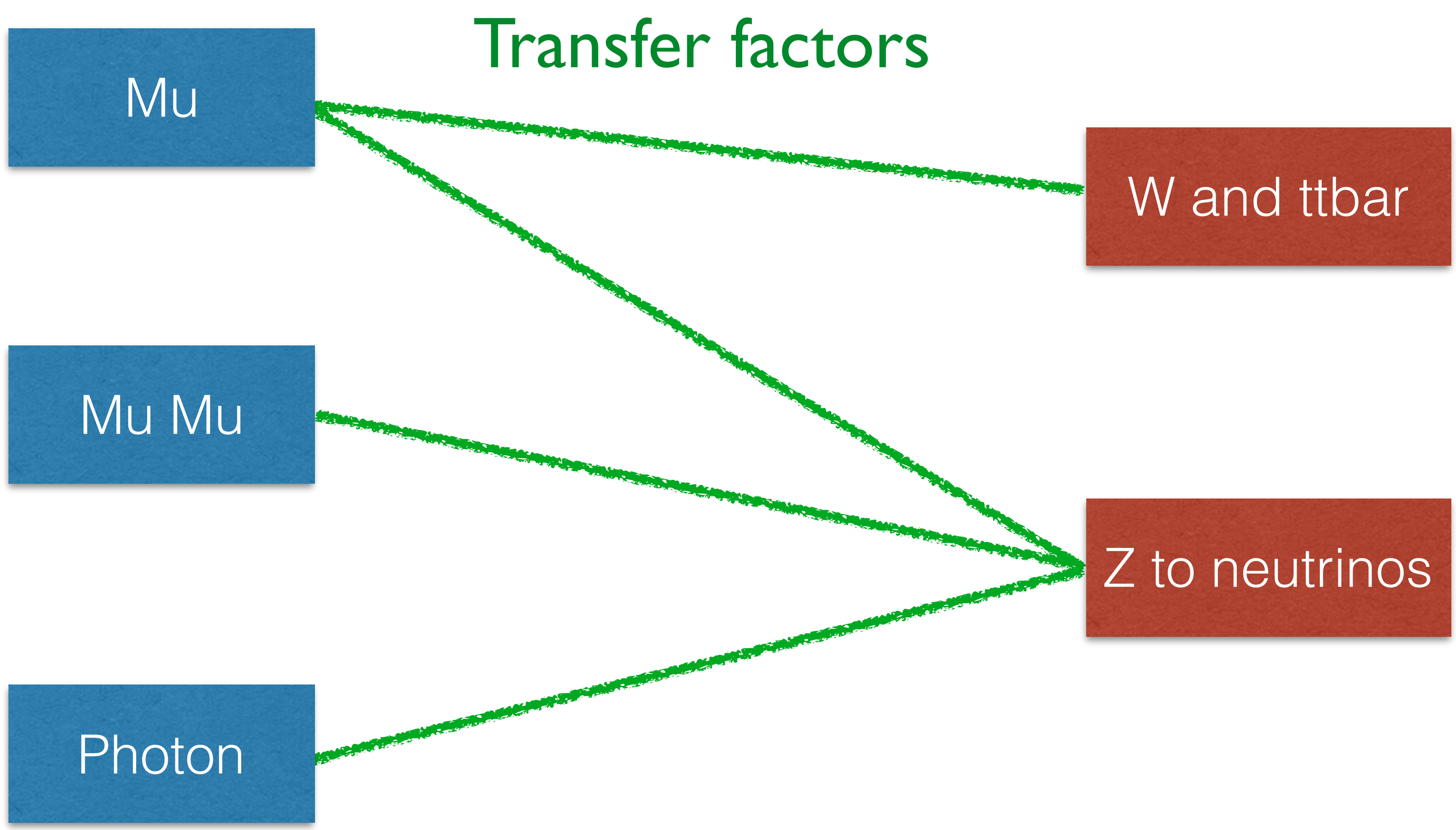
Control regions

Signal region

Predominantly W and ttbar with high statistics

Predominantly Z with small uncertainties

High statistics but QCD has to be controlled



Determination of systematic uncertainties on the remaining non-multijet backgrounds from simulation

- To search for significant excesses in hadronic events with missing energy must take account of all systematic uncertainties
- Vary known experimental and theoretical uncertainties by $\pm 1\sigma$ and propagate to the background prediction
- Propagate all relevant statistical uncertainties

Systematic source	Uncertainty in transfer factor [%]			
	$\mu + jets \Rightarrow t\bar{t}/W$	$\mu + jets \Rightarrow Z \to \nu\bar{\nu}$	$\mu\mu + jets \Rightarrow Z \to \nu\bar{\nu}$	$\gamma + jets \Rightarrow Z \to \nu\bar{\nu}$
Corrections applied to simulation:				
Jet energy scale	1–5	1–5	1–5	1–5
b-tag efficiency / mistag	1–5	1–5	1–5	1–5
Lepton scale factors	1–3	1–3	1–3	-
Pileup	0–2	0–2	0–2	0–2
Signal trigger efficiency	1–2	1–2	1–2	1–2
Muon trigger efficiency	2	2	2	-
Photon trigger efficiency	-	-	-	1–2
Top quark p_T	1–10	1–30	1–10	-

Closure tests

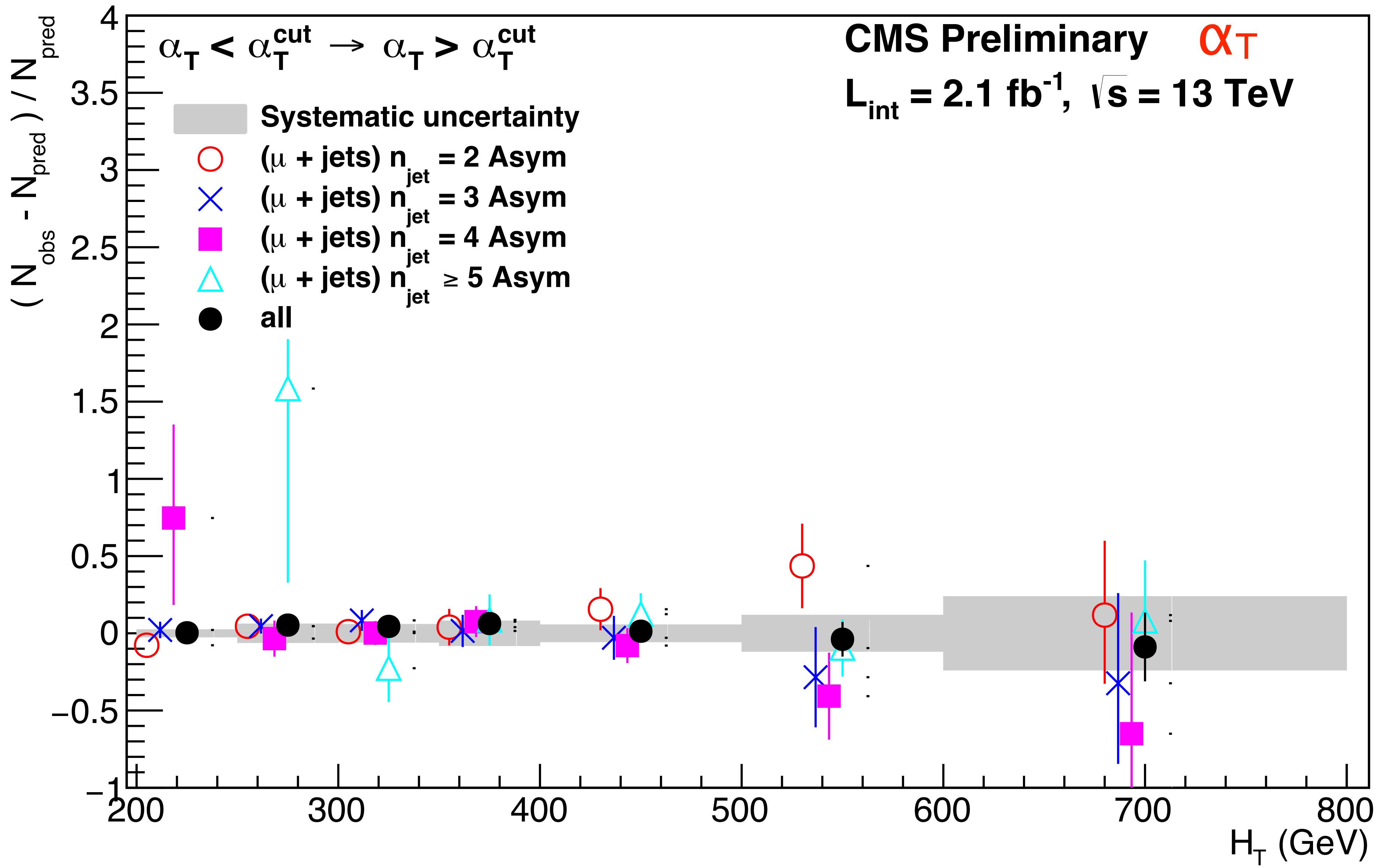
- Additional **data driven** tests that do not rely on simulation
- Confront data yield in control (sub-)sample against the prediction from another (sub-)sample (binned in n_{jet}, H_T)
- e.g. testing the α_T modelling with the single muon control sample (per category):

$$N_{pred}^{\alpha_T > x} = \frac{N_{MC}^{\alpha_T > x}}{N_{MC}^{\alpha_T < x}} \times N_{obs}^{\alpha_T < x}$$

- Any significant bias or trend (as a function of N_{jet} and H_T) should be evident and can be investigated
- In the absence of bias we can use closure tests to derive systematic uncertainties
- With appropriately chosen closure tests this information can be used to determine data-driven systematic errors

Additional data driven systematic uncertainties

- E.g. use yields of events with low α_T to predict yields of events with high α_T in the single muon control sample
- Tests the modelling of the α_T variable



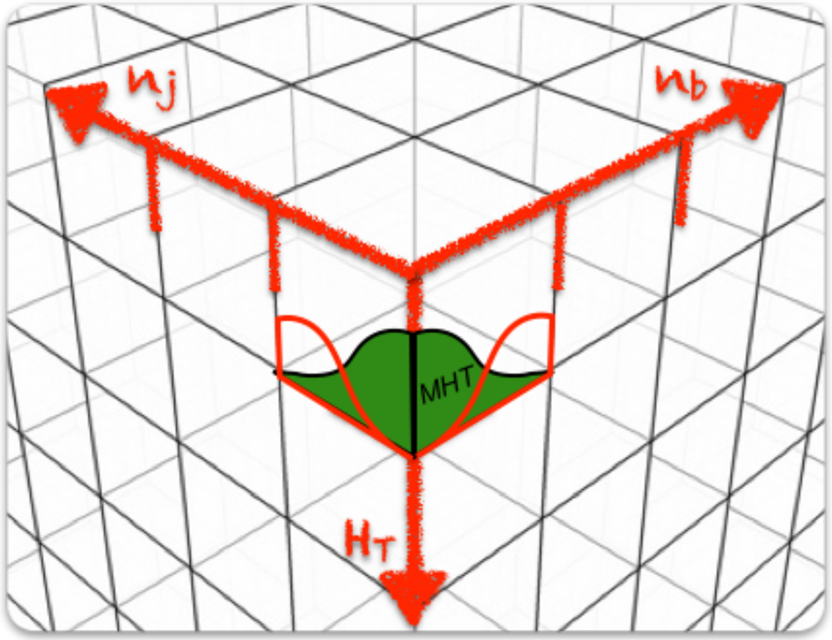
H_T^{miss} dimension

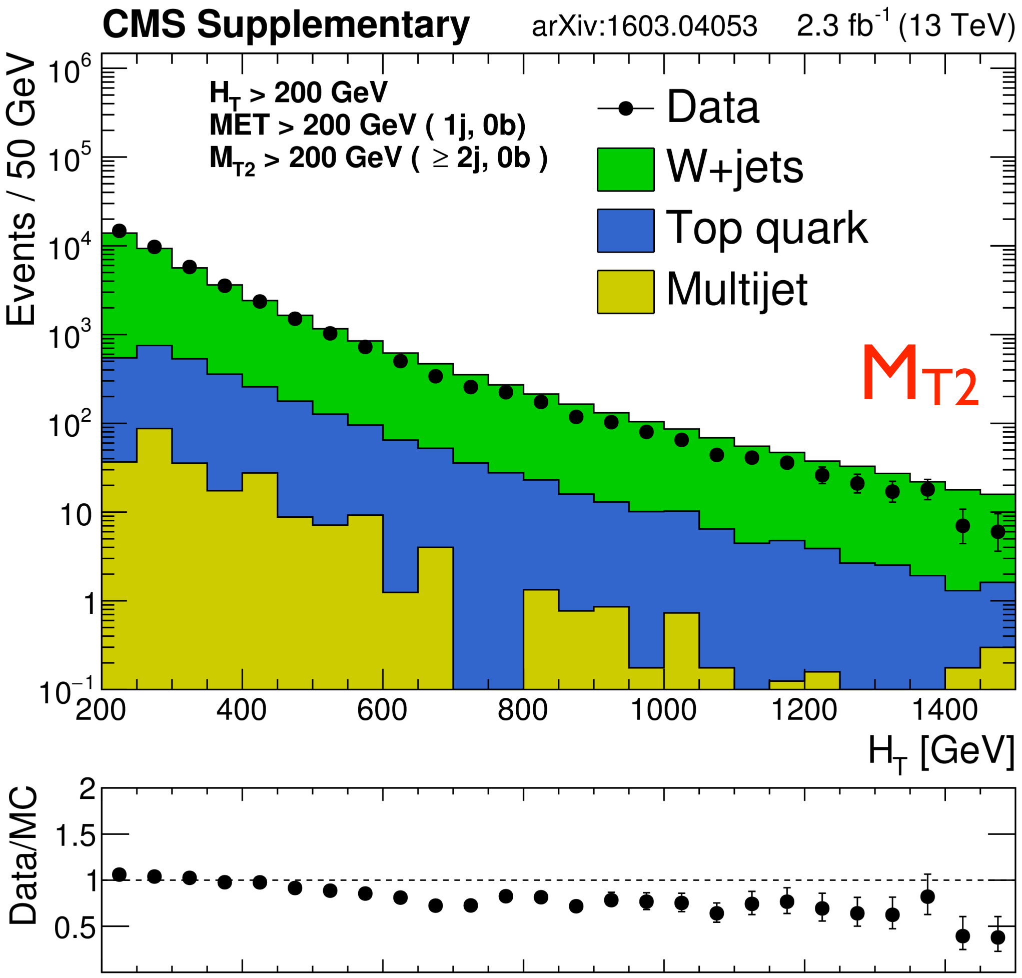
- In addition to the N_{jet}, N_b and H_T dimensions the shape of H_T^{miss} is used to discriminate signal from background

- Generally find that the mismodelling of event yields in the MC occurs as a function of the scale of the event, e.g. H_T

- As each bin is categorised based on these type of variables the H_T^{miss} distribution is taken directly from simulation in each bin

- The normalisation of this distribution comes from the data driven methods

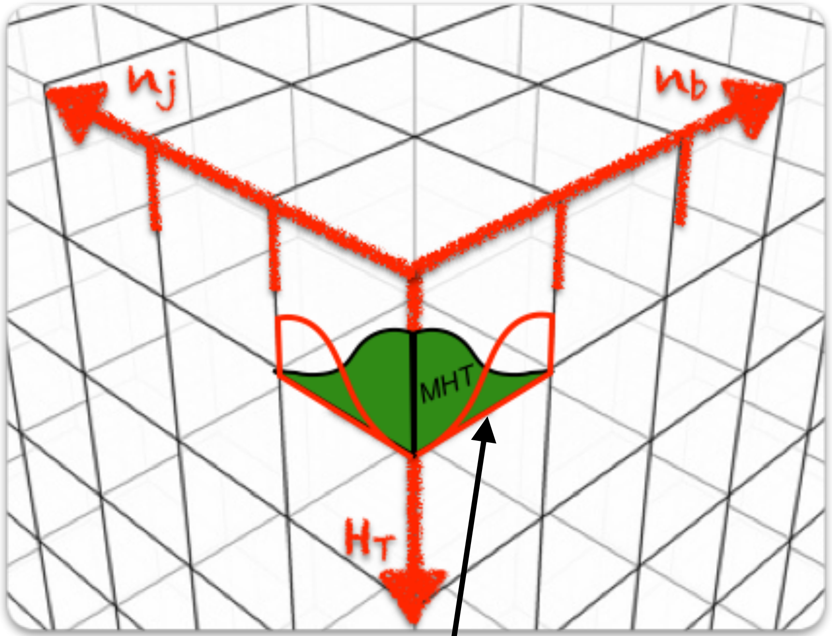




http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-003/index.html

H_T^{miss} dimension systematics

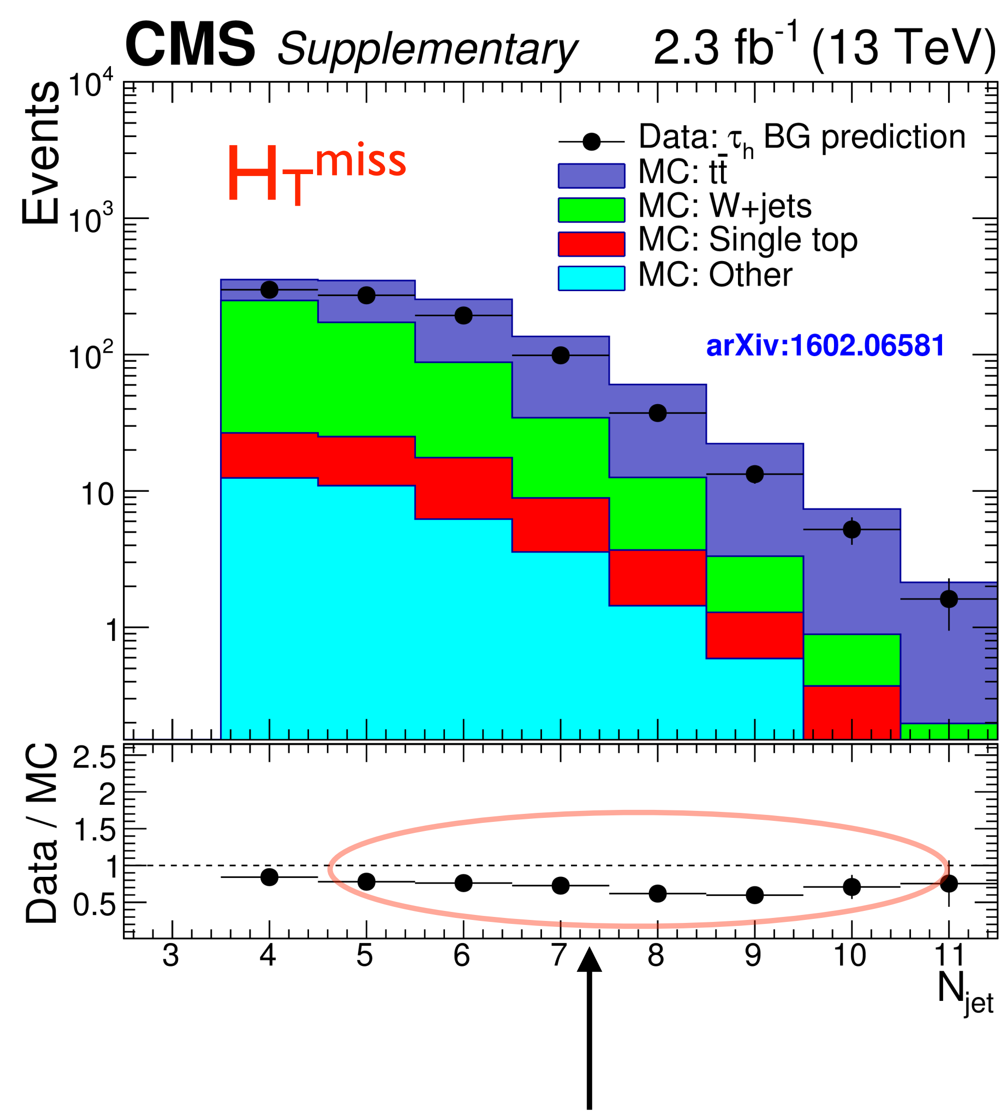
- In each (N_{jet},N_b,H_T) bin, the data/MC ratio is validated in multiple control regions and used to derive uncertainties in the H_T^{miss} shape
  - Fit linear orthogonal polynomial and check linear term is compatible with 0 over all bins
  - Propagate the fit uncertainties
- This data-driven systematic is included in addition to effect on H_T^{miss} of all known experimental and theoretical uncertainties
- Find that the fit is compatible with flat across control samples only when binning in N_{jet} and H_T



Data ~ MC

ISR modelling reweighting for signal models

- Major mismodelling effects in SM backgrounds are reduced with data driven methods
- Still want to correct the modelling of ISR in the MC for the signal models
- Define $N_{ISR\text{-}jets}$ as the number of jets not matched ($\Delta R<0.3$) to MC truth particles
- Derive corrections as a function of $N_{ISR\text{-}jets}$
  - e.g. in an inclusive 2-lepton ttbar sample and check in 1-lepton sample
- Propagate uncertainties in the derivation of the reweighting

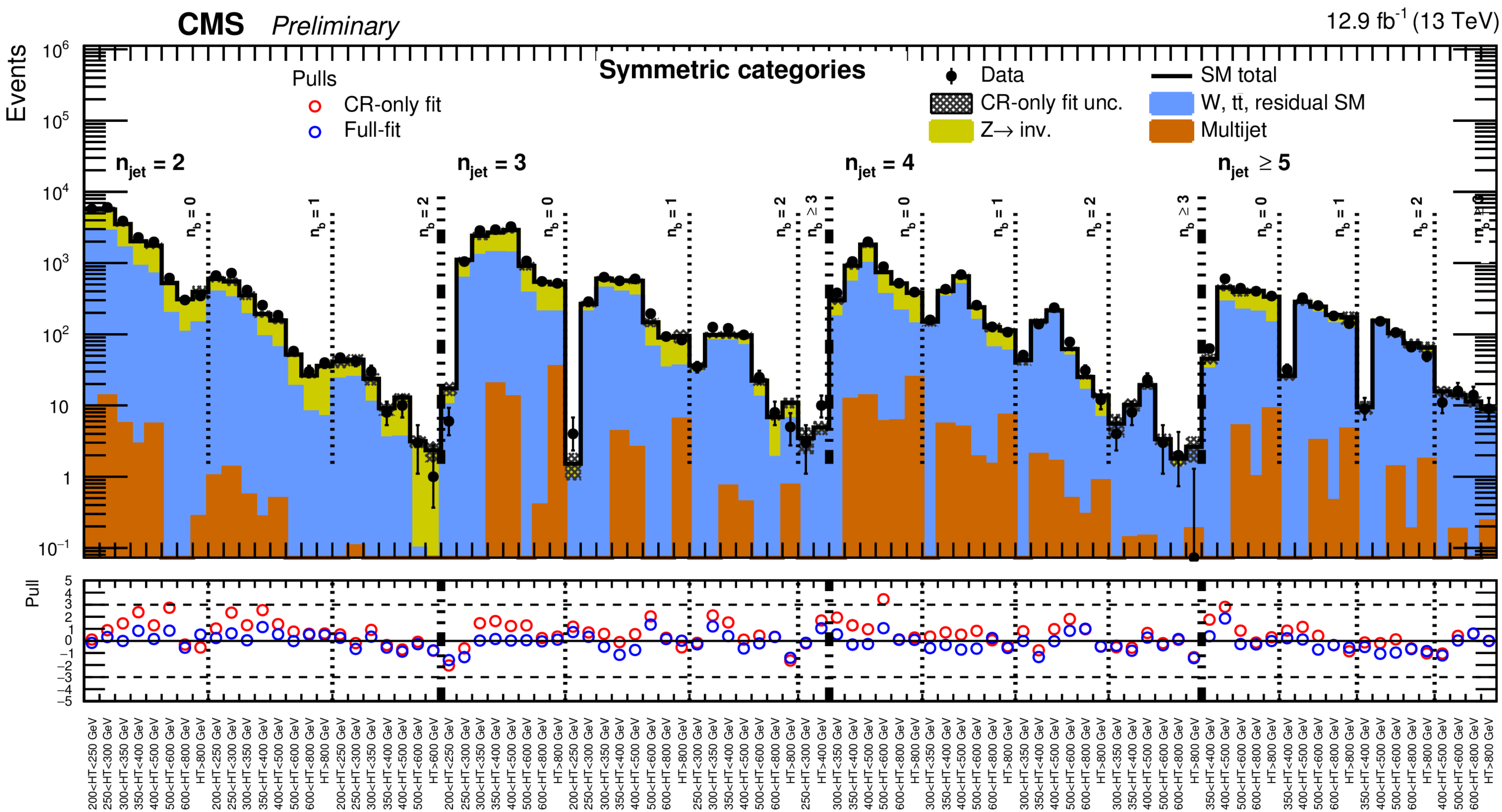


Corrections fix this kind of mismodelling in signal models

http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-002/index.html

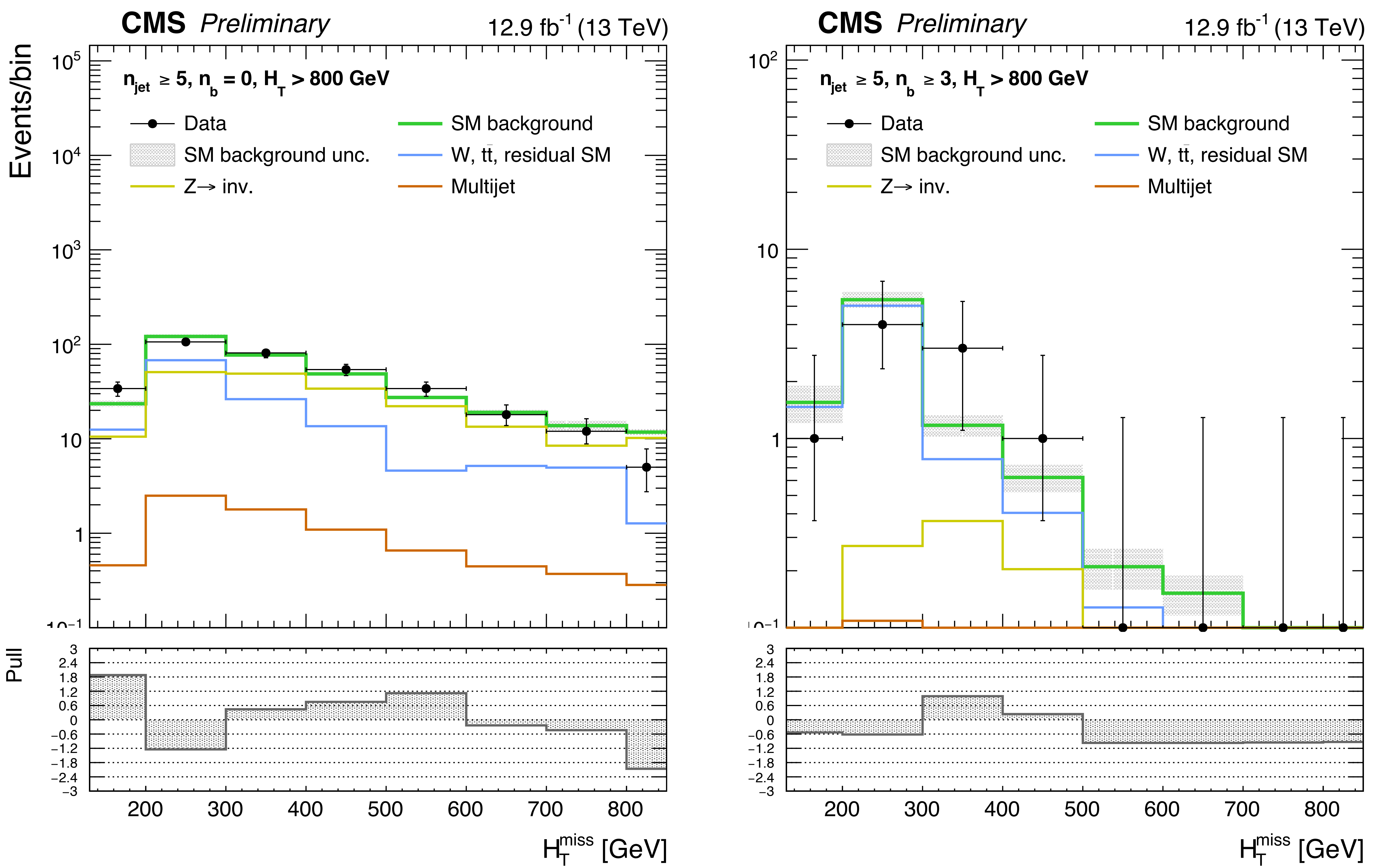
12.9/fb results - HT bins

- A maximum likelihood fit is performed simultaneously across all categories with systematic uncertainties floated as nuisance parameters



http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS-16-016/index.html

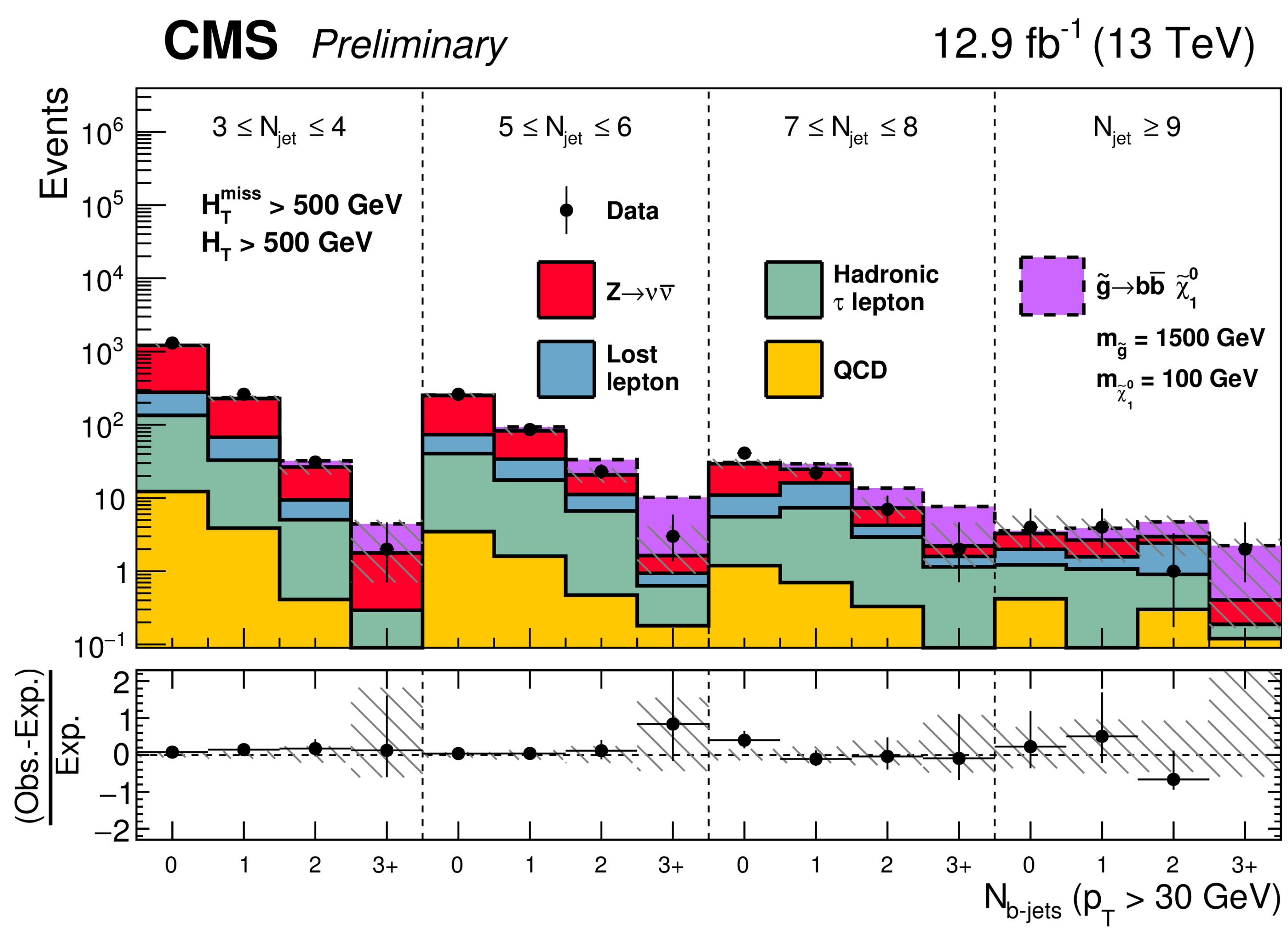
12.9/fb results - MHT dimension



http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS-16-016/index.html

Examples of other BSM+jets analyses at CMS

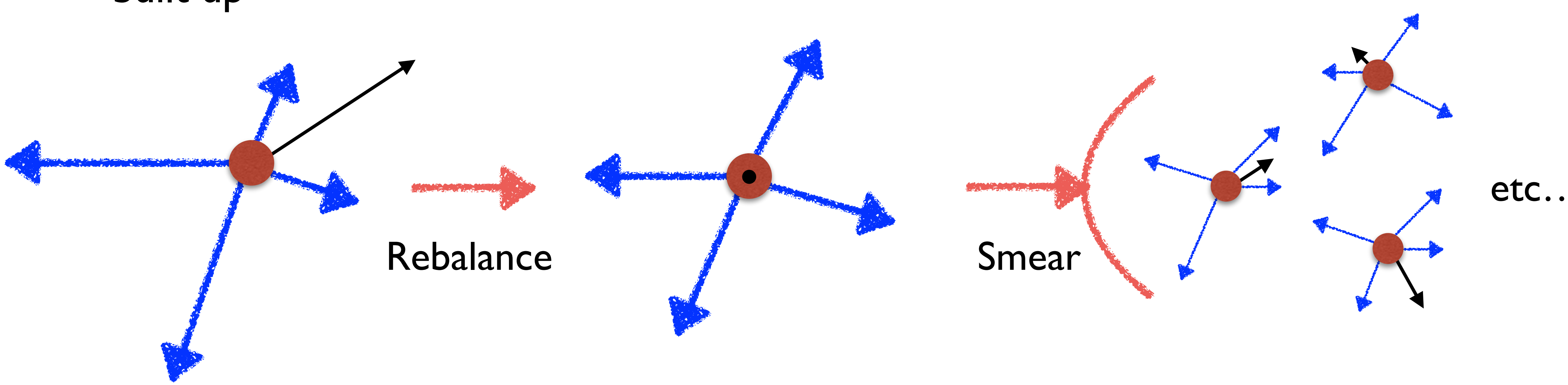
- H_T and H_T^{miss} search
  - A canonical jets+E_T^{miss} search
- **Search variable: H_T^{miss}**
- **Binned** in jet and b-jet multiplicity
  - **N_{jet}**: 3-4, 5-6, 7-8, 9+
  - **N_b**: 0, 1, 2, 3+
  - In each of these bins bin in H_T and H_T^{miss}

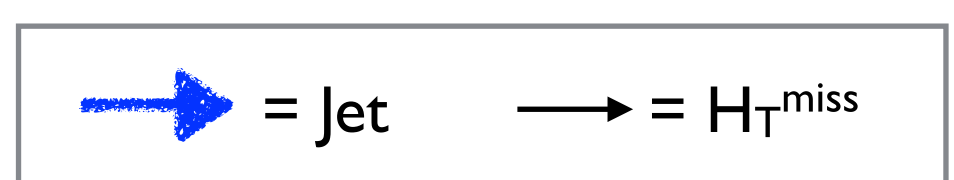


http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS-16-014/index.html

Alternative QCD background method: “rebalance and smear”

- Technique to gain more statistics for QCD events with E_T^{miss} that comes from mismeasurement
- Multi-jet events are collected with low threshold prescaled triggers and “rebalanced” by varying the jet p_T to obtain an event with minimal E_T^{miss}
  - Carried out with a maximum likelihood fit that takes into account the jet response function
- The rebalanced events are then smeared by varying the p_T of all the jets by a random number taken from the jet response function
- Each event is smeared multiple times and allows a large statistics sample to be built up

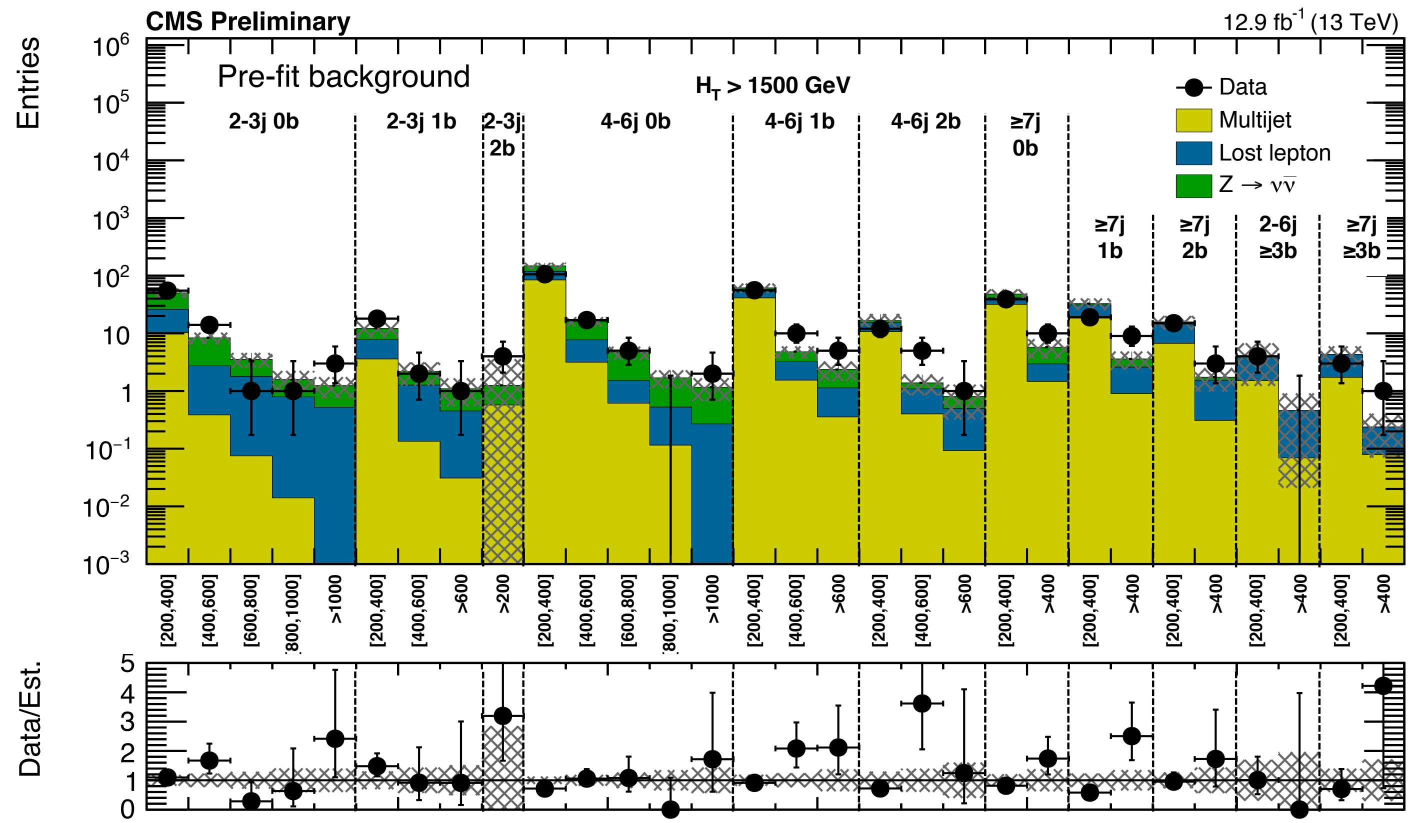


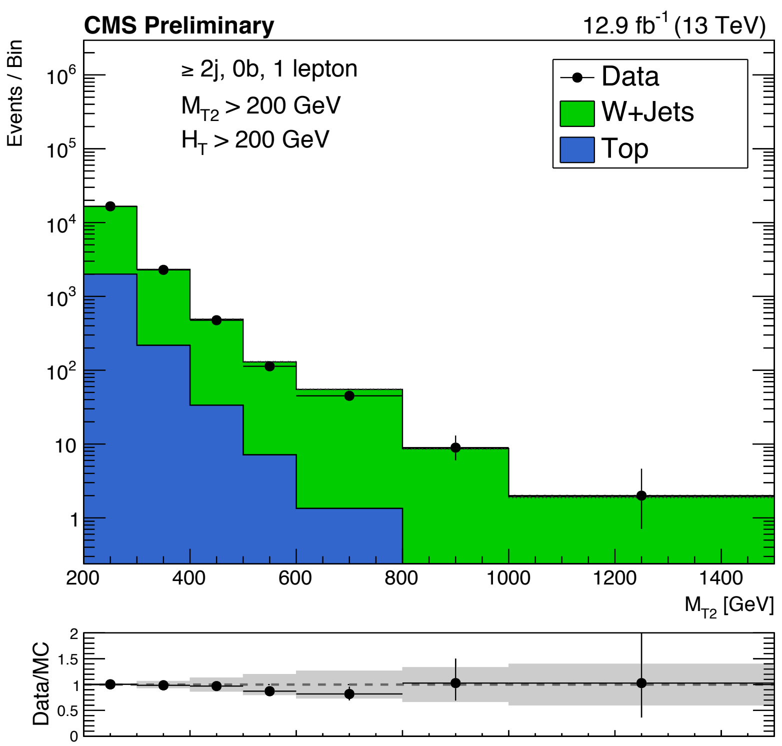


10.1007/JHEP06(2014)055

Examples of other BSM+jets analyses at CMS

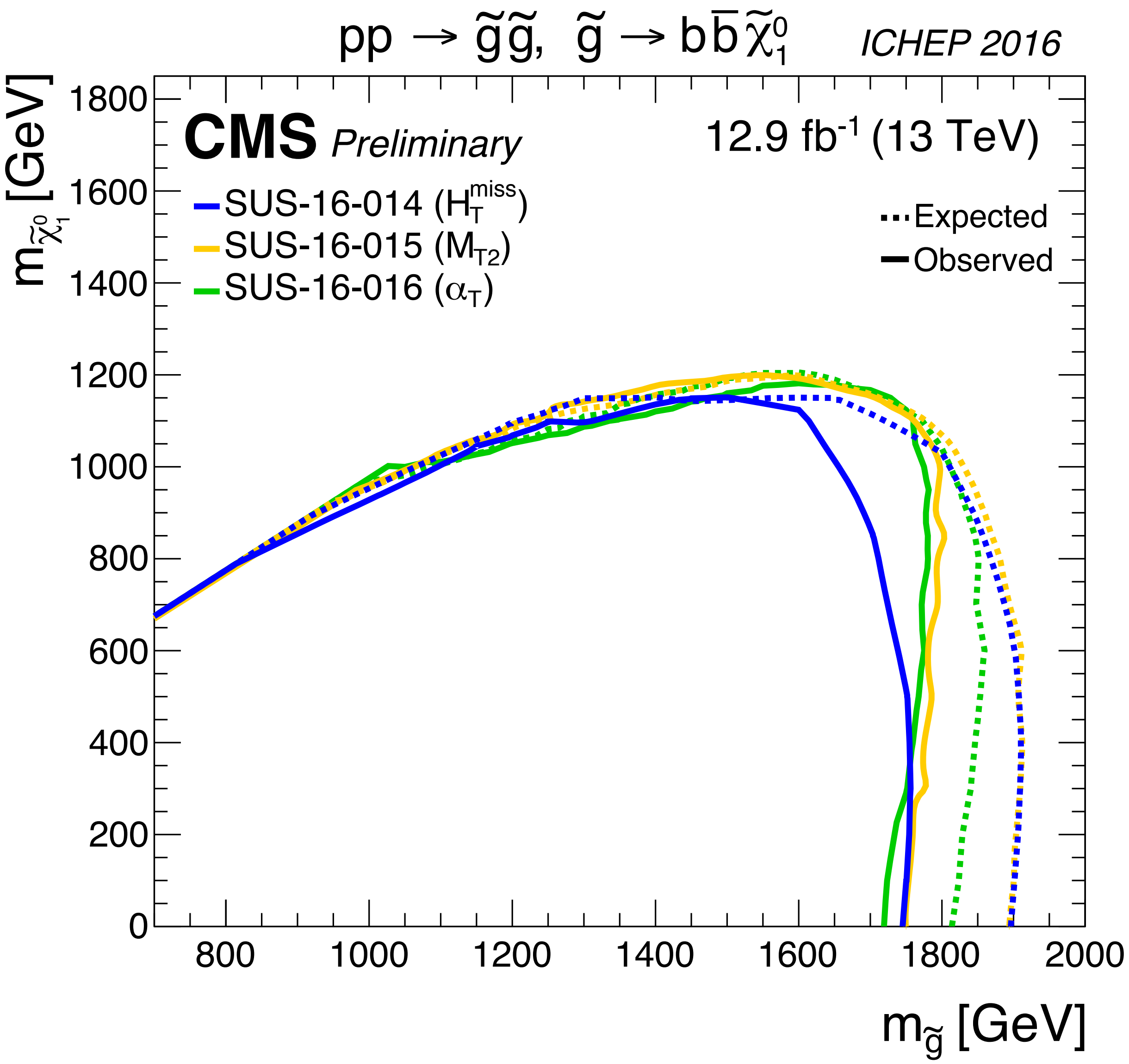
- M_{T2} search
  - Optimised for pair-produced new physics with WIMPs
- **Search variable:** M_{T2} ('stransverse mass', E_T^{miss}-like)
- **Binned** in H_T, N_{jet} and N_b
  - H_T: 200, 450, 575, 1000, 1500+ GeV
  - N_{jet}: 1, 2-3, 4-6, 7+
  - N_b: 0, 1, 2, 3+
- In each of these regions look at the tails of M_{T2}

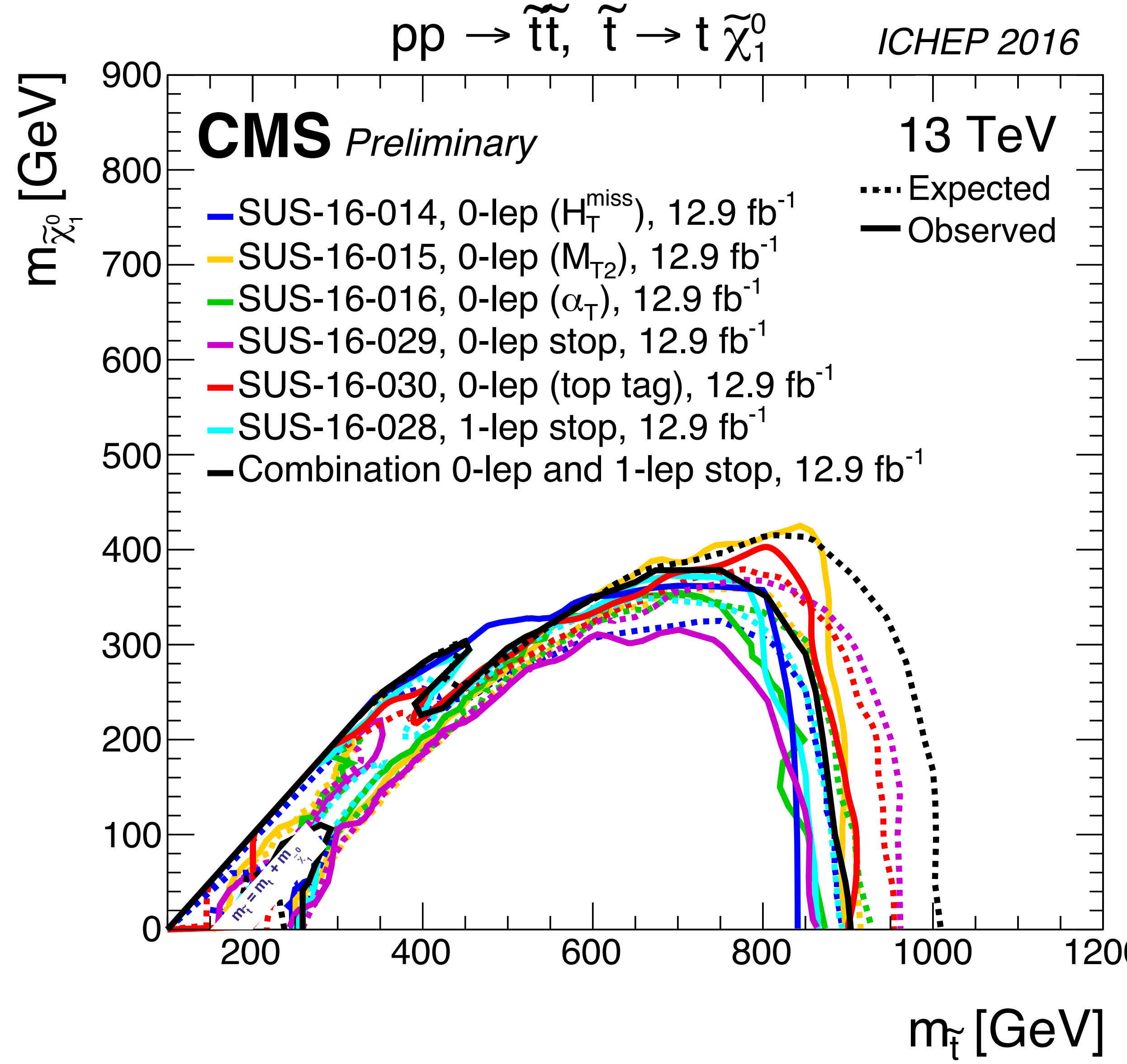


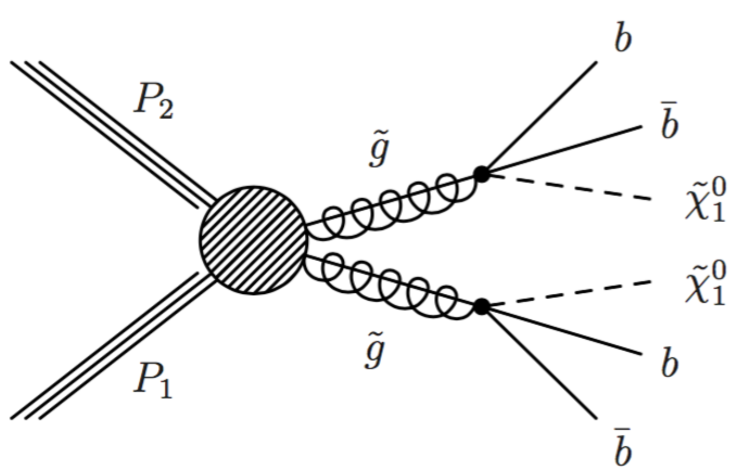


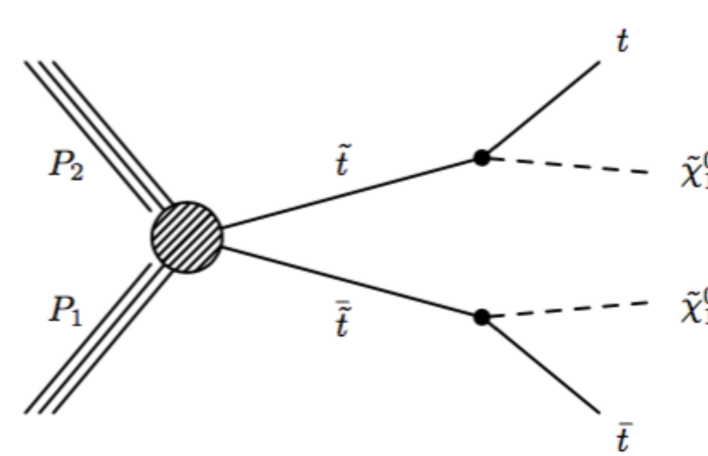
http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS-16-015/index.html

SUSY results



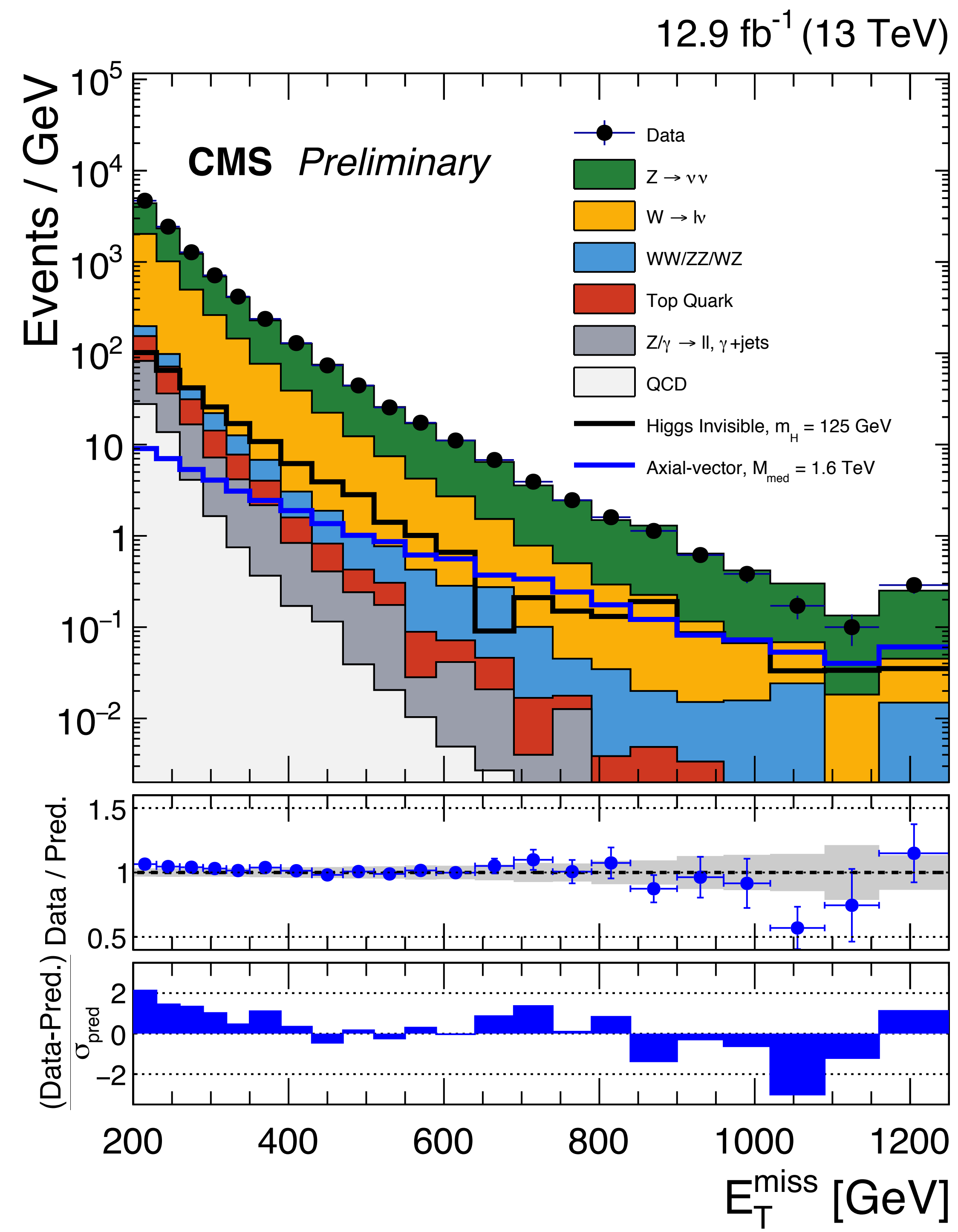






Examples of other BSM+jets analyses at CMS

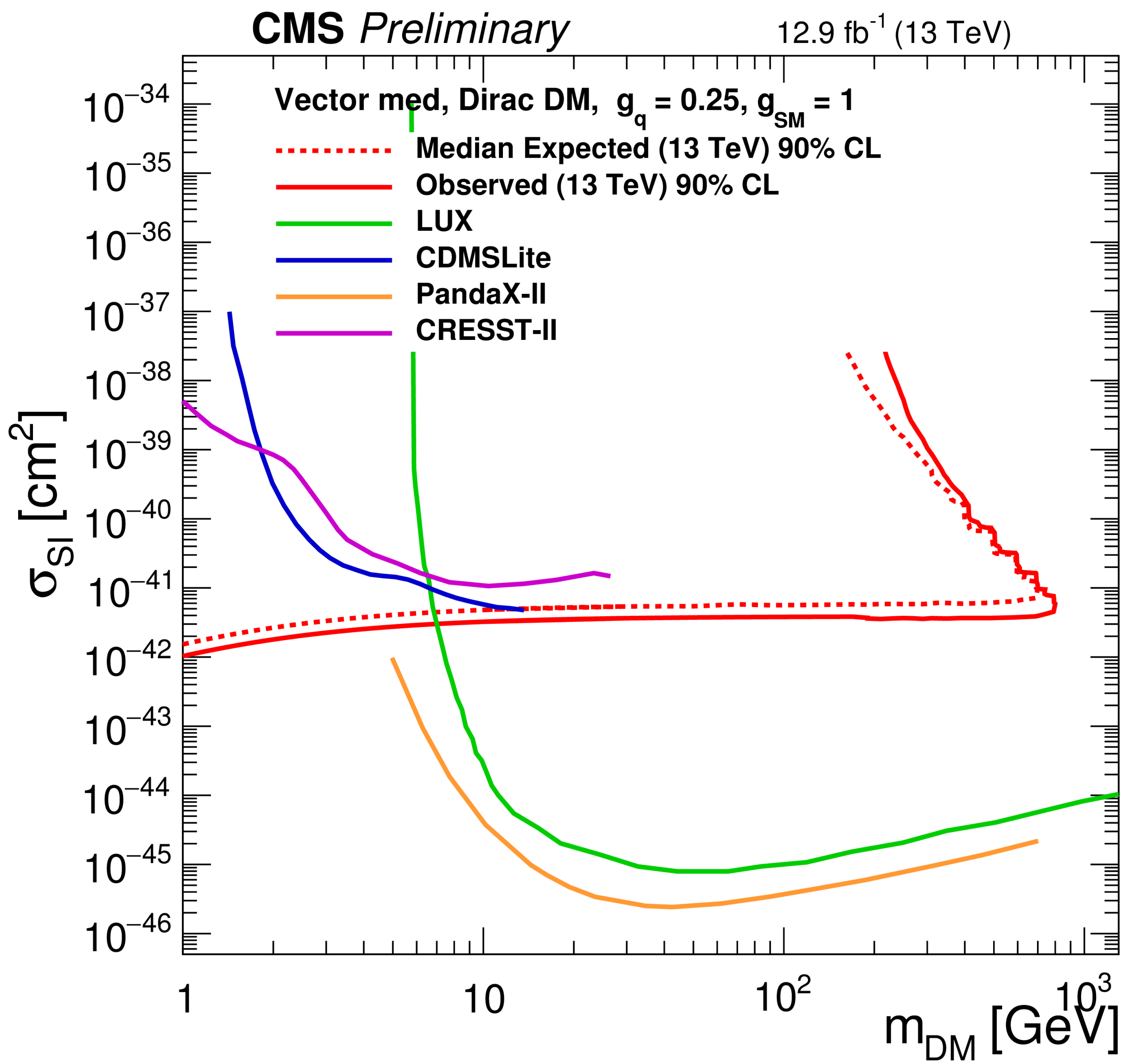
- "Mono-jet" analysis
  - A search for invisible systems with ISR
- **Search variable: E_T^{miss}**
  - Require E_T^{miss} > 200 GeV
- Require:
  - >= 1 ak4 jet with pT>100GeV
  - or >=1 ak8 jet with mass 65-105GeV
- Fit background and signal predictions to E_T^{miss} in data

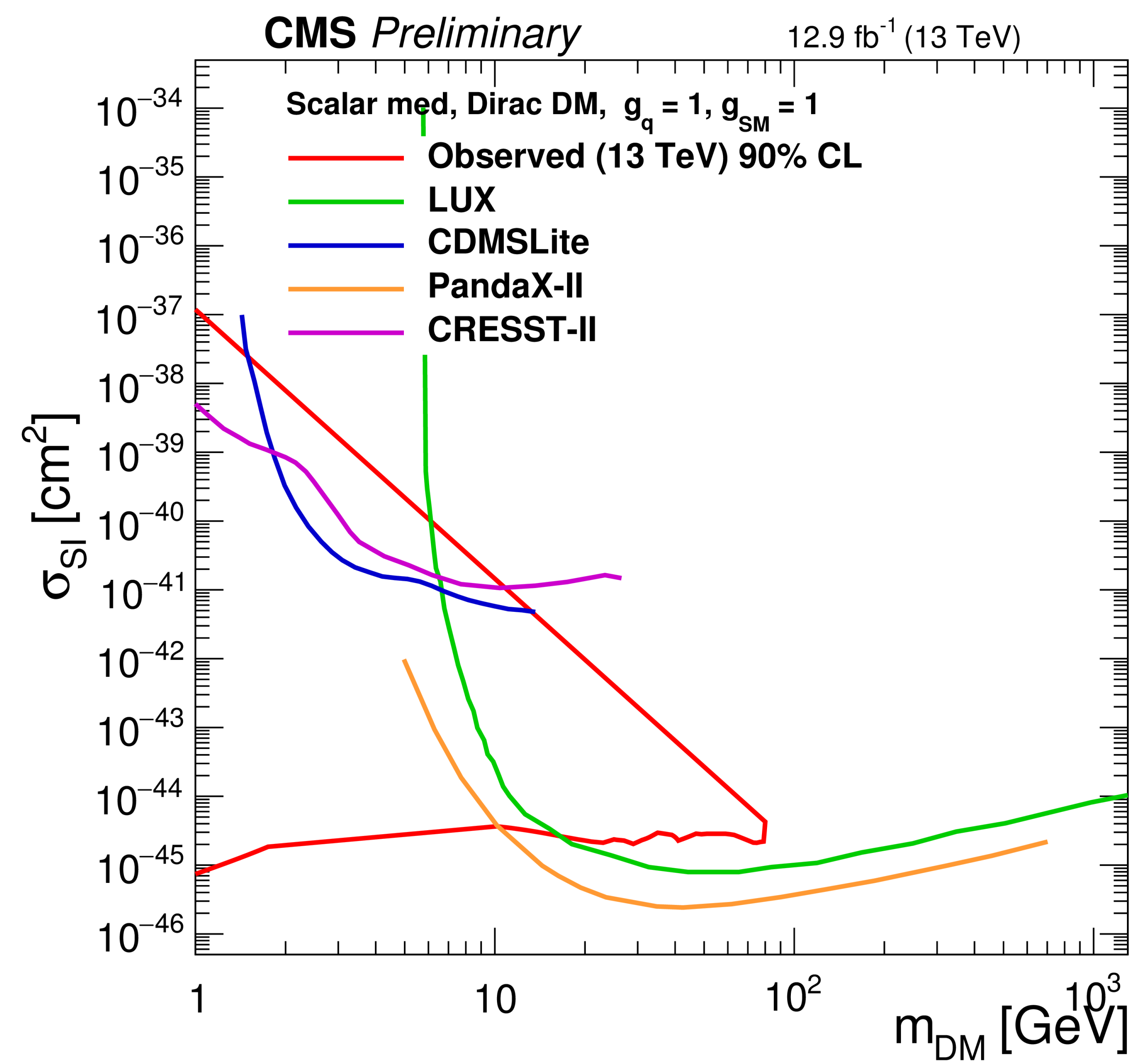


http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/EXO-16-037/index.html

Dark matter results

- Current public results from the mono-jet analysis
- Accelerators provide complimentary reach with direct detection experiments





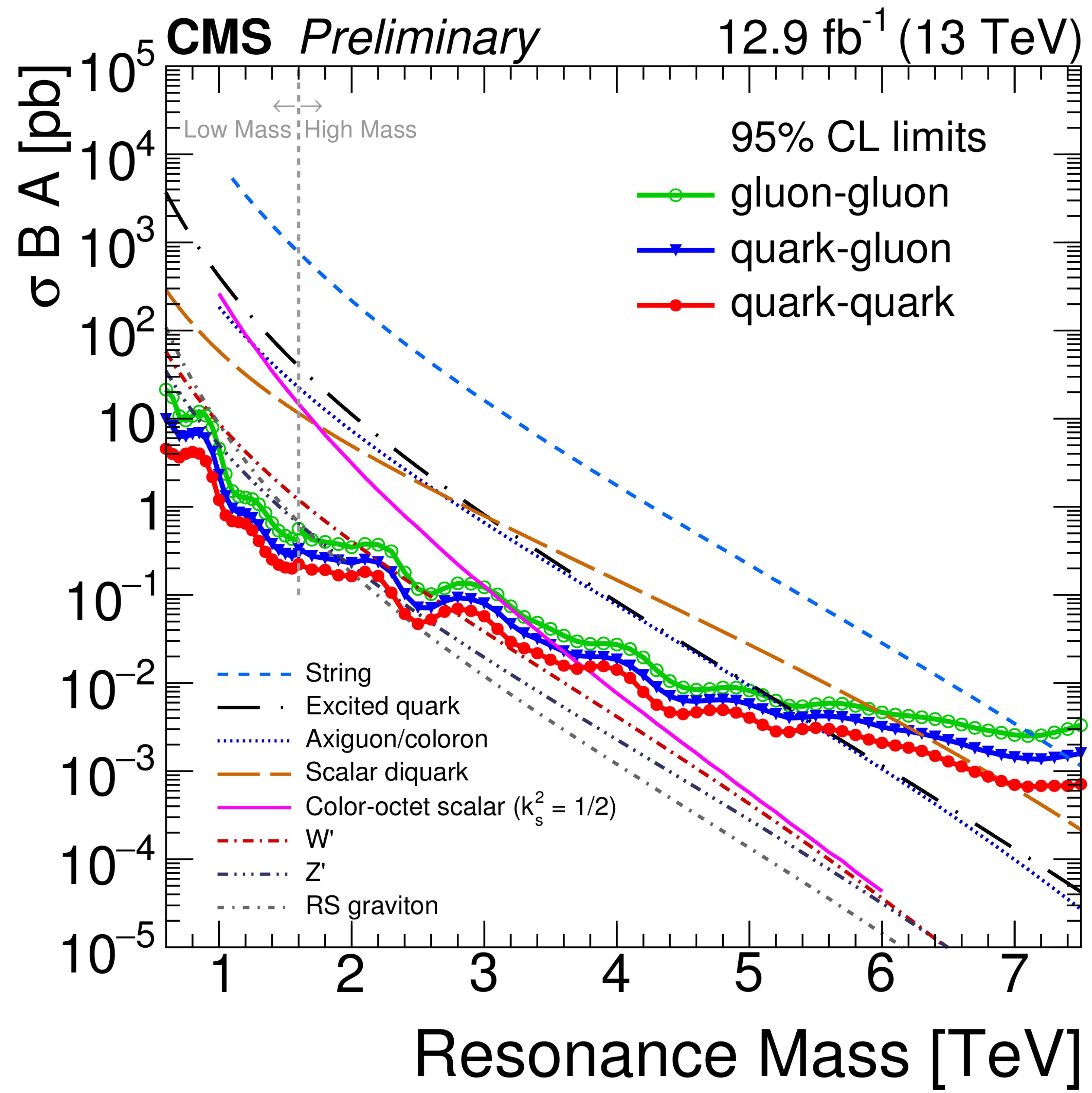
Conclusions

- BSM searches with jets are critical for us to explore the unknown beyond the standard model
- A big challenge is presented by very high QCD multi-jet cross-sections which is typically controlled with dedicated variables
- Standard model backgrounds with genuine missing energy, and their systematic uncertainties, are predicted with data driven methods wherever possible
- In current SUSY and DM analyses **no excess** of events above the predicted Standard Model backgrounds is observed
- The picture will become clearer as Run 2 progresses
- Fingers crossed for a glimpse of something new!

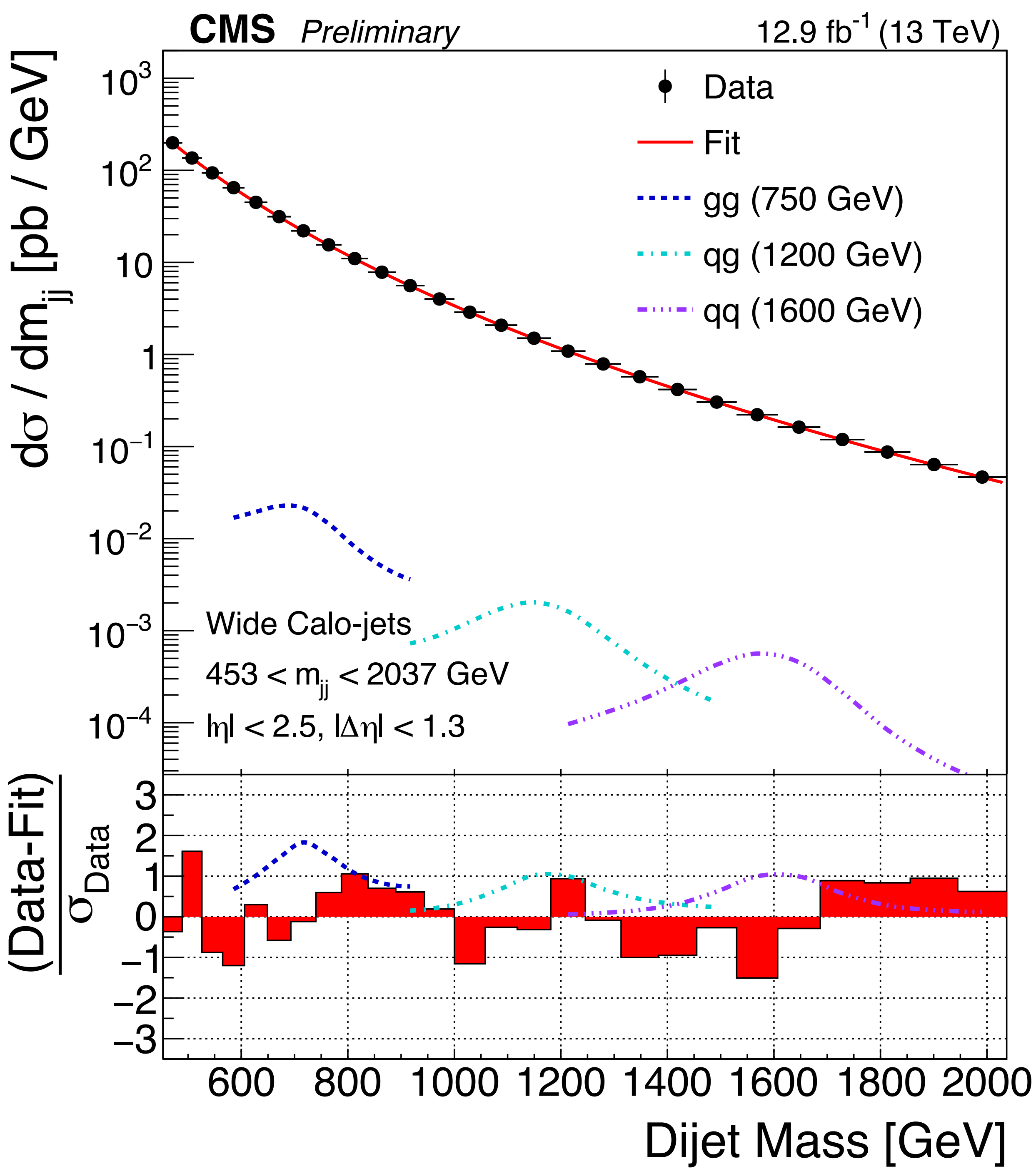
Backup

Di-jet resonance search - latest results

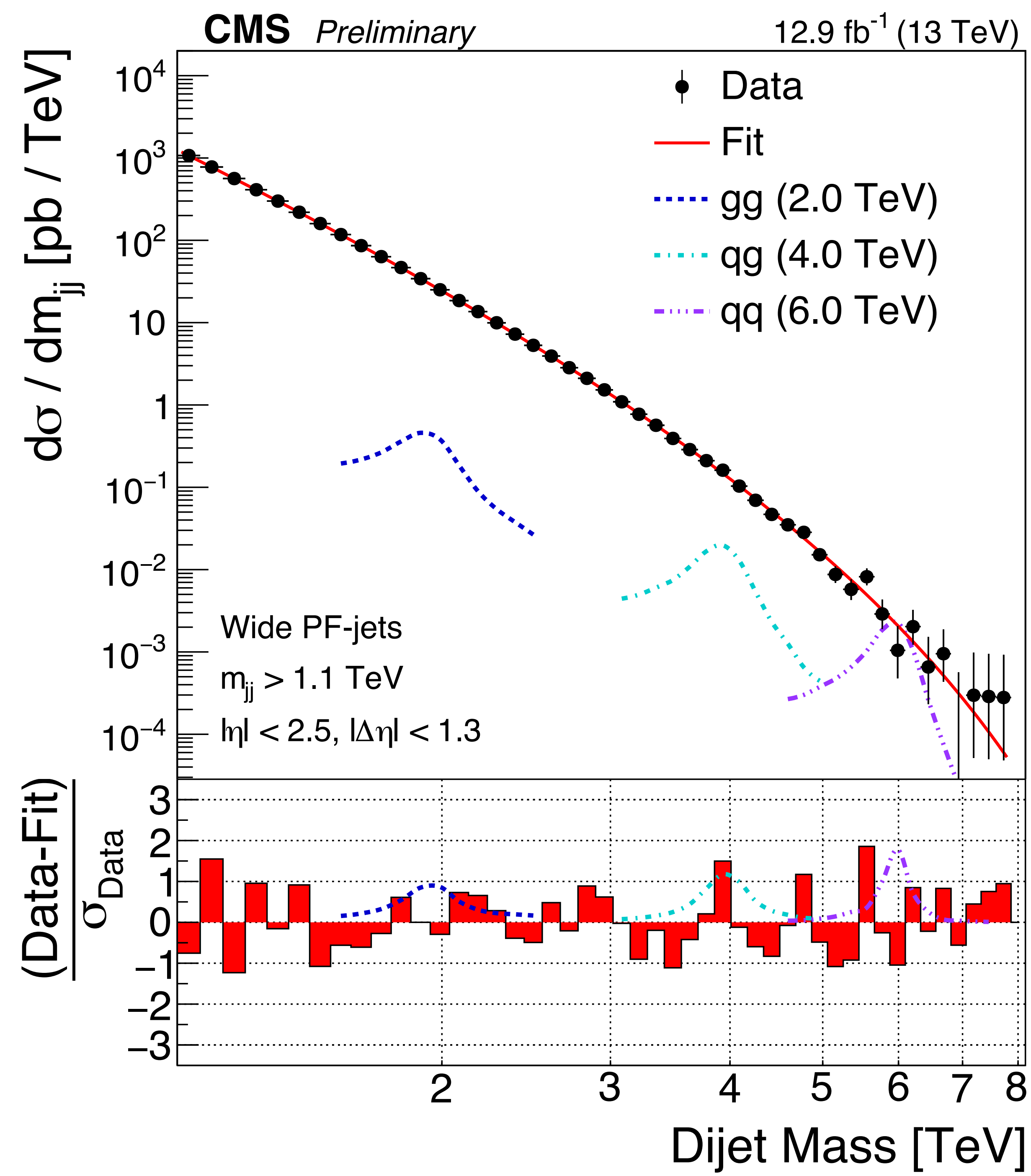
- Di-jet resonance search fits smoothly falling multi-jet background and hunts for resonances above the background
- No statistically significant deviations from background observed in latest Run 2 data



Di-jet resonance search - latest results

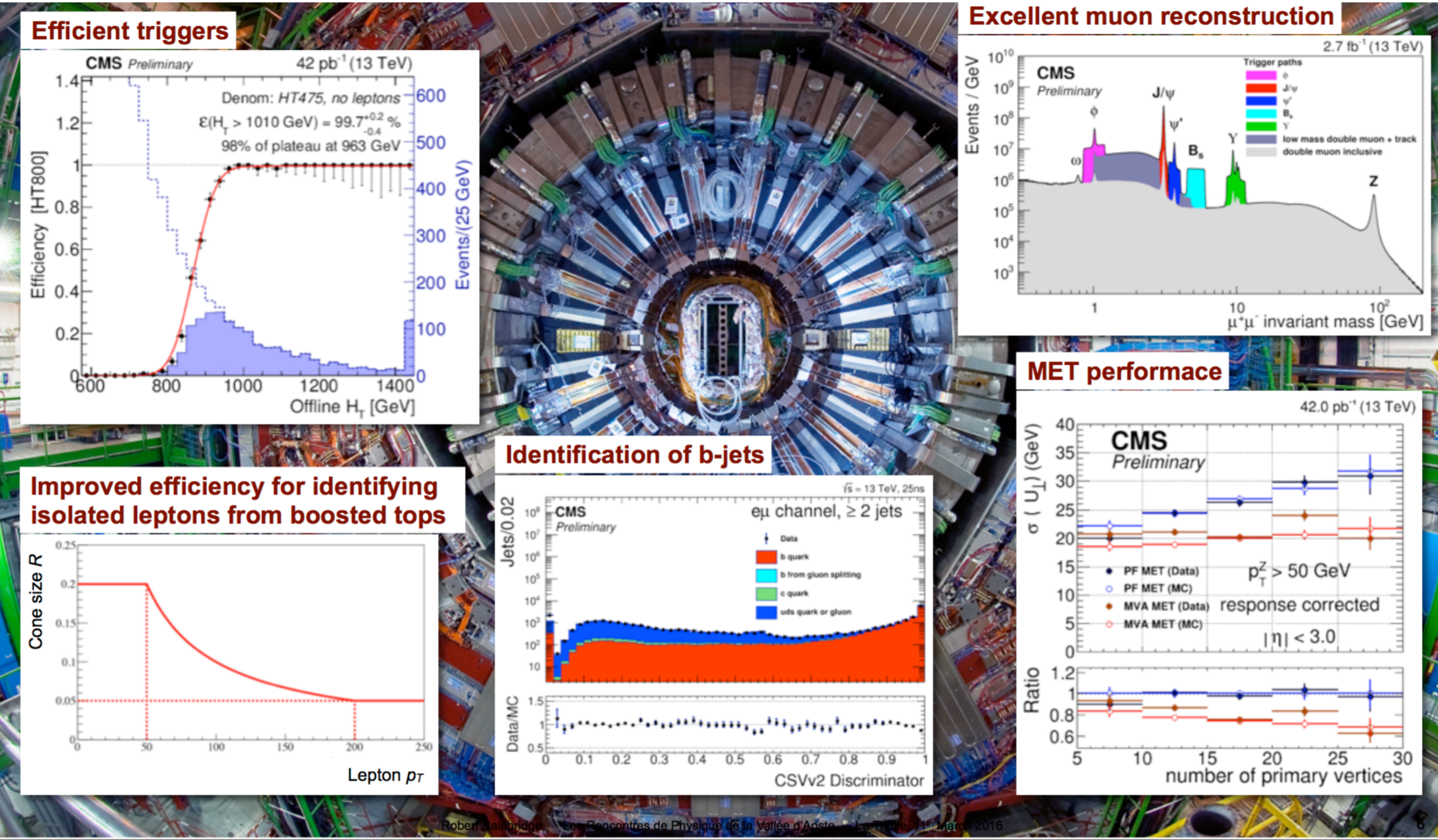


Low mass analysis



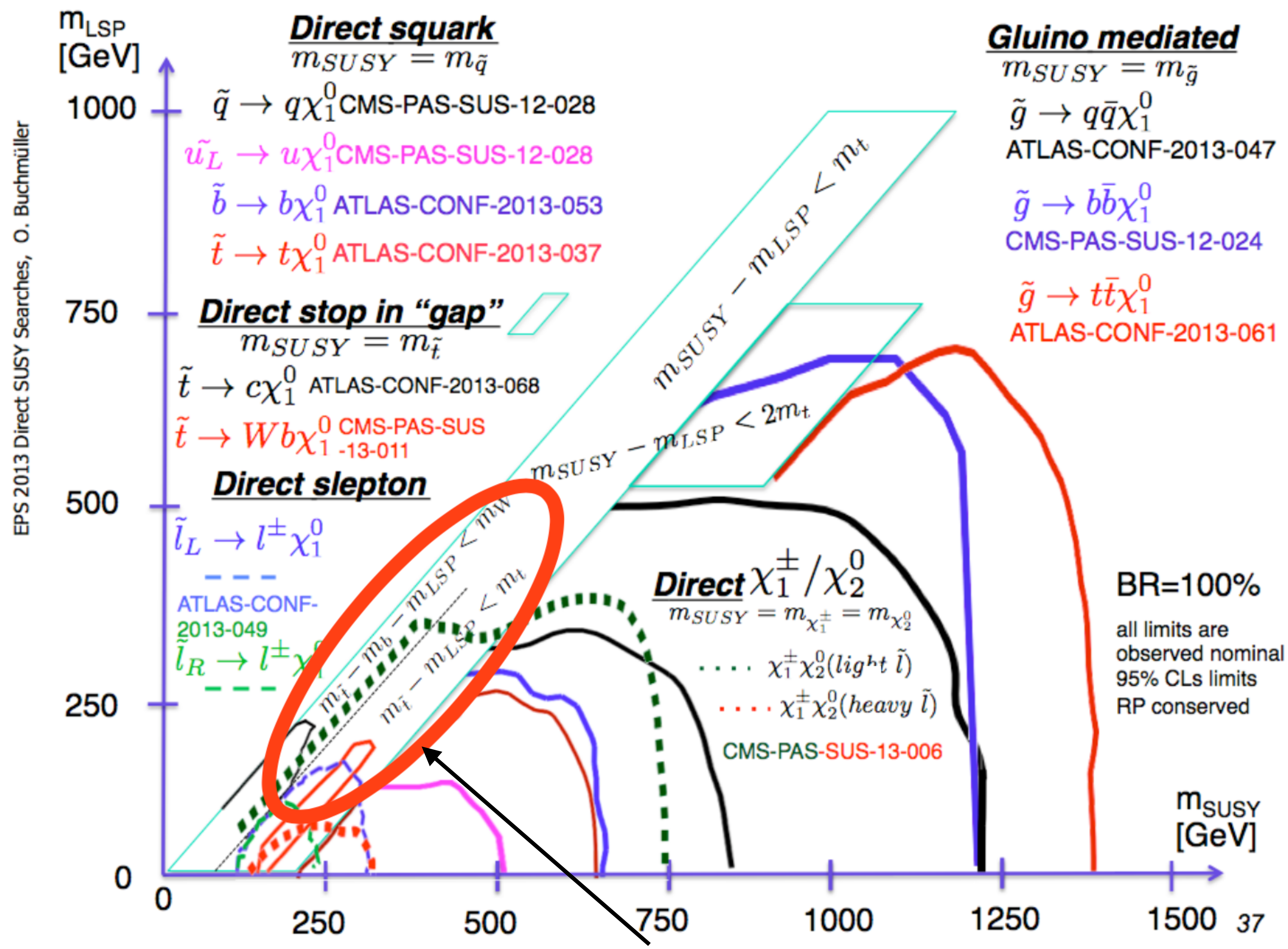
High mass analysis

CMS for SUSY searches



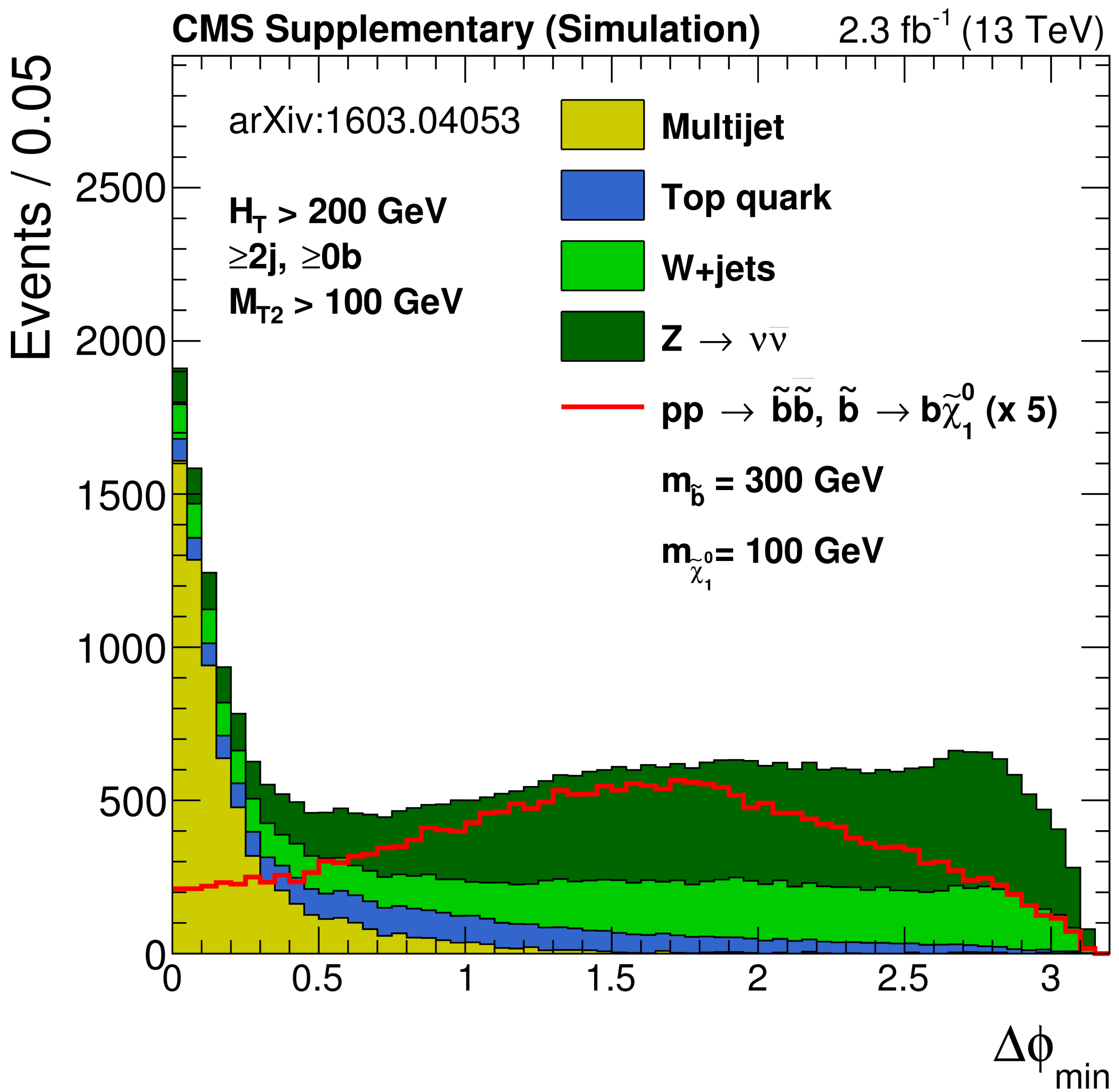
Compressed SUSY

- If SUSY is manifest with small mass splittings we have a compressed spectrum
- This results in decays to soft SM particles and only a small proportion of the MET is visible
- Challenging region that RA1 is well suited to



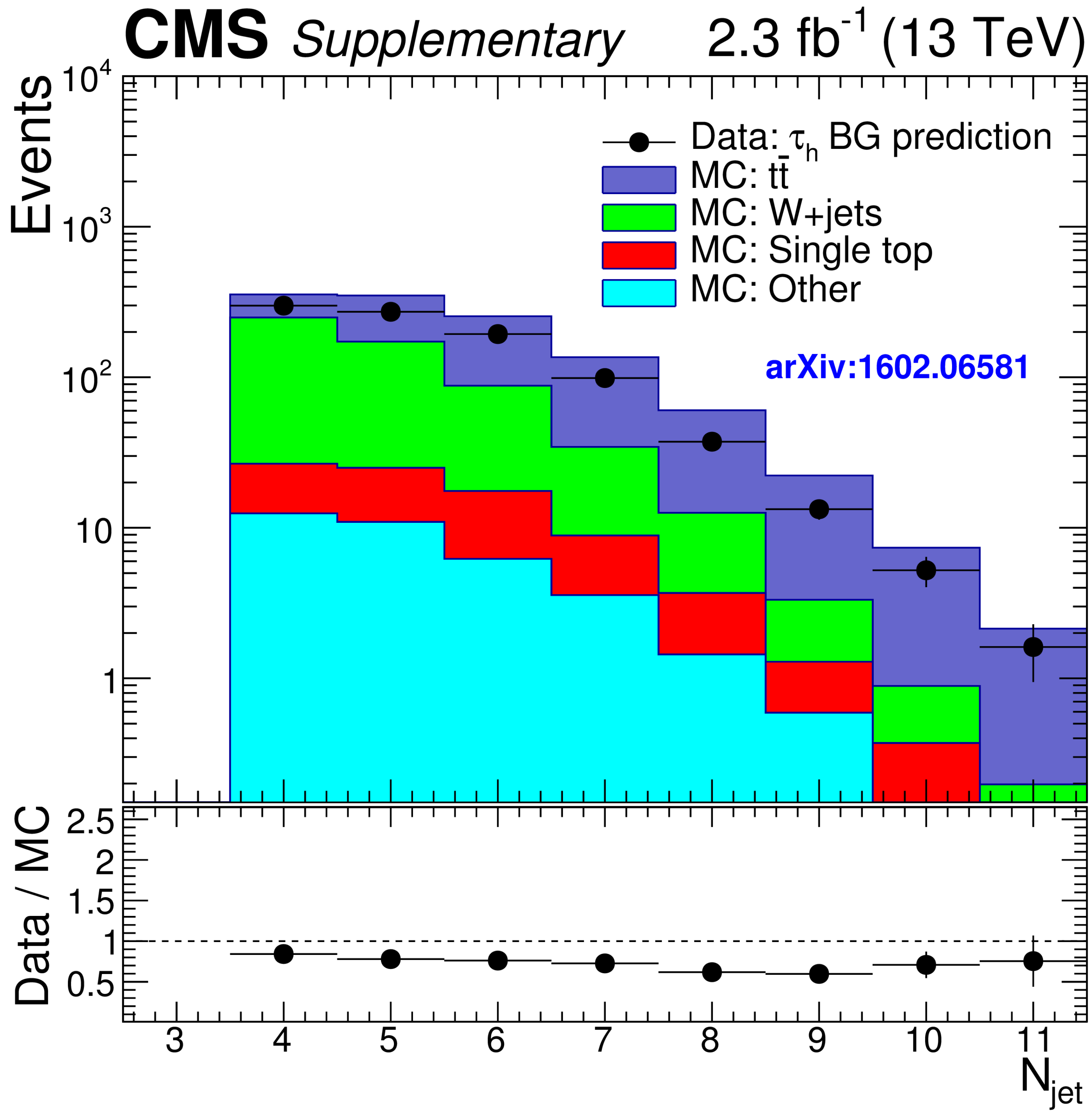
Limits don't extend above $m_{LSP} \approx 300$GeV on the diagonal

Δϕ distribution (MT2 analysis)



http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-003/index.html

Njet distribution (RA2b analysis)



http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-002/index.html

QCD background estimation

- Philosophy
  - –QCD is the most difficult background to measure...
  - –Cut hard (AlphaT + dPhi*), assume conservative systematics on negligible contribution

- **QCD estimate**: based on QCD-enriched data sideband, independent per (Njet,HT)
  - –Method relies on QCD-enriched (and signal depleted) sideband: MHT/MET >1.25
  - –Data counts, $\boldsymbol{N^{corr}}$, in sideband corrected to remove non-QCD contribution
    - –New: now use mu+jets control sample to predict the non-QCD background component
  - –Ratio, $\boldsymbol{R_{QCD}}$, of MC QCD events satisfying or failing the MHT/MET < 1.25 requirement
  - –Product of $\boldsymbol{N^{corr}}$ and $\boldsymbol{R_{QCD}}$ gives estimate for QCD contamination vs (Njet,HT)

- **R_{QCD} validation**: ratio from MC is validated in dPhi* data sideband (dPhi* < 0.5)
  - –Construct double ratio R_{QCD}^{data} / R_{QCD}^{MC}: should be unity, independent of dPhi*, Njet, HT

- **QCD in the tails**: check MHT and Nb distributions to ensure no “enrichment”
  - –Distributions taken from MC, integrated over other variables and regions in HT

Likelihood model

i = $H_T/n_{jet}/n_b$ bin

signal strength

$$L(\mu, \underline{\theta}(\mu)) = \prod_i (L^i_{had}(\mu, \underline{\theta}(\mu)) \times \prod_{control} L^i_{control}(\mu, \underline{\theta}(\mu)))$$

nuisances

- **Binned analysis** - L is a product of **poisson terms** of the likelihood of the observation in each bin given the expected counts from background and signal processes (+ systematics)
- Likelihood broken into two parts per $H_T/n_j/n_b$ - had (signal region) and control
    - L^i_{had} contains the signal and background yields per MH_T bin and relevant (shape) systematics on the MH_T distribution within the signal bin
    - $L^i_{control}$ contains the prediction of the dominant background yields by the control regions inclusive in MH_T including the relevant (log normal) systematics
- Signal contribution in signal and control regions (with relevant systematics) included
- Set limits based on value of μ excluded with CLs > 95% derived using LHC test statistic*

*ATL-PHYS-PUB-2011-11

Sideband corrections

- Take highest-precision cross section calculations and determine process-specific corrections relevant for our signal region phase space
    - Improves closure of data-driven checks (minimal impact on predictions)
- Use data sideband defined by MHT (αT for photon) cut in control regions
- Derive corrections with simultaneous fit over sidebands over all HT/njet/nb bins (using a floating parameter per process)
    - Don't need to use cuts on nb/njet to enrich samples

Process	Sideband	Selection	Corrrection
γ + jets	$0.50 < \alpha_{\mathrm{T}} < 0.52$	γ + jets	1.33 ± 0.03
W + jets	$100 < H_{\mathrm{T}}^{\mathrm{miss}} < 130\,\mathrm{GeV}$	μ + jets	1.13 ± 0.01
Z + jets	$100 < H_{\mathrm{T}}^{\mathrm{miss}} < 130\,\mathrm{GeV}$	$\mu\mu$ + jets	0.99 ± 0.02
$\mathrm{t\bar{t}}$ + jets	$100 < H_{\mathrm{T}}^{\mathrm{miss}} < 130\,\mathrm{GeV}$	μ + jets, $\mu\mu$ + jets	0.86 ± 0.01

Full list of signal systematics

Systematic source	Type	Correlated	Typical magnitude (%)
Luminosity	Normalisation	Yes	6.3
Monte Carlo statistics	Norm. + shape	No	1–50
Jet energy scale	Norm. + shape	Yes	3–10
b-tag efficiency scale factors	Norm. + shape	Yes	5–40
Lepton scale factors	Normalisation	Yes	1–5
Pile-up	Norm. + shape	Yes	0–5
Trigger efficiency	Norm. + shape	Yes	0–4
Initial state radiation	Norm. + shape	Yes	1–30
Modelling of $H_{\mathrm{T}}^{\mathrm{miss}}$	Normalisation	Yes	1–5

Example suite of tests

The closure tests are designed to probe the modelling of specific physics effects, several examples are given below

- **W/tt composition**
  - $n_{jet}=2 \rightarrow n_{jet}=3$
  - $n_{jet}=3 \rightarrow n_{jet}=4$, …
- **Lepton ID modelling**
  - Tight-Loose μμ → Tight-Tight μμ
- **Z/gamma ratio**
  - μμ+jets→γ+jets
- **Z→inv with b-tags**
  - μ+1b → μμ+1b
- **Consistency between W,Z bosons**
  - μ+jets→γ+jets
  - μ+jets→μμ+jets
- **W polarisation effects**
  - μ^++jets→μ^-+jets
- **α_T and $\Delta\phi^*$ modelling**
  - “core” → “tail”
- **b-tagging SFs**
  - 0 b-tag→ 1 b-tag
  - 1 b-tag→ 2 b-tags

(in addition to dedicated independent study)

- **Lepton/photon efficiency and SF**
  - μ+jets→μμ+jets
  - μ+jets→γ+jets

Full systematic list with data driven tests

Systematic source	Uncertainty in transfer factor [%]			
	$\mu + jets \Rightarrow t\bar{t}/W$	$\mu + jets \Rightarrow Z \to \nu\bar{\nu}$	$\mu\mu + jets \Rightarrow Z \to \nu\bar{\nu}$	$\gamma + jets \Rightarrow Z \to \nu\bar{\nu}$
Corrections applied to simulation:				
Jet energy scale	1–5	1–5	1–5	1–5
b-tag efficiency / mistag	1–5	1–5	1–5	1–5
Lepton scale factors	1–3	1–3	1–3	-
Pileup	0–2	0–2	0–2	0–2
Signal trigger efficiency	1–2	1–2	1–2	1–2
Muon trigger efficiency	2	2	2	-
Photon trigger efficiency	-	-	-	1–2
Top quark p_T	1–10	1–30	1–10	-
Derived from closure tests in data:				
W/Z ratio	-	4–15	-	-
Z/γ ratio	-	-	-	6–11
W/$t\bar{t}$ composition	4–30	-	-	-
W polarisation	2–10	2–10	-	-
α_T / $\Delta\phi^*_{min}$	3–30	3–30	3–30	-