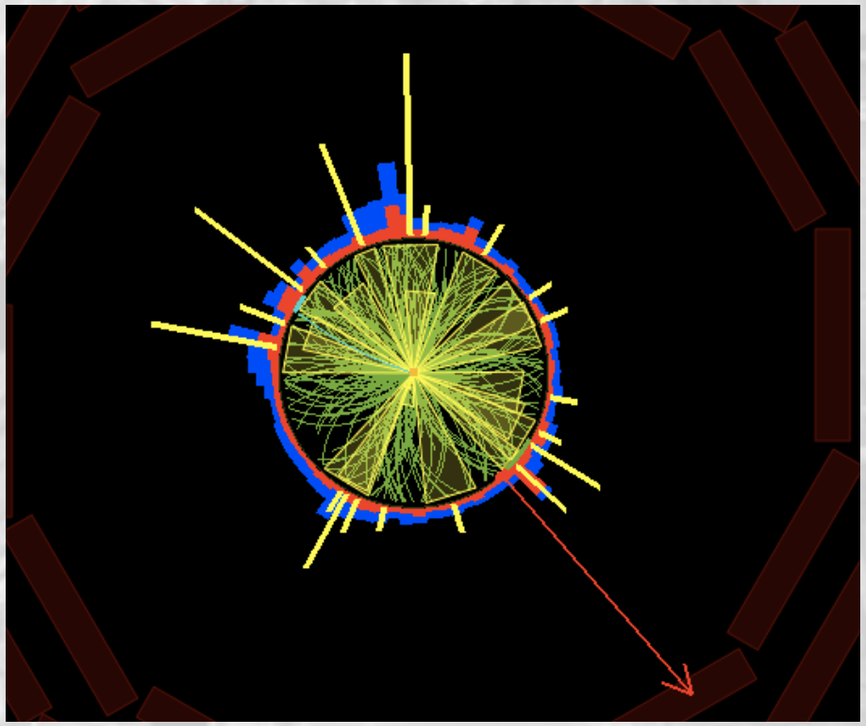
Searches for new physics with multijet events at CMS

Maurizio Pierini
CERN

LHC and Jet Physics

- In a hadronic environment there are easier things than jets to deal with (e.g., isolated leptons or photons)

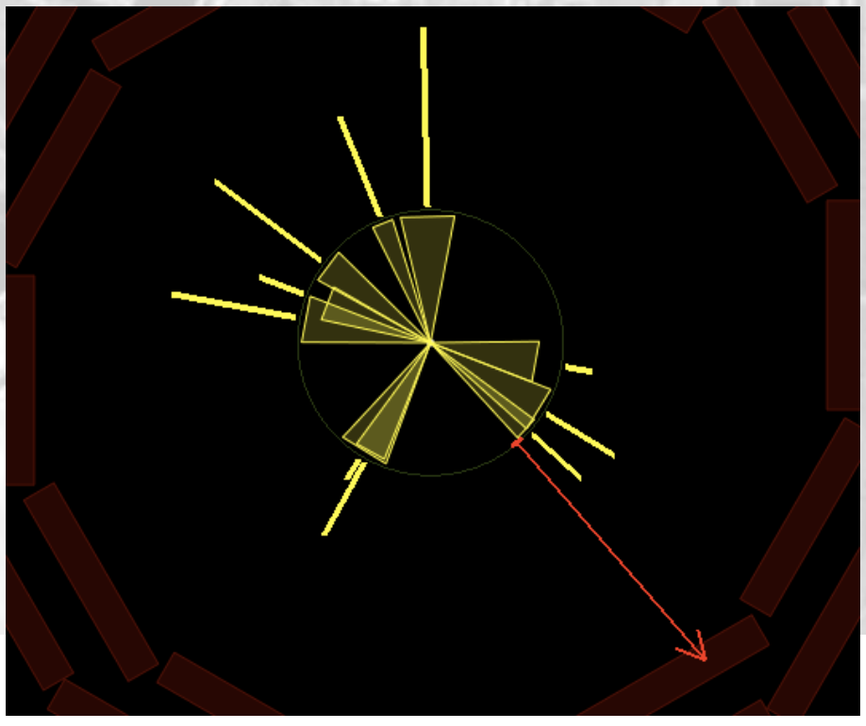
- On the other hand, anything else than jets is expensive (low BRs and/or acceptance). A large fraction of the signal (e.g., signals with W or Z) could be hiding in multijet events



- Even in events with leptons, the signal vs. bkg discrimination could be in the jet kinematic and jet properties

- LHC detectors+theory progresses made our life easier: jet physics is now a kind of precision physics

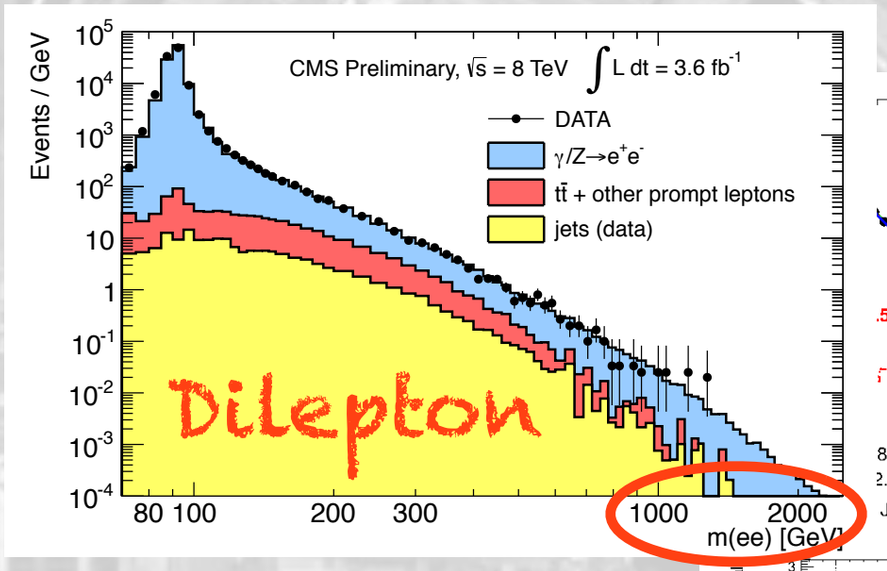
- After all, once we know our detector and we apply some cleanup, life is not that hard...

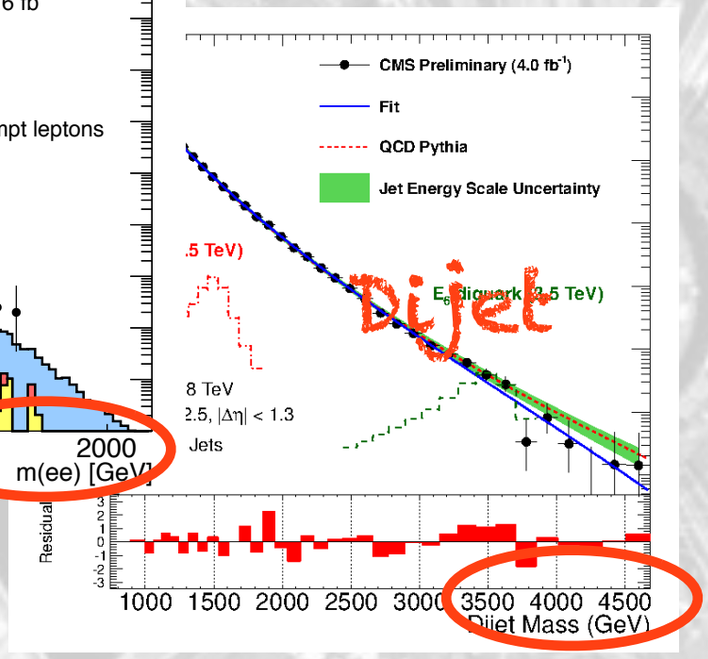


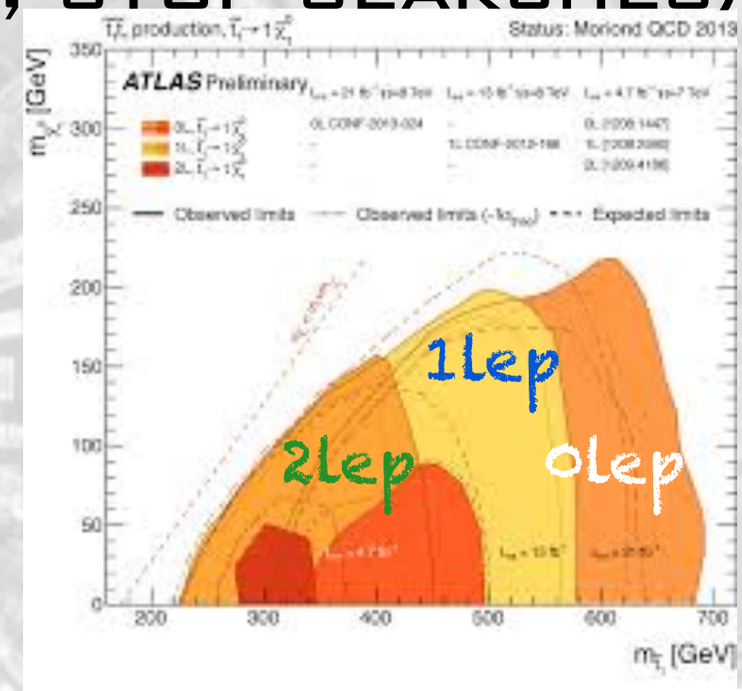
The Advantages

- Multijet events provide complementary information (e.g. strongest limits with less integrated luminosity)

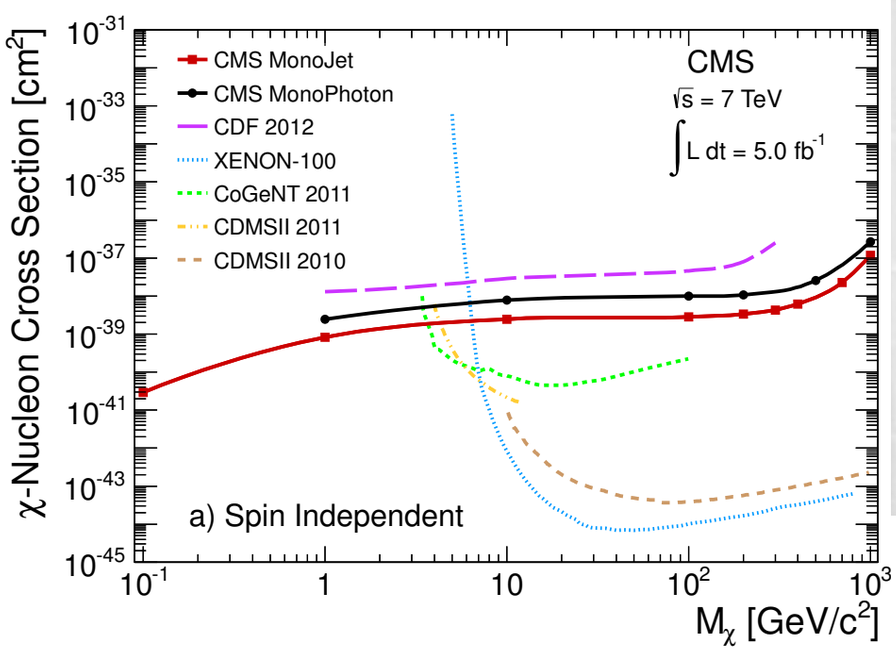
- Often, jets are important for S vs. B discrimination even in presence of leptons (e.g., stop searches)







- Jets are the way to go (e.g. DM search with monojet)



- Bottom line, keeping an inclusive approach cannot be worse, particularly when it is not exactly clear what to look for

Three examples

Three examples of interesting/challenging/new things that we did with jets during Run1 and we should keep doing (possibly even better) for Run2

- **Data scouting:** Dijet resonance searches
- **Jet substructure:** search for heavy resonances decaying to WW/WZ/ZZ searches
- **Jet kinematic variables:** SUSY search as a bump hunt with Razor

Needed "Tools"

- The high rate: at high peak luminosity the LHC is a "junk" maker. This is why we have a trigger system. Easy with rare objects (e.g. e,μ). When you want to look at jets and all you get is jets, you need to make choices that come with a price

Trigger

- The majority of the high-pT LHC events come with jets. Everything is a background to deal with

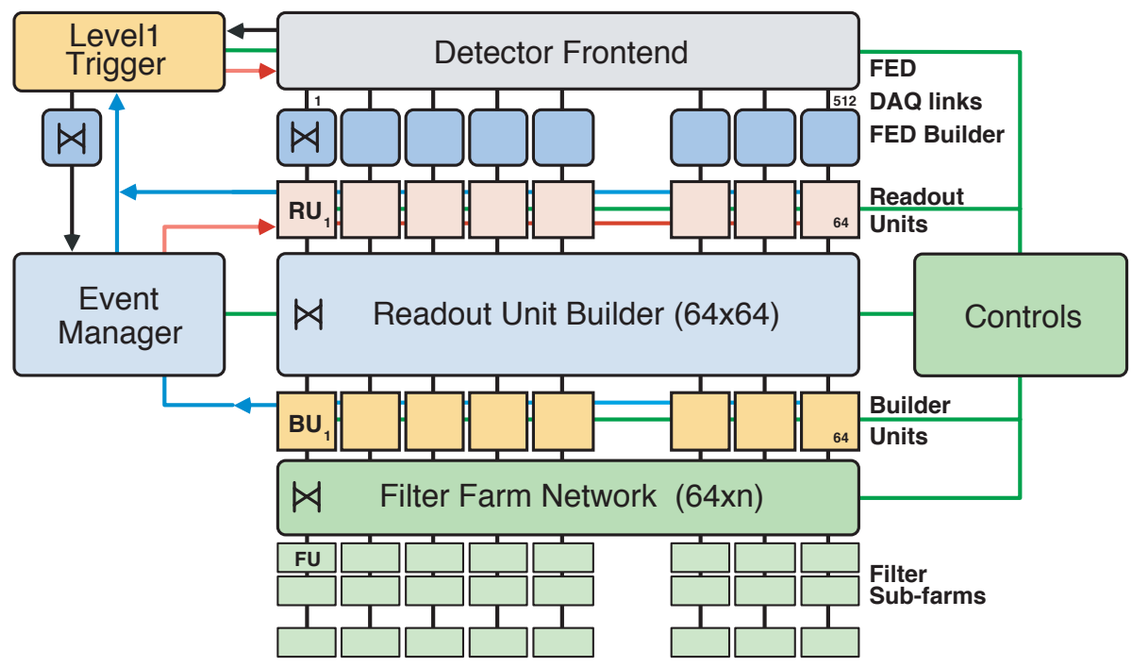
Kinematic

- Jets are not only a way to reconstruct quarks & gluon. A jet could come from boosted W,Z,H,top with overlapping daughters. New opportunities, which come with new needs (good granularity to resolve the substructure)

Particle Flow

The CMS Trigger

- Hardware L1 trigger interfaced to the detector (standard)
- No L2/L3 trigger. Instead, software-based trigger running on PC farm
- Running online a faster version of the offline reconstruction
- In principle (and in practice) one could run the analysis selection online
- This is what we do to keep the analyses as loose as possible

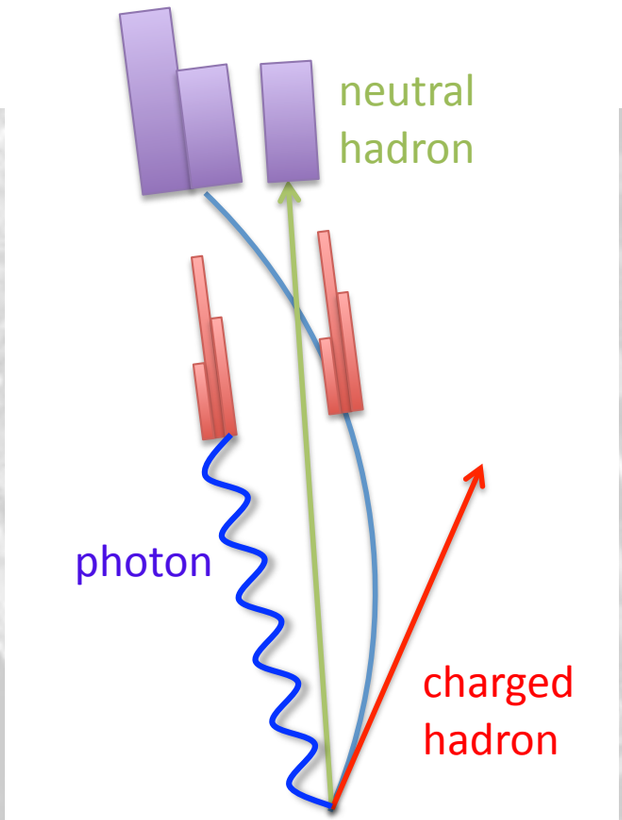


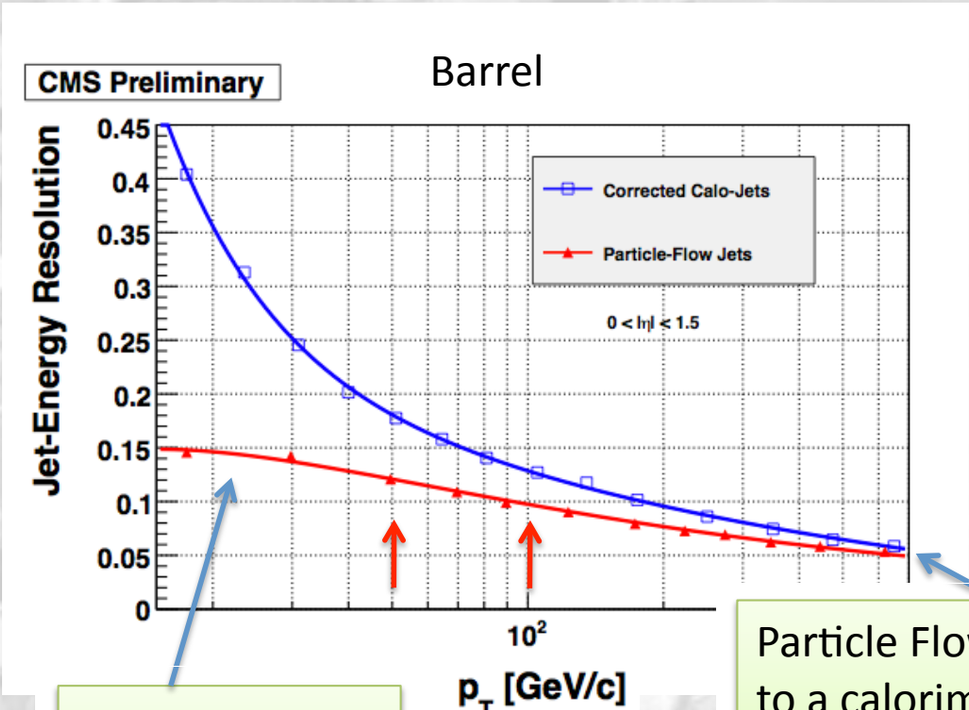
~ 350 Hz of physics taken and reconstructed

~ 600 taken and parked for next year (not enough CPU @T0)

Particle Flow

- Combine the information from all detectors to reconstruct single particles
- Provides lists of particles (e,m,g, charged and neutral hadrons)
- Improves HCAL resolution with tracker
- Replace the HCAL granularity with tracker granularity (important for jet substructure)





Factor 2 improvement at low pT

Particle Flow converges to a calorimetric measurement at high pT

Dijet Resonance Searches

The Fastest LHC Search

Easy to say

- Select two jets at large pT, regardless of the rest of the event
- Look for a bump on a falling invariant-mass spectrum

But life is harder than that

- Mass resolution is not that good even at gen level
- At large pT, quarks and gluons radiate and make it worse
- The dijet rate is very high at trigger level. Cannot look for light resonances with small cross sections

The Fastest LHC Search

Easy to say

- Select two jets at large pT, regardless of the rest of the event
- Look for a bump on a falling invariant-mass spectrum

But life is harder than that

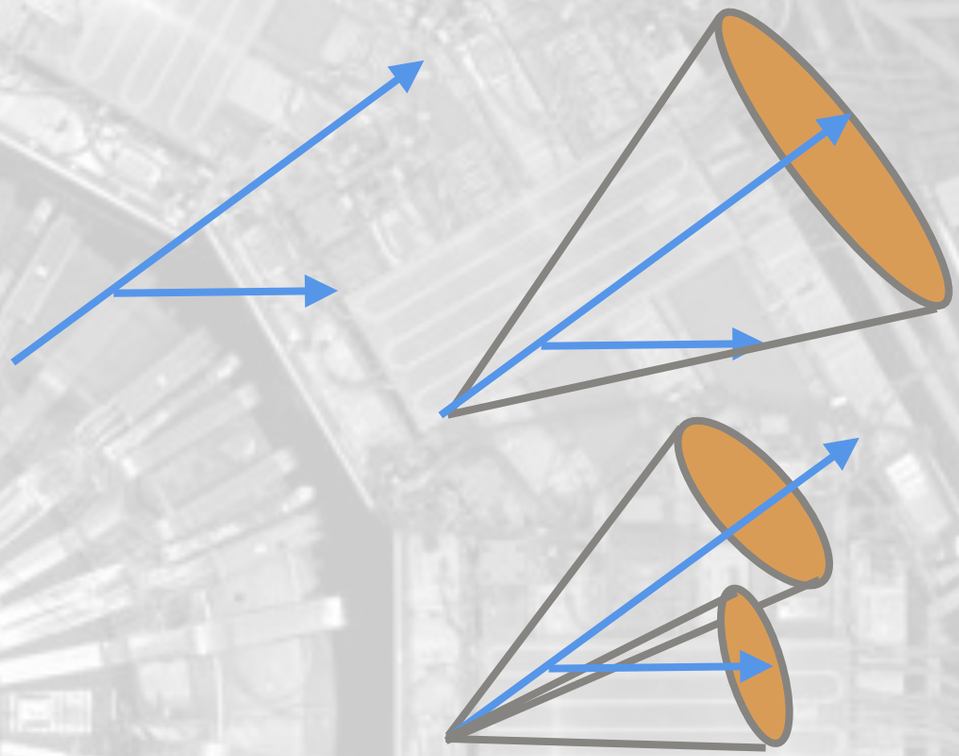
- Mass resolution is not that good even at gen level
- At large pT, quarks and gluons radiate and make it worse
- The dijet rate is very high at trigger level. Cannot look for light resonances with small cross sections

The "wide" jets

- Hard radiation happens at small angles → The radiated q/g should produce a jet nearby

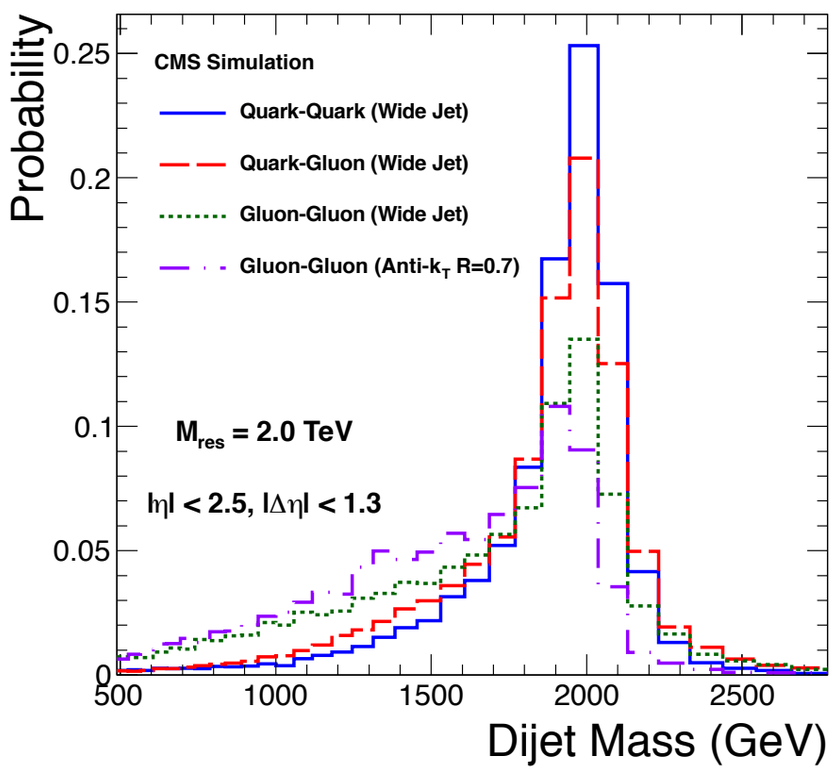
- One could use a larger cone

- Or try to recover the radiation looking for a jet nearby. We added to the initial jet all the jets within $\Delta R = 1.1$ and $pT > 40$ GeV (PU is a problem for softer jets)



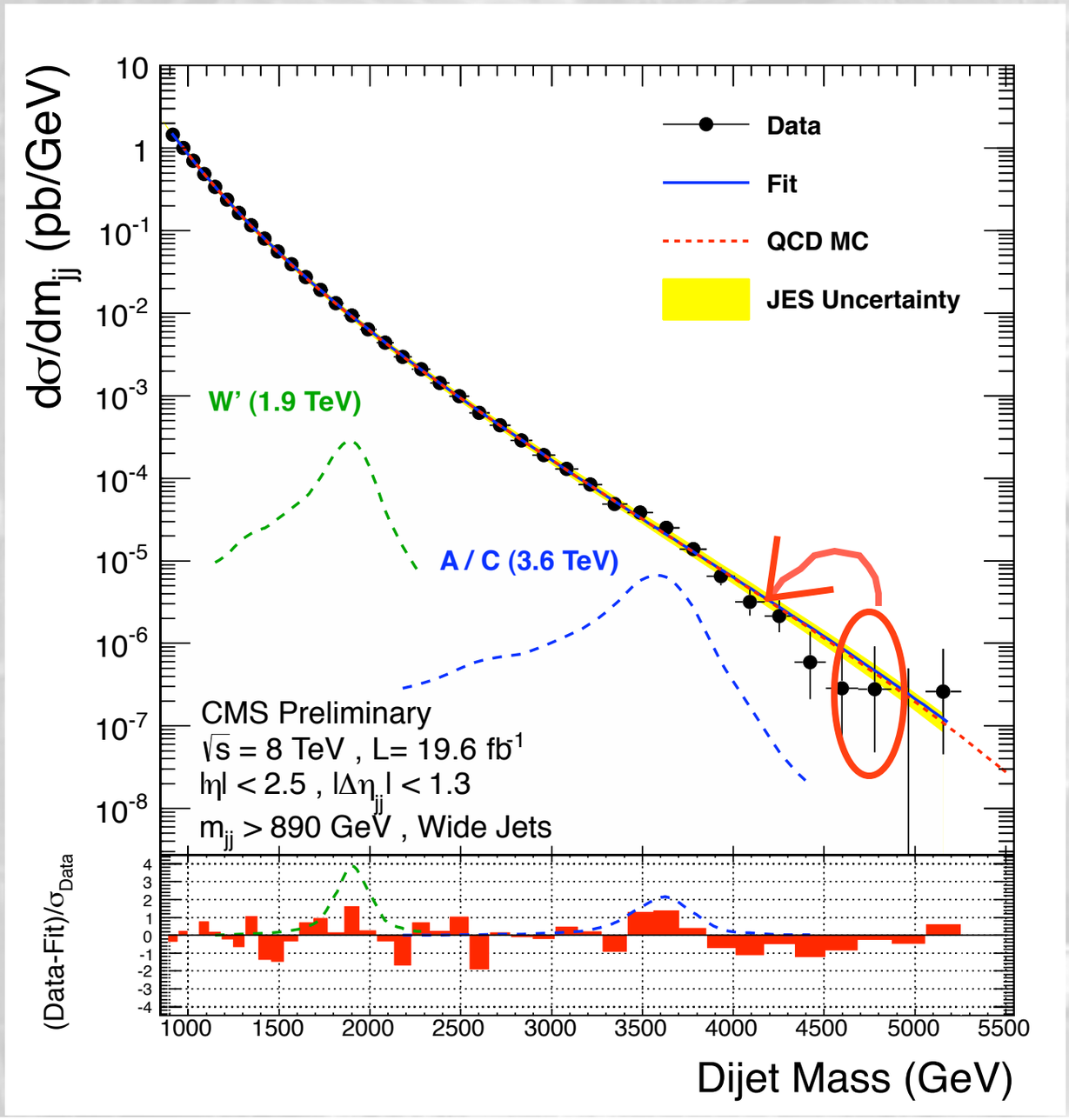
- A better solution would be an anti-kT clustering of the clustered jets, infrared safe (working on that)

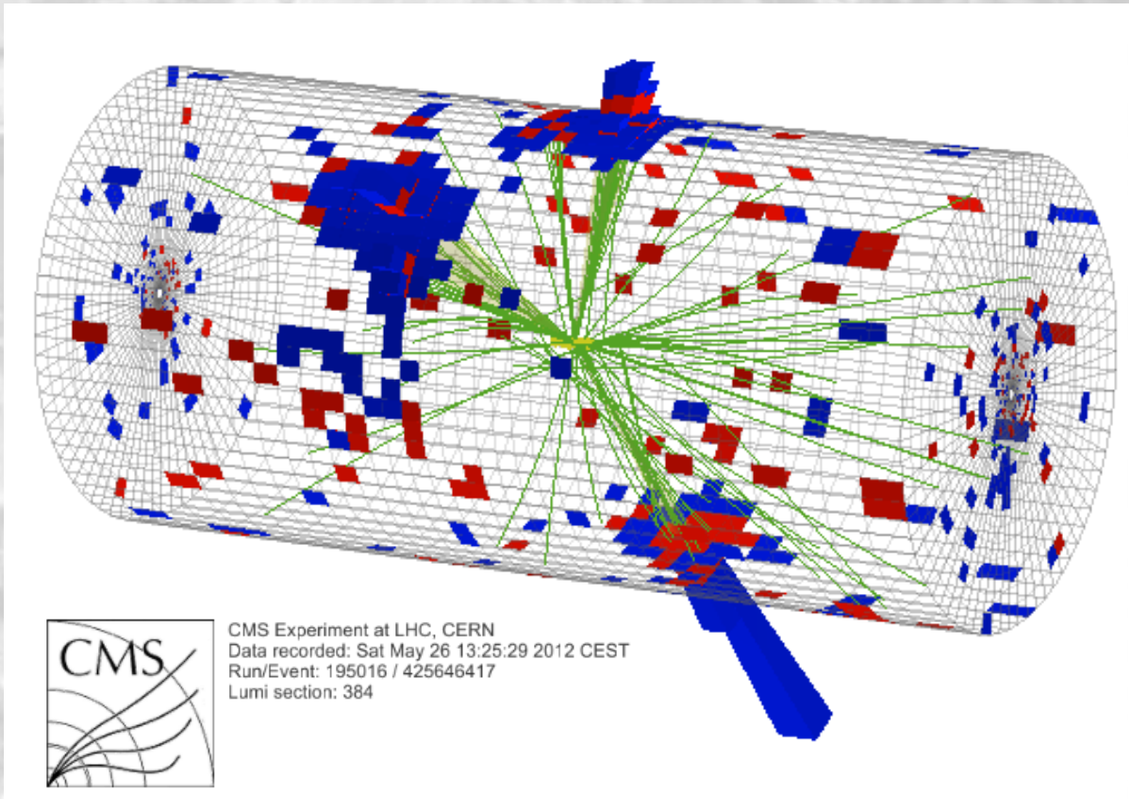
- Peaks moves at the right place. Resolution improves (particularly for gg resonances, as gluons radiate more)



A SELECTED EVENT

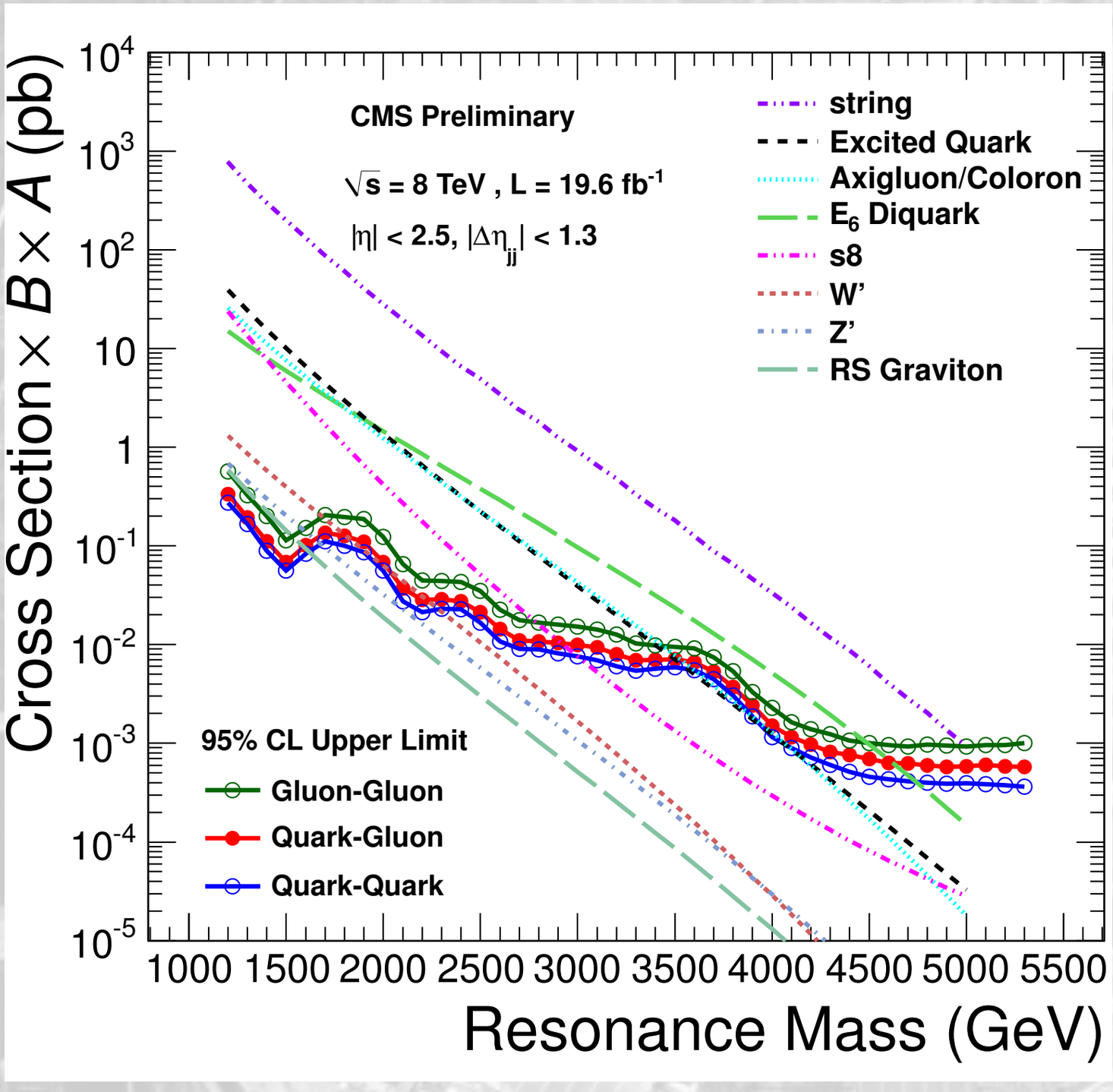
- MANY SIGNAL EVENTS (FROM SIGNAL MC) LOOK LIKE THIS (REAL DATA) EVENT

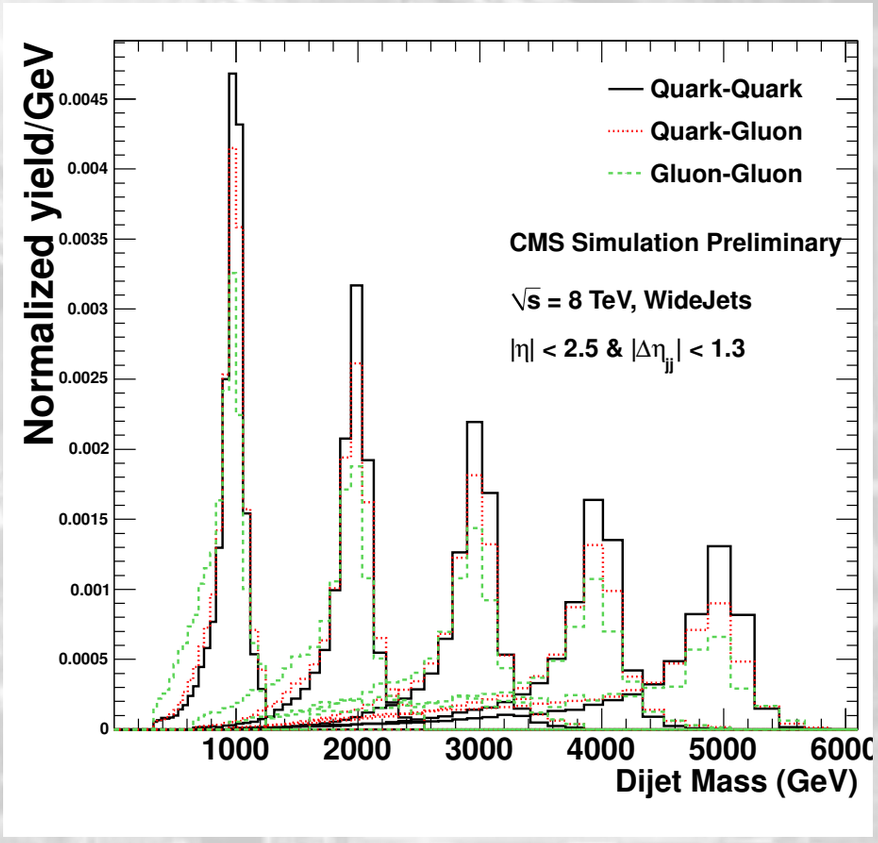




WHEN THE RADIATED JET IS MISSED, THE MASS IS UNDERESTIMATED AND THE SIGNAL EVENT MOVES TO A HIGHER-BACKGROUND BIN

8 TeV Results

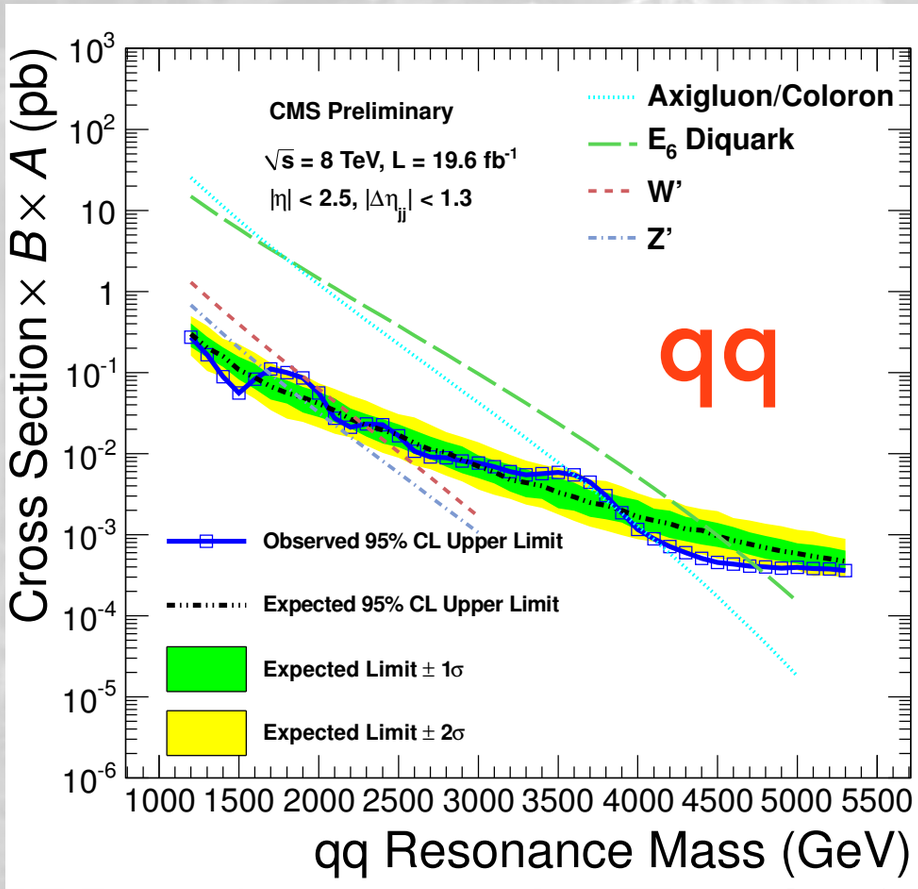


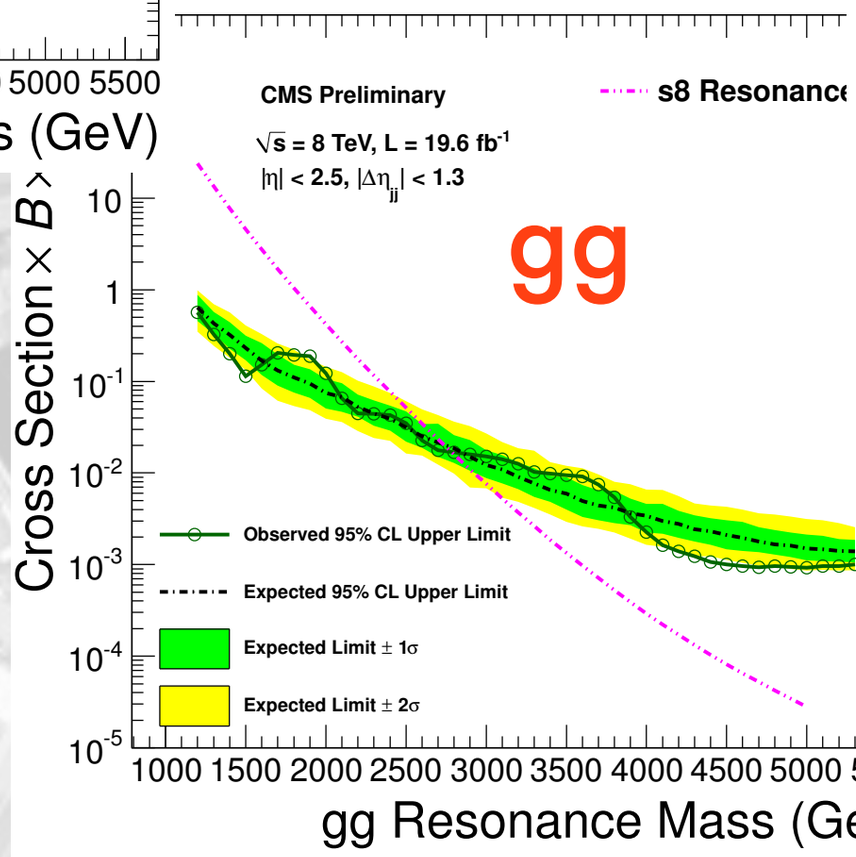


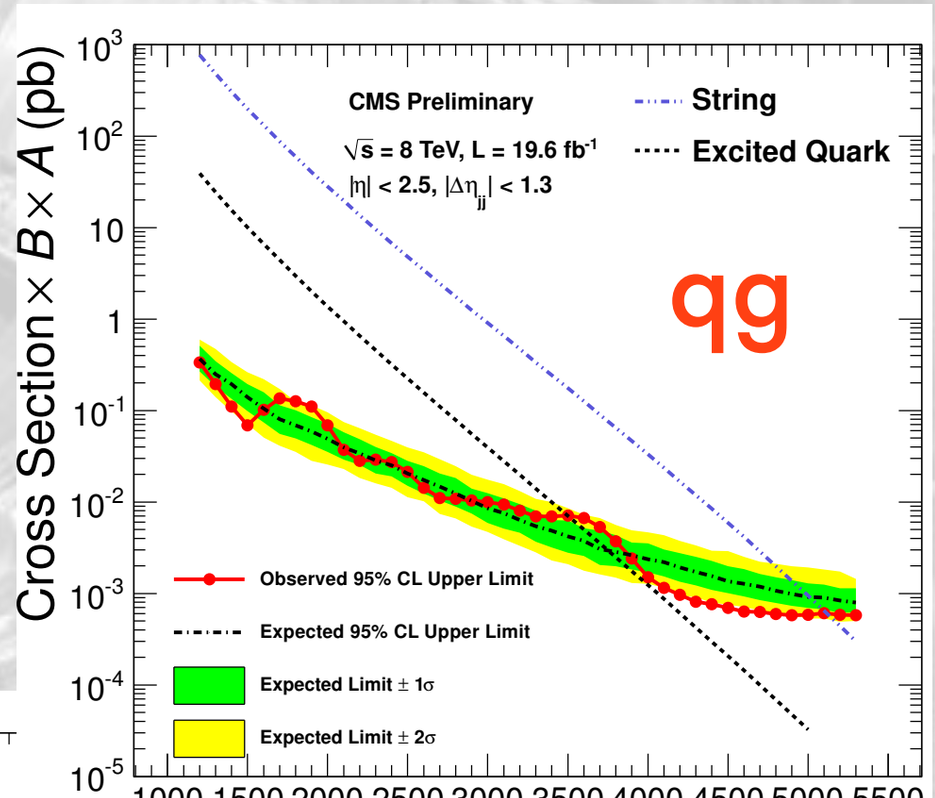
The limit depends on the final state, as the peak is wider when more gluons (more radiation) in final state

The largest "excess" is ~2σ around 3.5 TeV

8 TeV Results



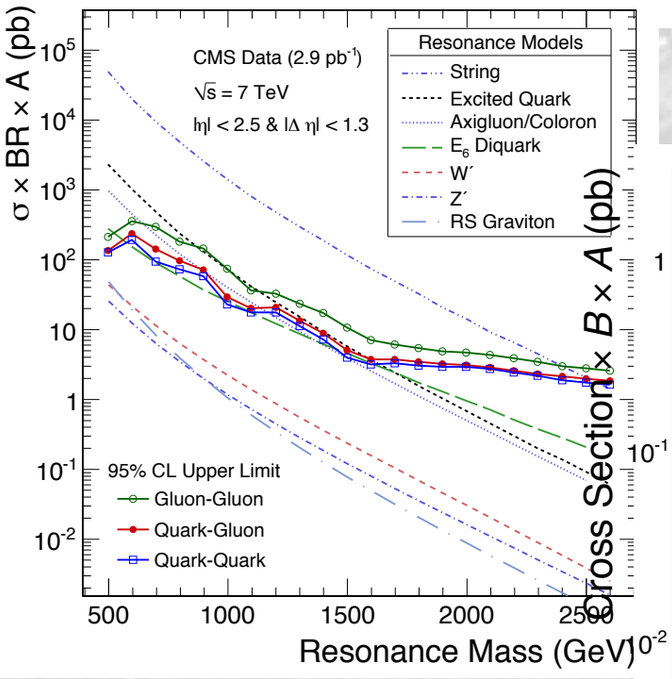




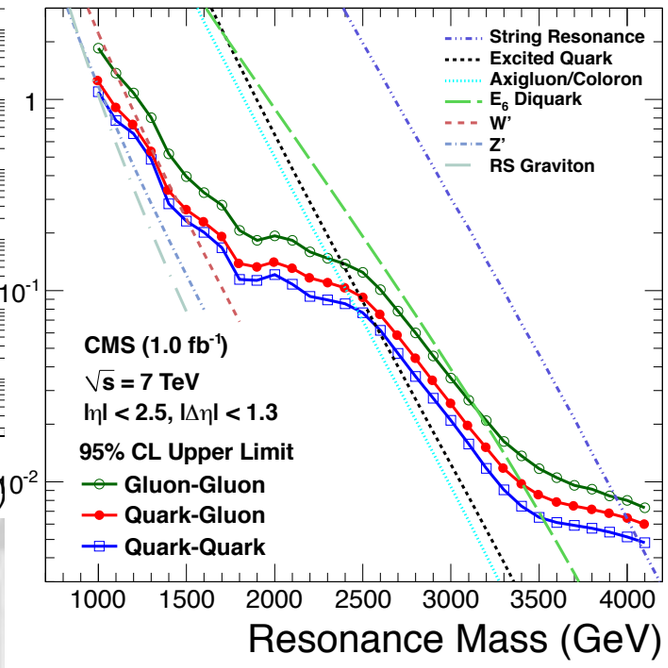
THE LIMIT DEPENDS ON THE FINAL STATE, AS THE PEAK IS WIDER WHEN MORE GLUONS (MORE RADIATION) IN FINAL STATE

Results vs Time

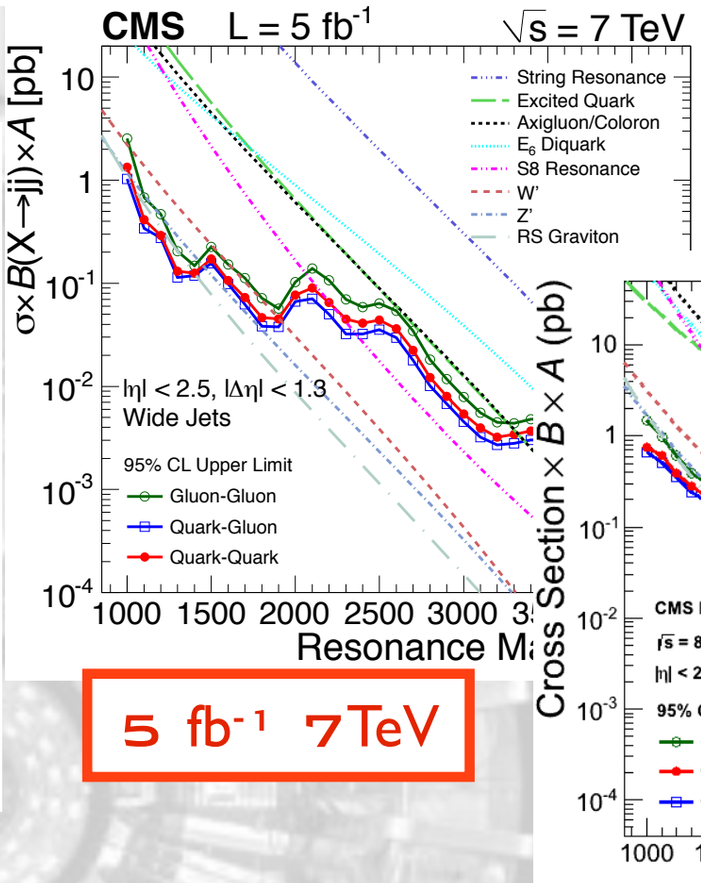
3 pb⁻¹ 7 TeV



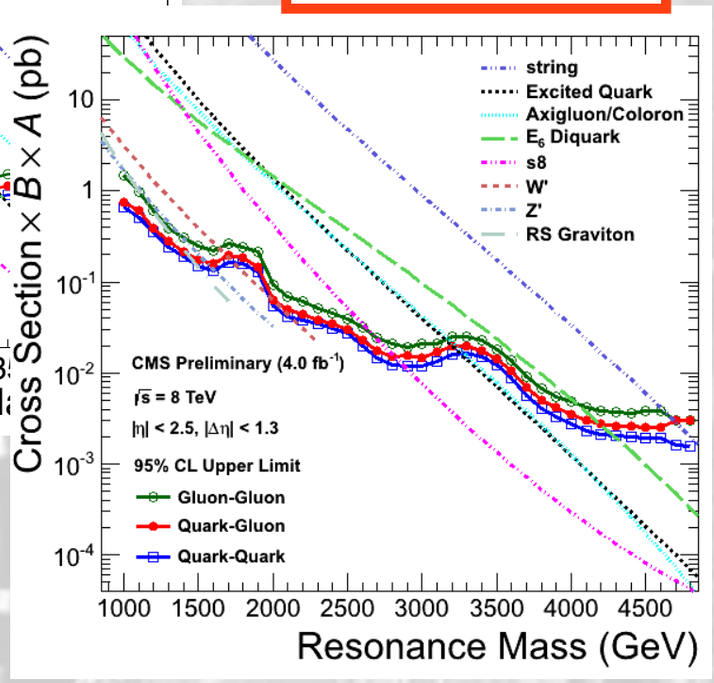
1 fb⁻¹ 7 TeV



5 fb⁻¹ 7 TeV



3 fb⁻¹ 8 TeV



E [ ]	L [ ]	minMass [ ]	σ minMass
7	0.003	400	~ 200
7	1	900	~ 1
7	5	1000	~ 1
8	3	1000	~ 0.6

- Unexplored territory left behind when the luminosity increased

- Trigger improvements compensated the (slower) rate increase after

The Fastest LHC Search

Easy to say

- Select two jets at large pT, regardless of the rest of the event
- Look for a bump on a falling invariant-mass spectrum

But life is harder than that

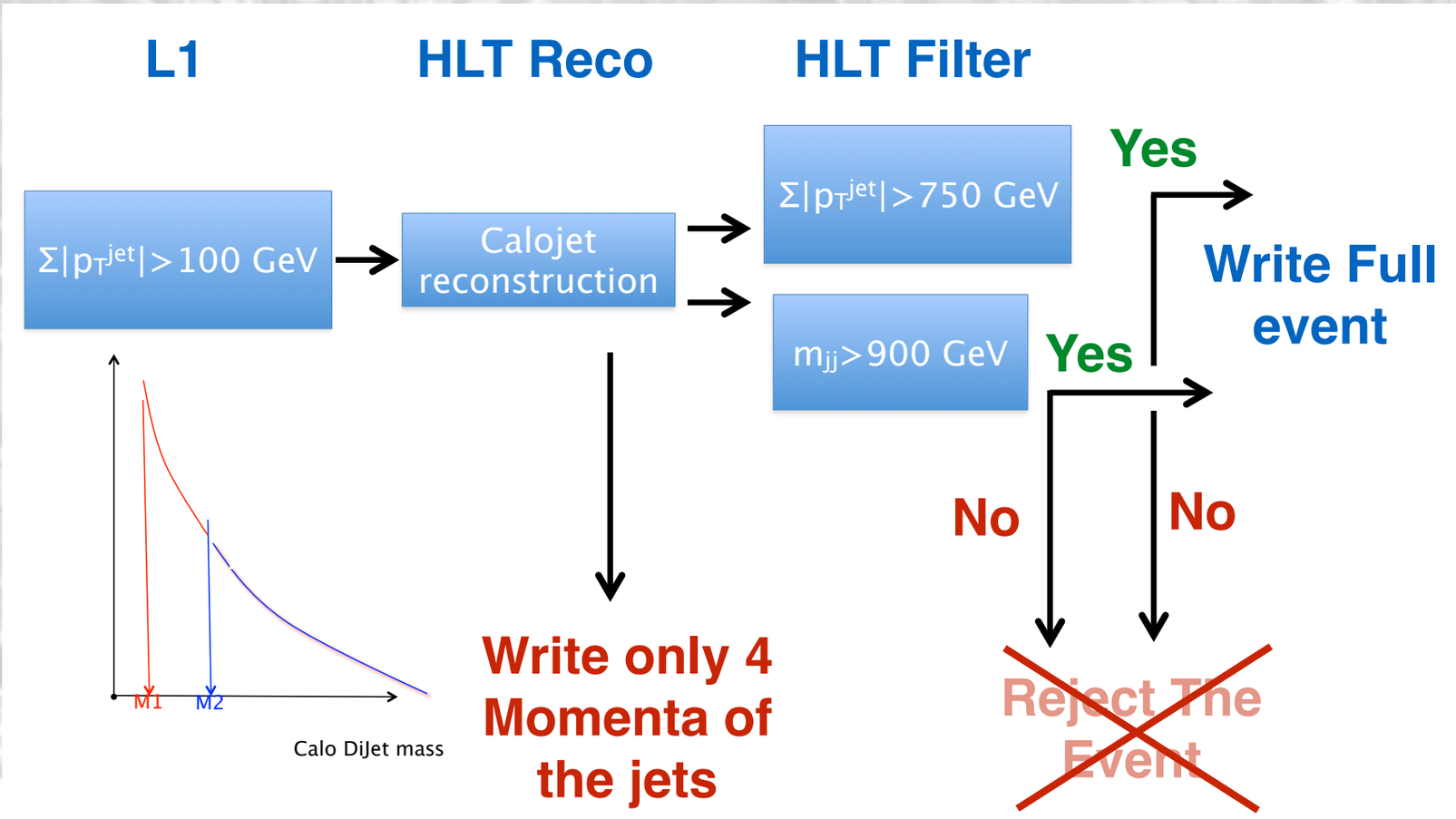
- Mass resolution is not that good even at gen level
- At large pT, quarks and gluons radiate and make it worse
- **The dijet rate is very high at trigger level. Cannot look for light resonances with small cross sections**

The Dijet Data Scouting

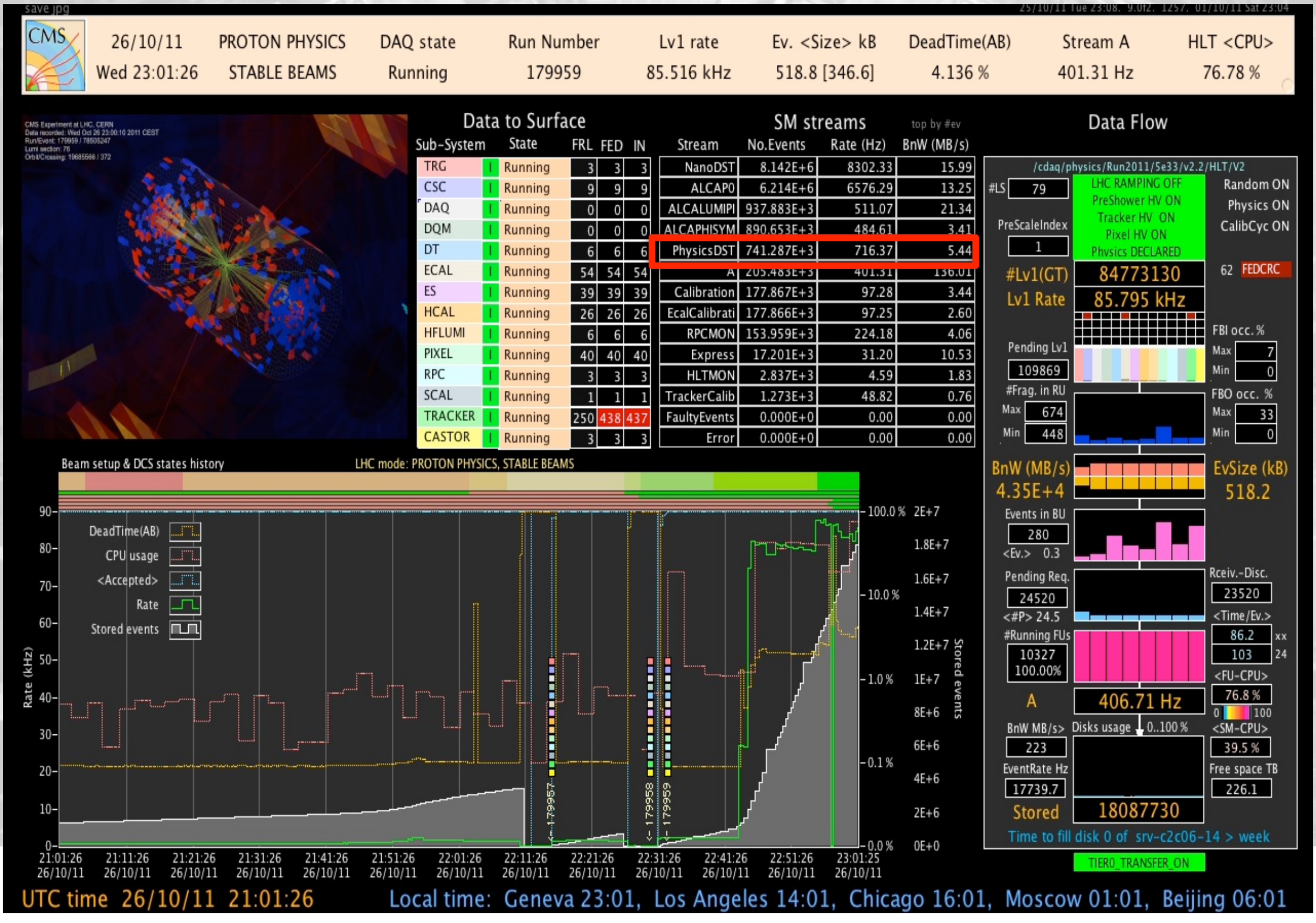
- The bottle-neck is the I/O (MB/s) and not the rate (event/sec)
- We only need the jet pT, and some minimal information to reject noise
- We write much more than that
- **THE COMPROMISE**

Accept worse resolution @HLT

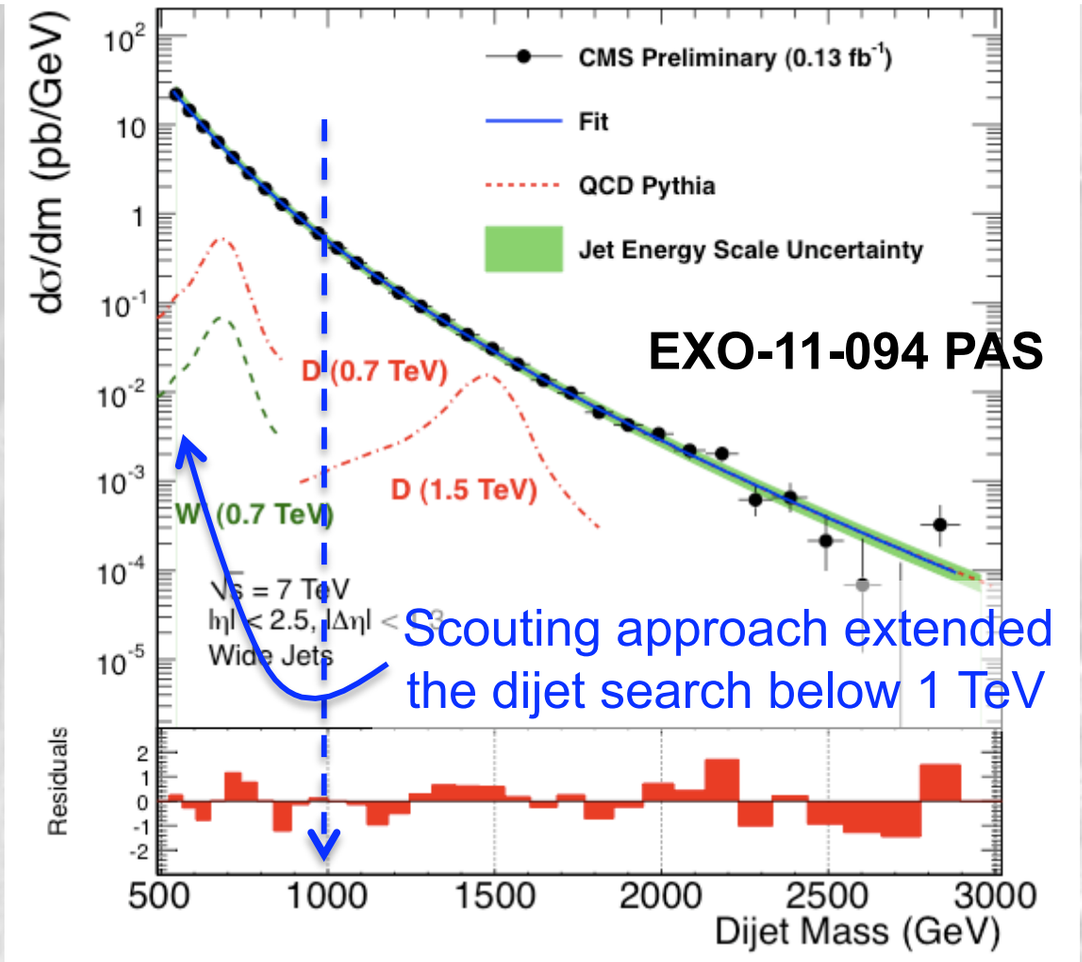
Write only minimal information (4momenta jets&leptons) <10 Kb/s

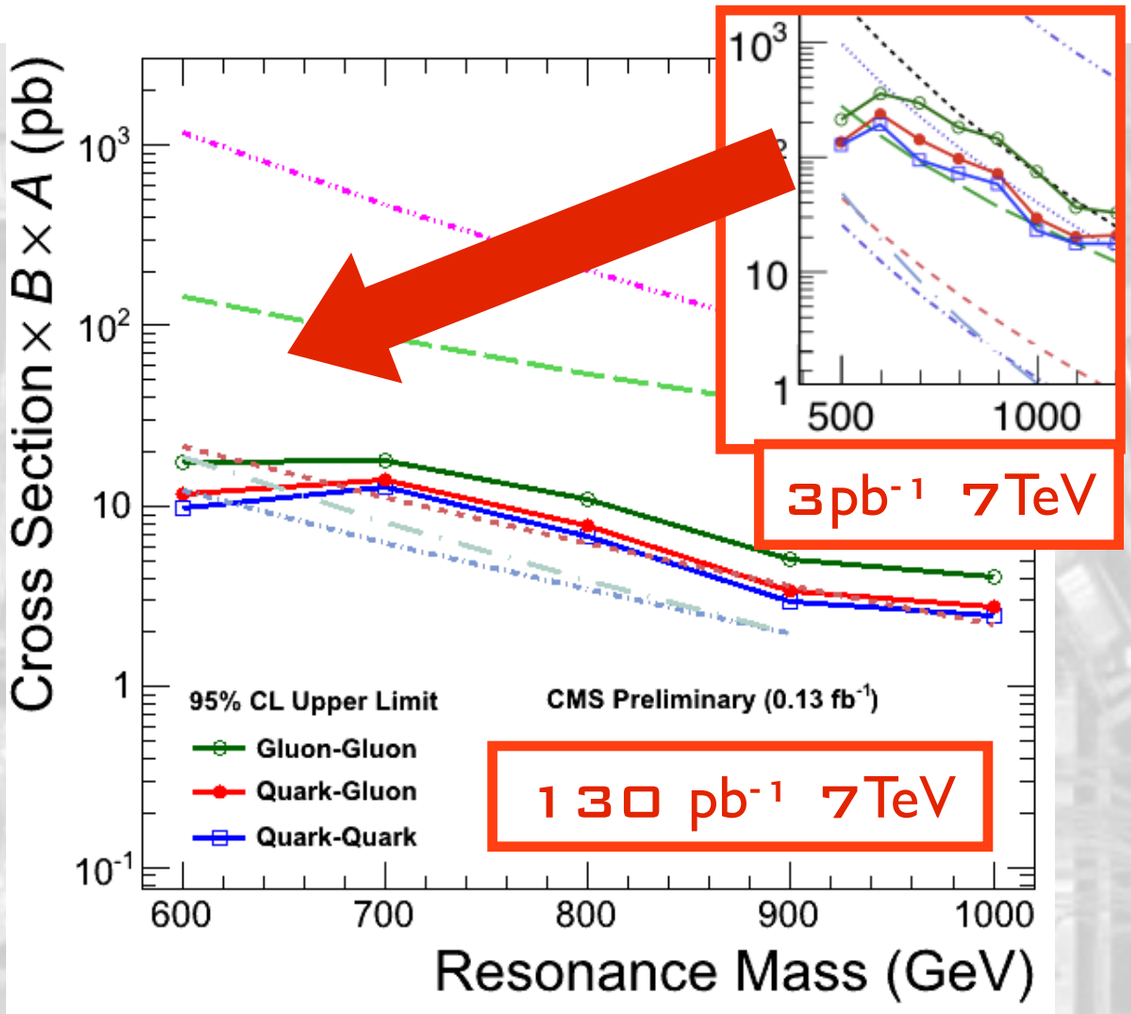


O(1 kHz) DATA OUT



The Dijet Data Scouting





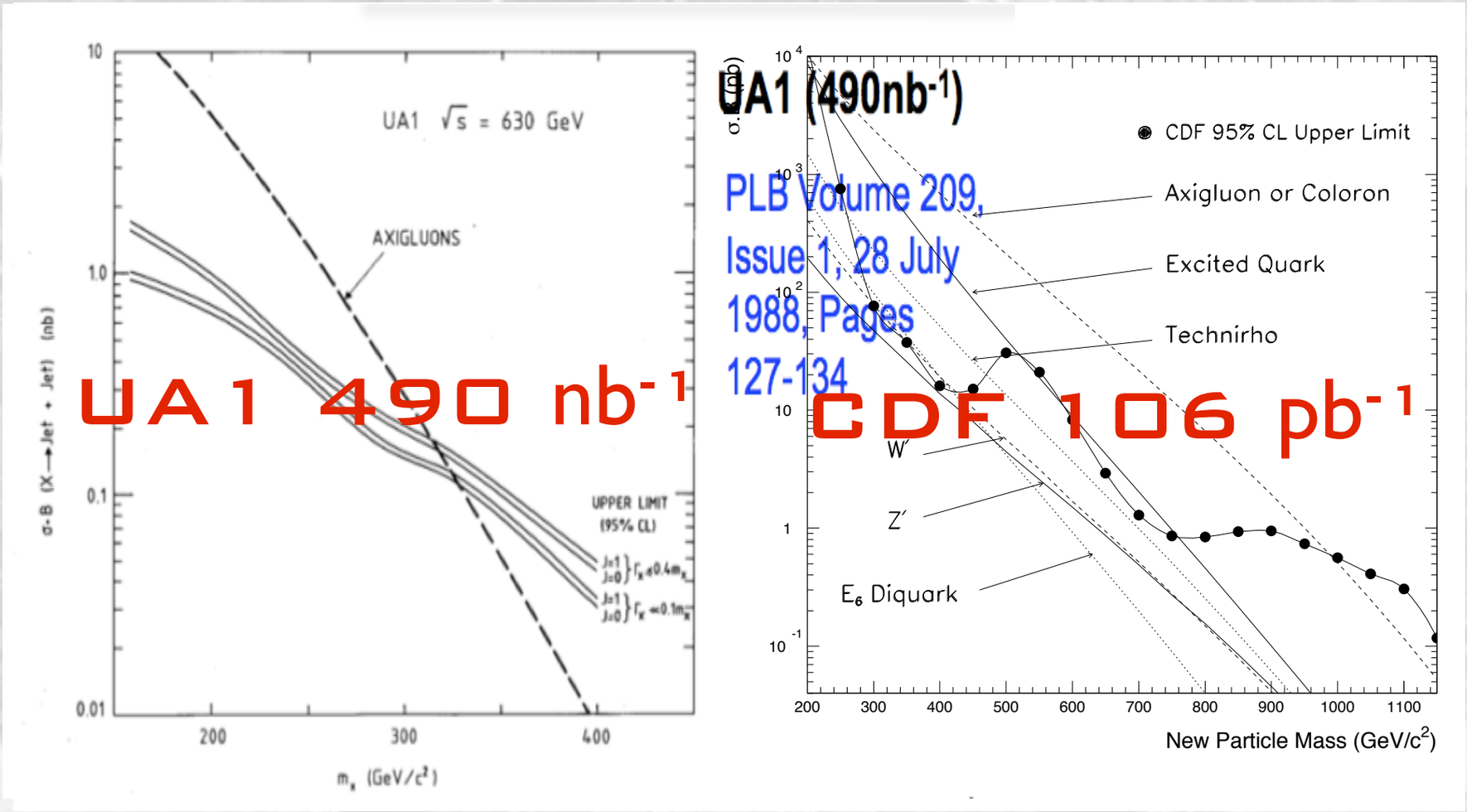
16 hour Run at the end of 2011 run (7TeV)

Collected ~4 times the statistics we had in 2010 (35 pb^{-1}) with equivalent trigger

Improved the limit published in 2010 by one order of magnitude

Back To the '80s/'90s

There is a region of the σ vs mass plane which was not probed since UA1,UA2 and Tevatron Run1



Scouting can be used to explore that region

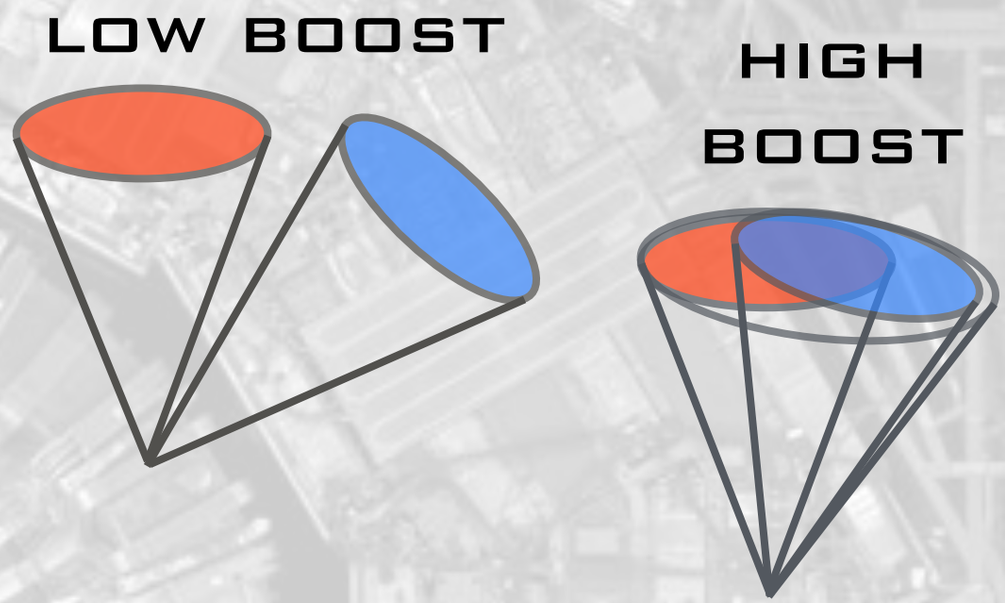
Collected ~18 fb^{-1} with scouting switched on

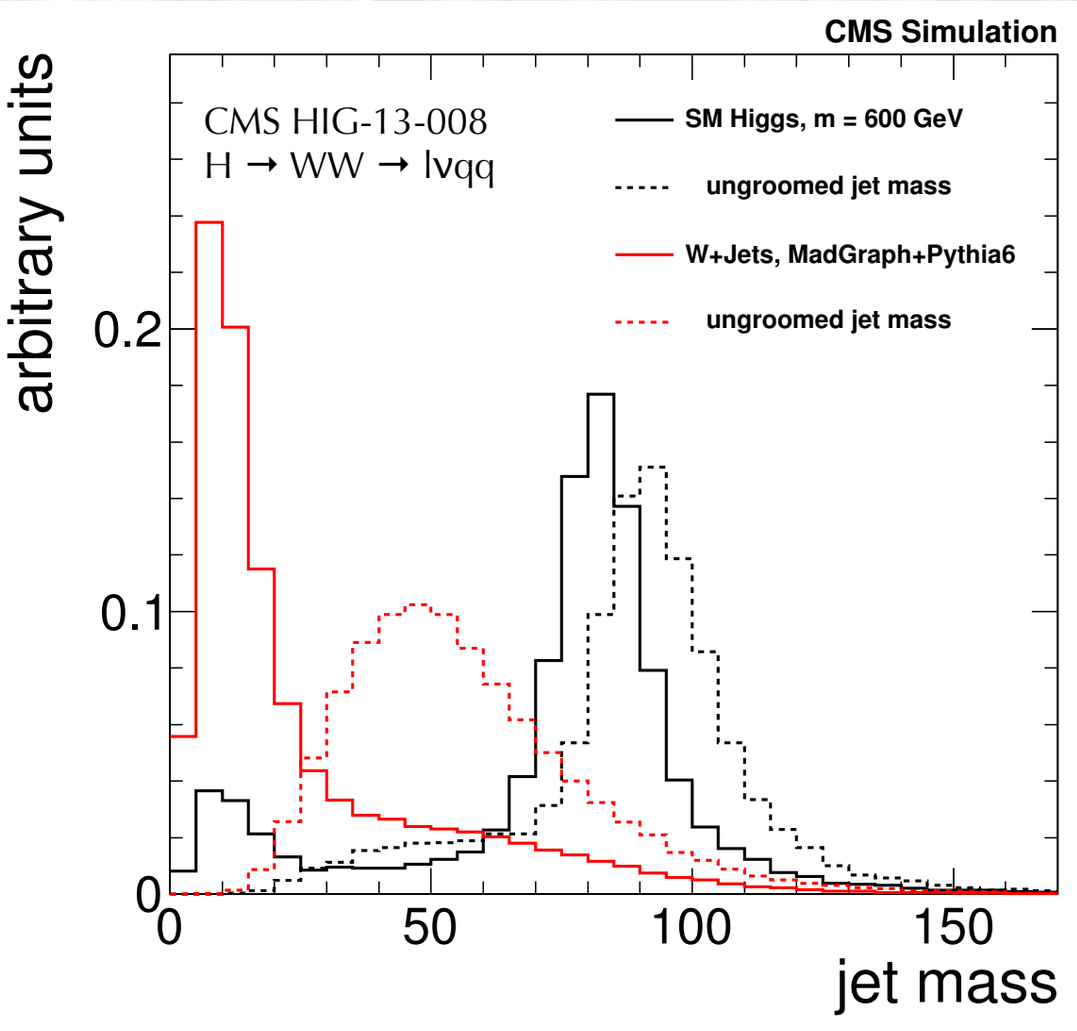
Results expected by the Summer

WW/WZ/ZZ Resonance Searches

From Jets to boosted Vs

- The dijet analysis is sensitive to whatever looks like jets
- In particular, to boosted W/Z (e.g. RS bulk gravitons)
- A simple cut on the jet mass would remove the obvious QCD background
- The jet mass for a QCD jet is a pT measurement, so it is not quite useful (~ V mass peak for high-enough pT)
- This is why grooming techniques are introduced



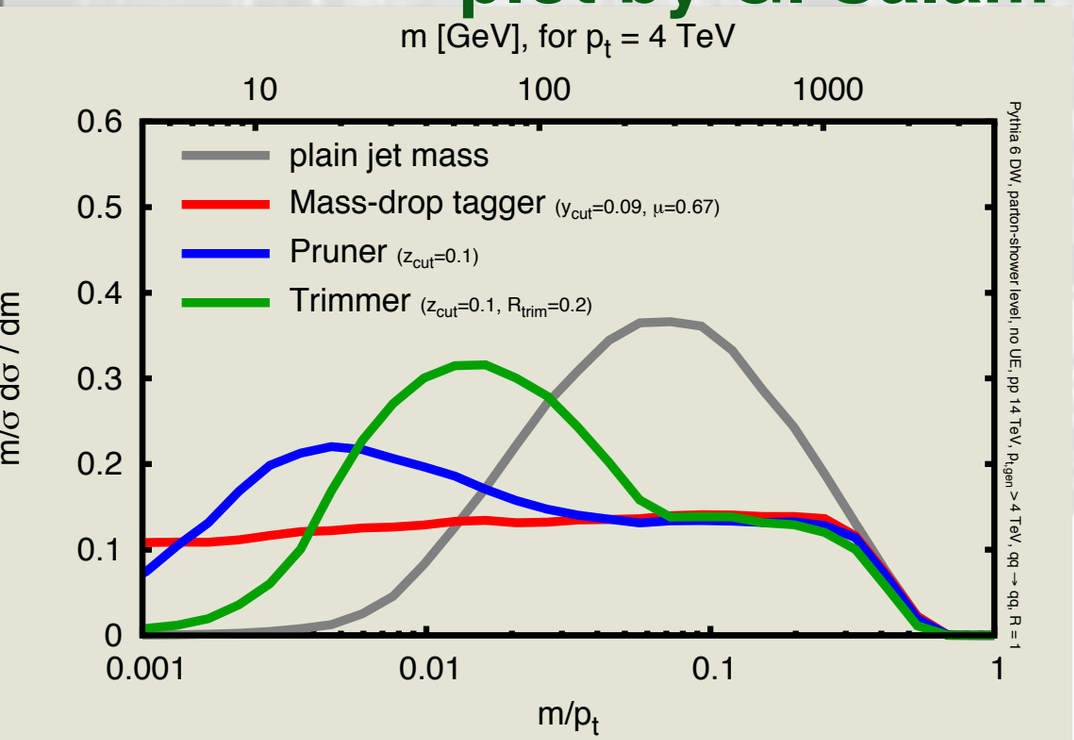


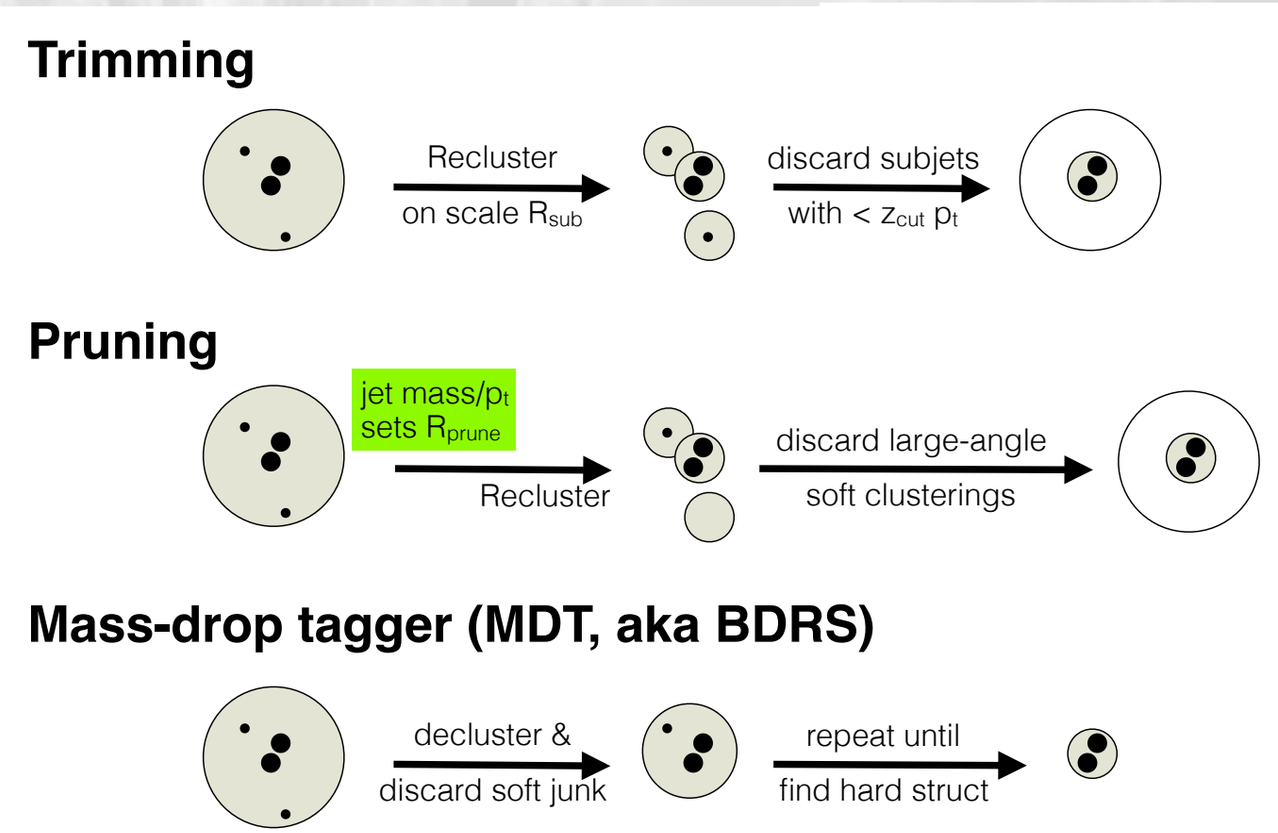
Jet Grooming

- With jet grooming we mean a set of techniques to "clean" the jet from soft radiation
- Once this is done, the jet mass relates to the mass of the parton seeding the showering
- Moreover, these techniques attenuate the effect of PU (which looks like soft radiation)

- Several techniques:

plot by G. Salam





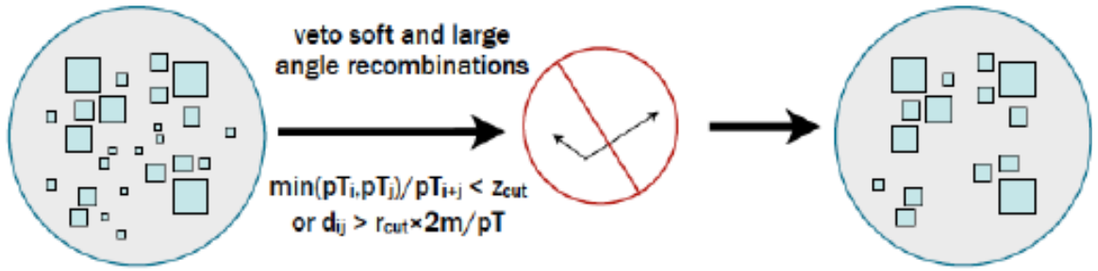
Pruning

- Start from a large-radius jet(CA with R=0.8 for CMS)

- Recluster the jet constituents and evaluate the hardness and angular separation of the last recombination

hardness angular separation

$z = \frac{min\left(p_{T,i}, p_{T,j}\right)}{p_{T,JET}} > 0.1$	$\Delta R < 0.5 \frac{M_{JET}}{p_{T,JET}}$



- Remove the softest subject if conditions not satisfied

- Procedure removes soft and large-angle radiation from QCD. Improves the mass resolution

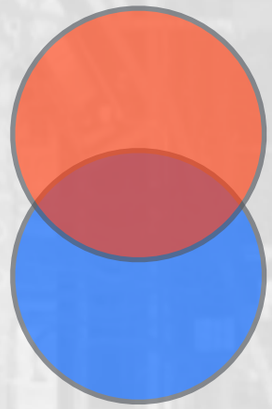
- Also mitigates the pile-up contamination

arXiv:0903.5081 arXiv:0912.0033

Substructure

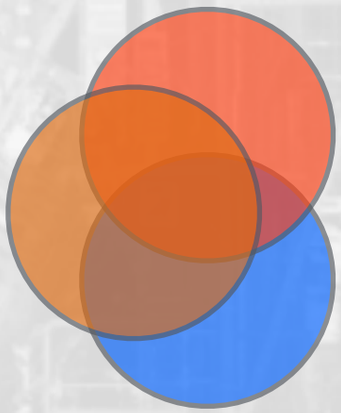
- Substructure consists in looking for clustering of jet constituents (subjets) inside the jet (i.e. “multipole” vs “monopole” source of QCD radiation)

- Growing literature on this subject. About one new variable proposed every month. Dedicated series of conferences (e.g. Boost)



- The name of the game is in good momentum and angular resolution

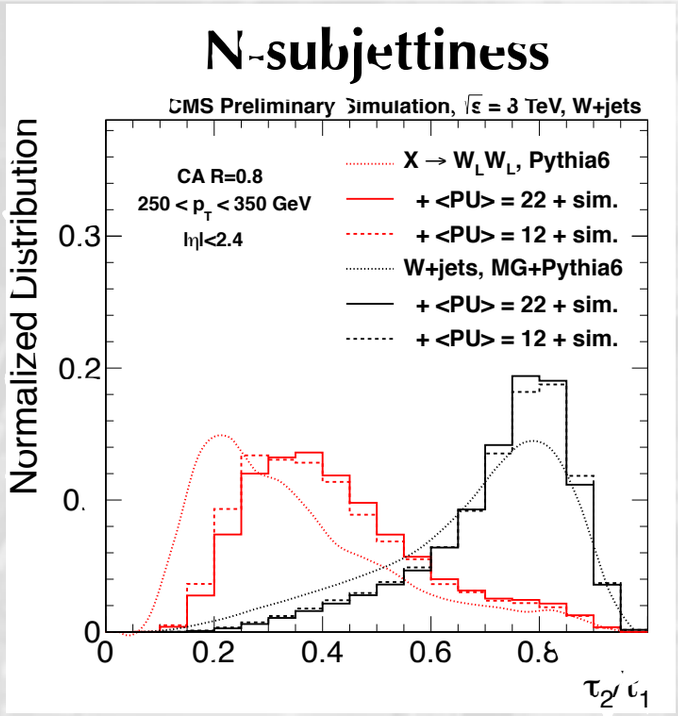
- More emphasis in the calculability of the variables, in order to improve our understanding of this growing subject

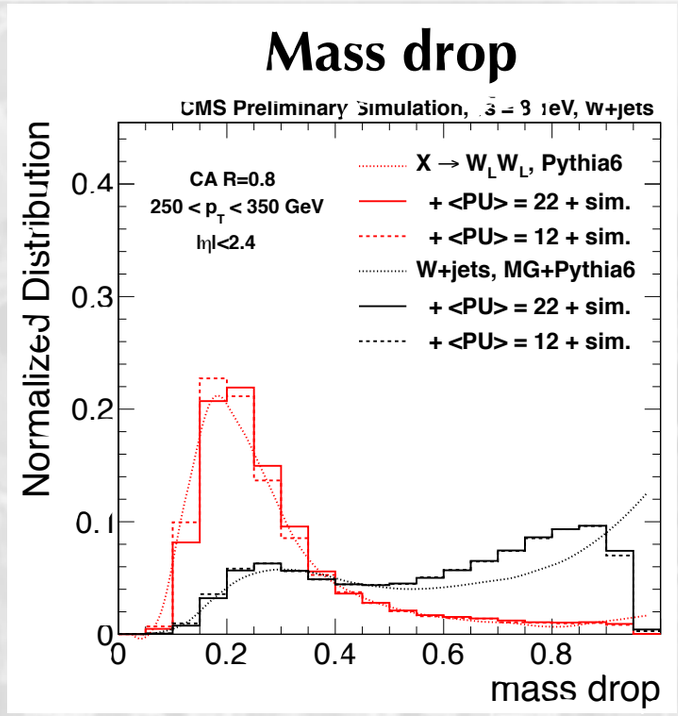


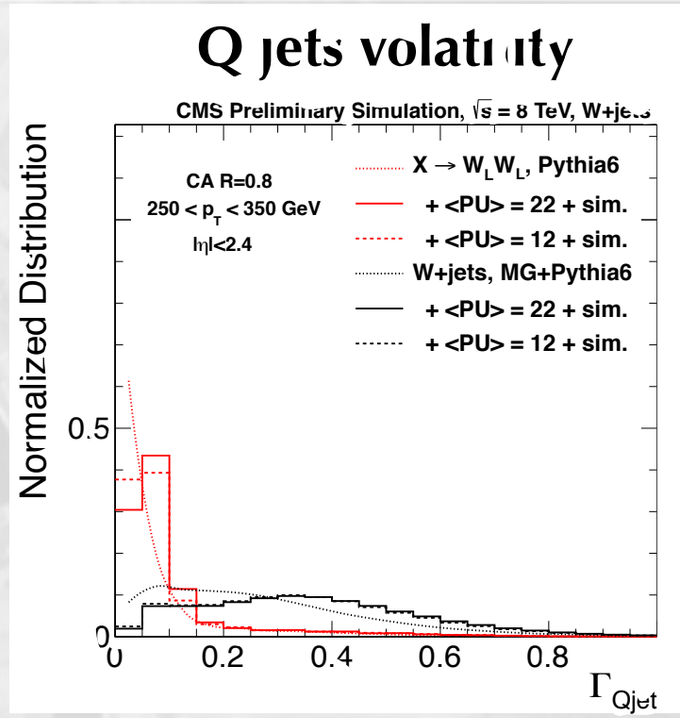
- Also important is the effect of PU (which reduces the discrimination power)

S vs B after detector effects

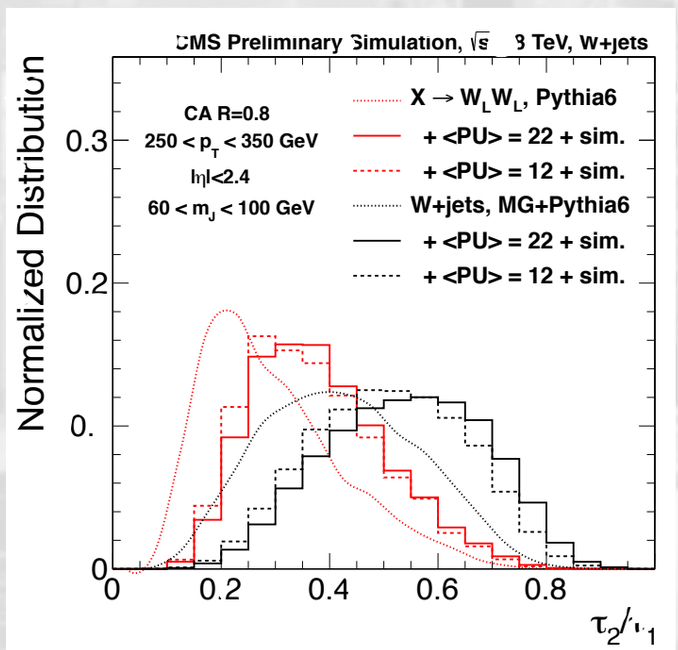
No W mass se ection

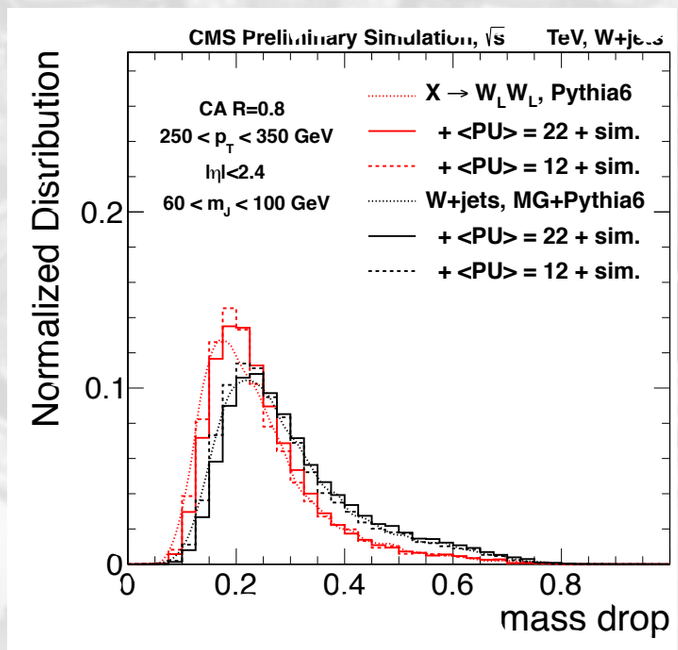


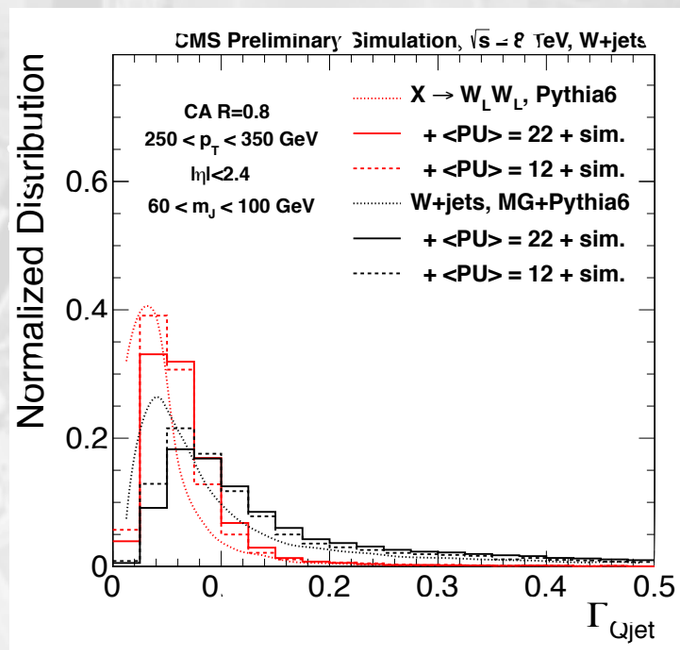




W mass selection



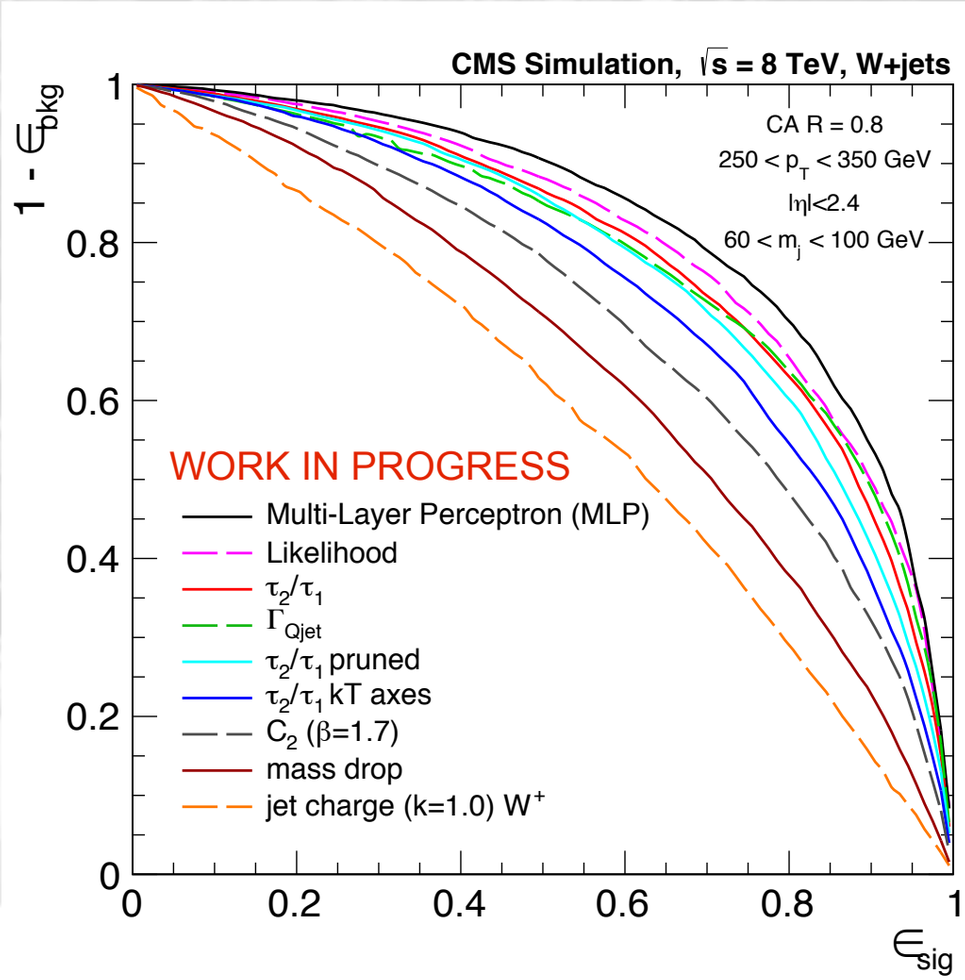




CMS-PAS-JME-13-006

Tagger Definition

- One wants to separate a two-prong from a single-prong angular distribution of the decay products around the jet axis
- Several variables "on the market" to exploit this difference
- CMS studies them and compared performances (also with MVA techniques). An obvious winner in performance vs simplicity compromise
- A fundamental ingredient here is PF: not only the energy resolution but also the angular resolution and the detector granularity at play. The integration of the tracker with the calorimeters provides this at optimal



CMS-PAS-JME-13-006

Jet N-Subjettiness

- N-subjettiness is smaller if the constituents of a jet can be arranged in N subjets

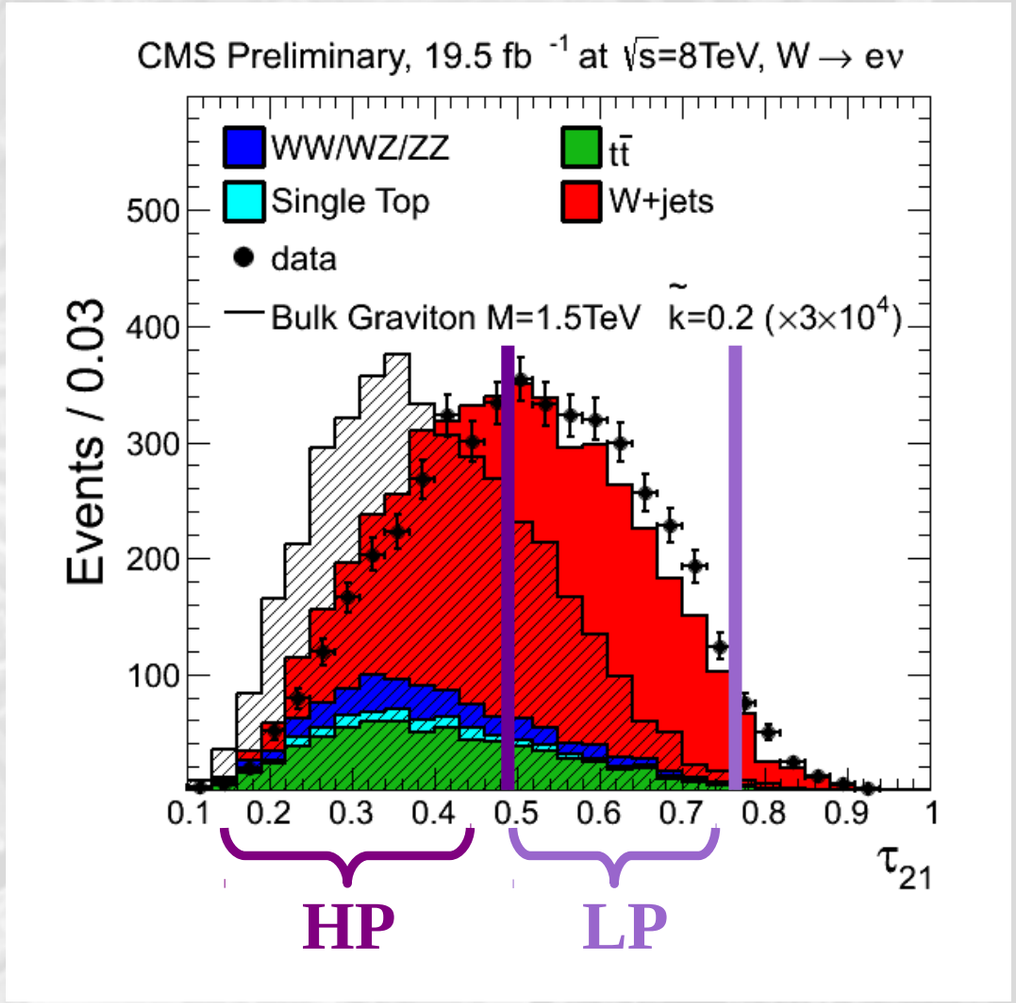
- One can use the variable to test different hypotheses (e.g. V vs top vs QCD jet)

- In real life, ratios are particularly useful to categorize events (High purity vs Low Purity) after a mass cut on the jet is applied

- Correlation with jet mass and PU effects tend to reduce the discrimination power

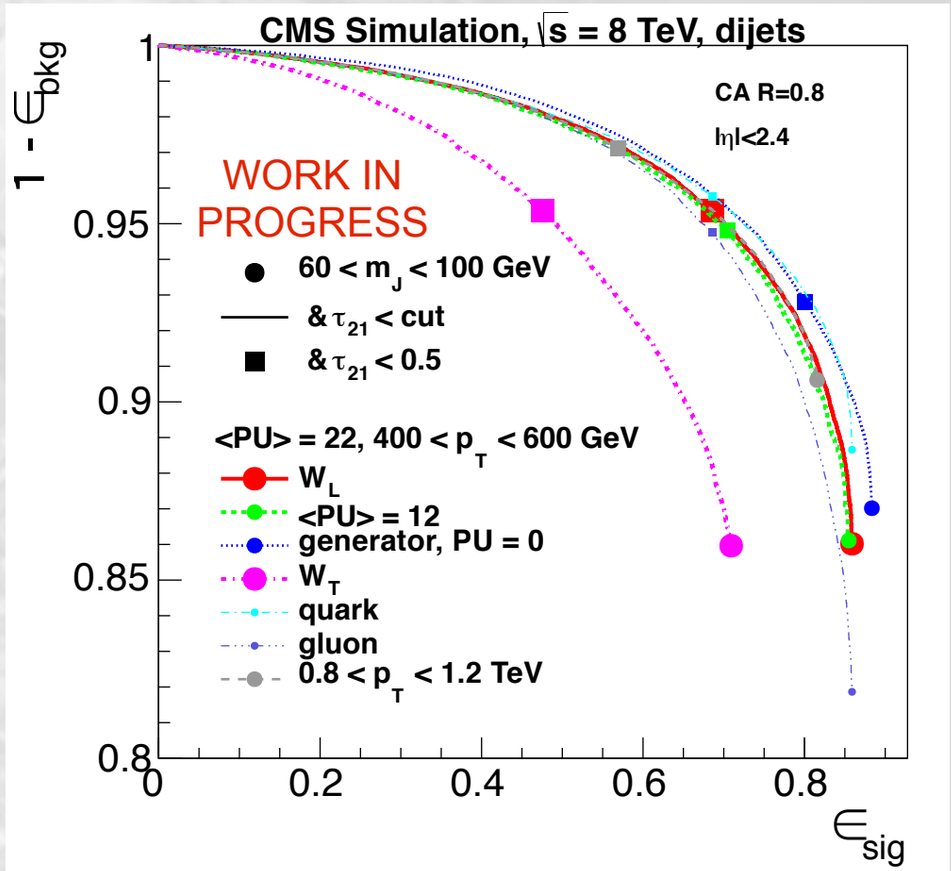
- Polarization also matters (e.g. separation more effective for V_L than V_T)

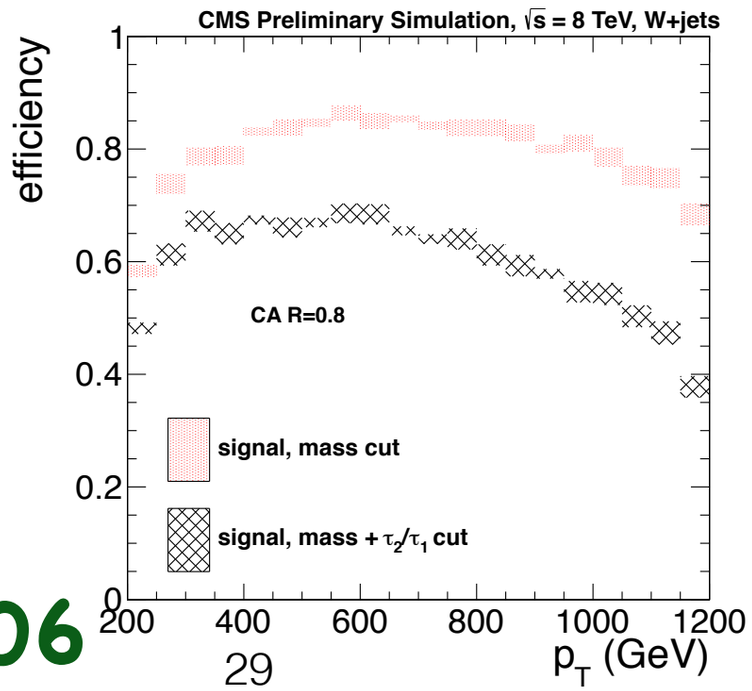
$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \, min\left\{ \Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k} \right\}$$

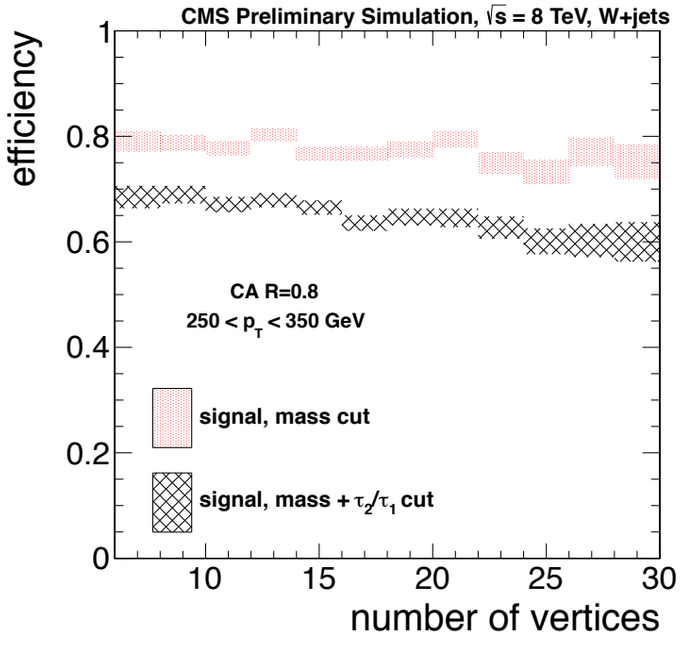


Tagging Efficiency

- Tagger consists of mass cut
  [60,100] GeV + τ_2/τ_1<0.5
- Efficiency is better for V_L than for V_T
- Similar discrimination power vs quarks and gluons
- Even a minimal Pileup (12 interactions/ BunchX) has an impact on bkg rejection vs signal efficiency

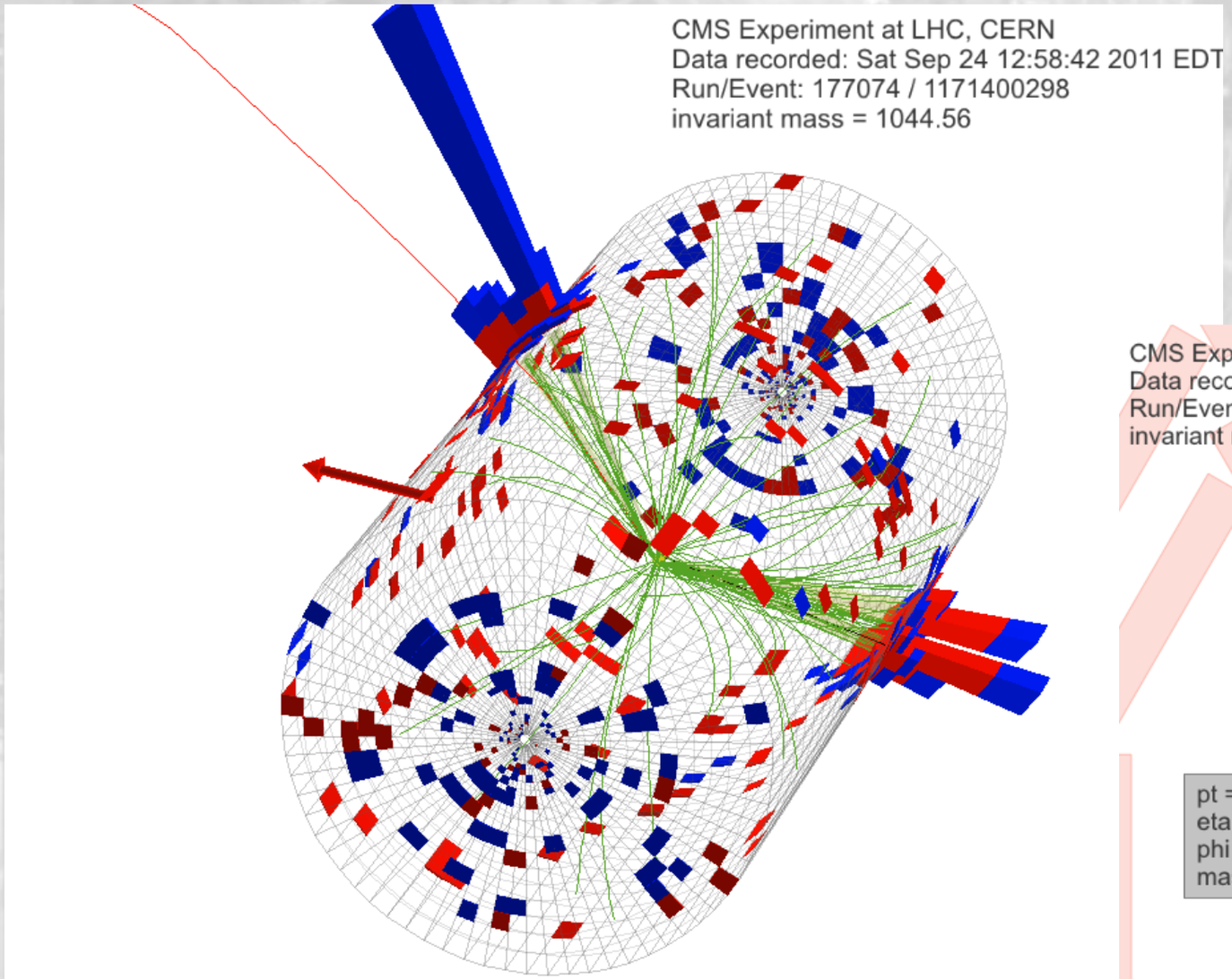


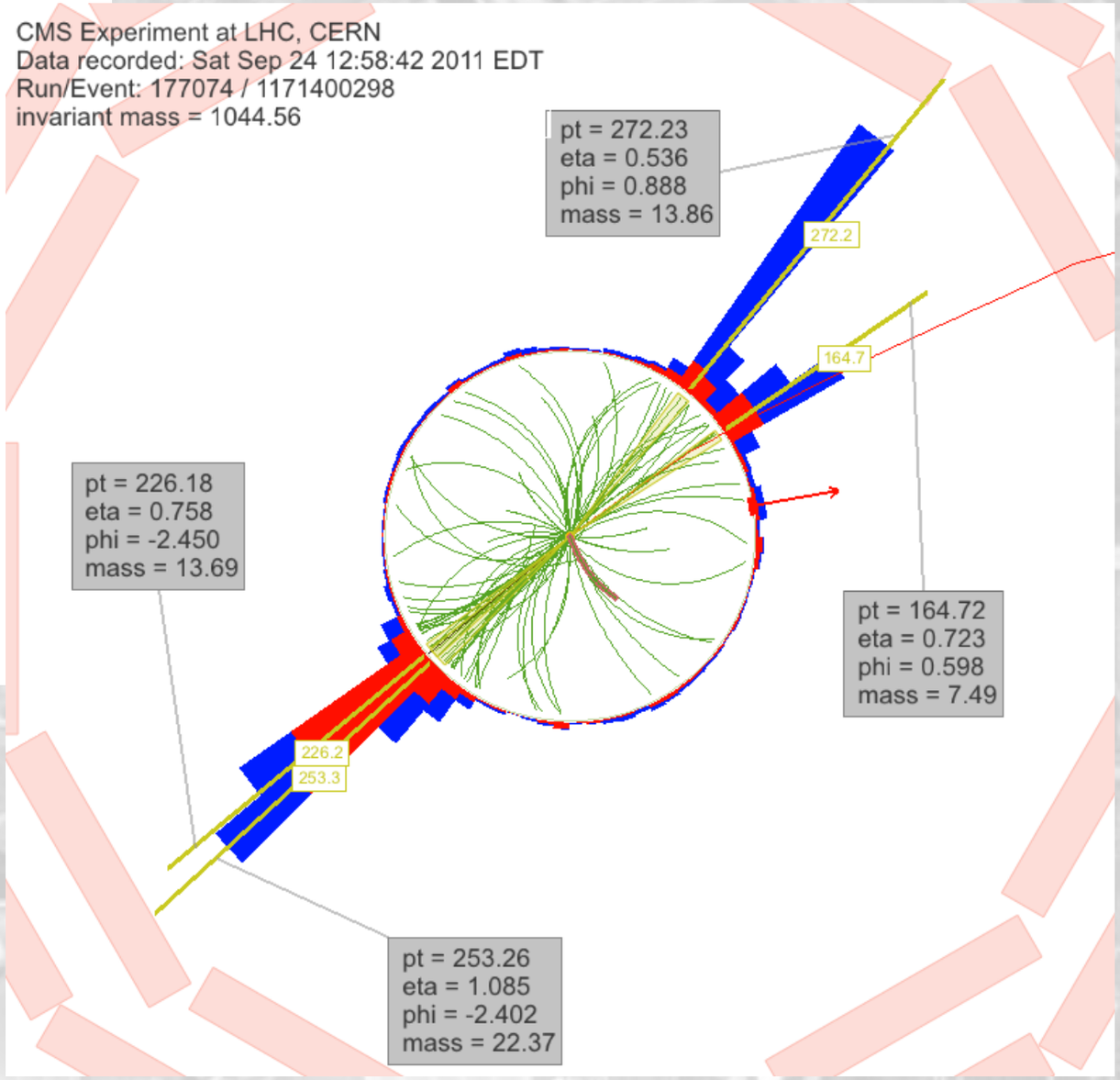




CMS-PAS-JME-13-006

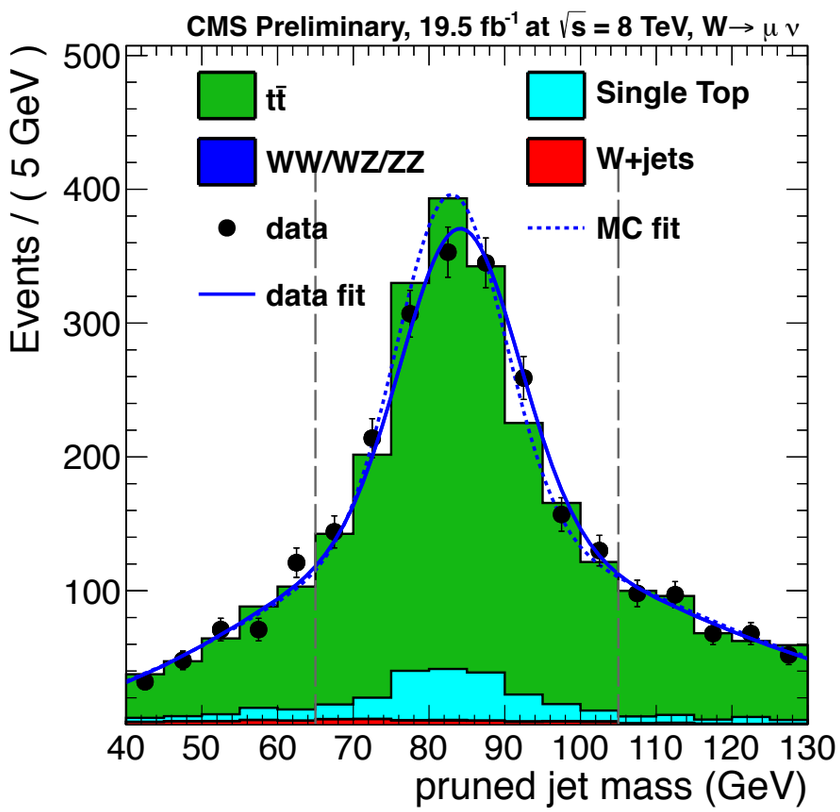
A DOUBLE-TAG EVENT

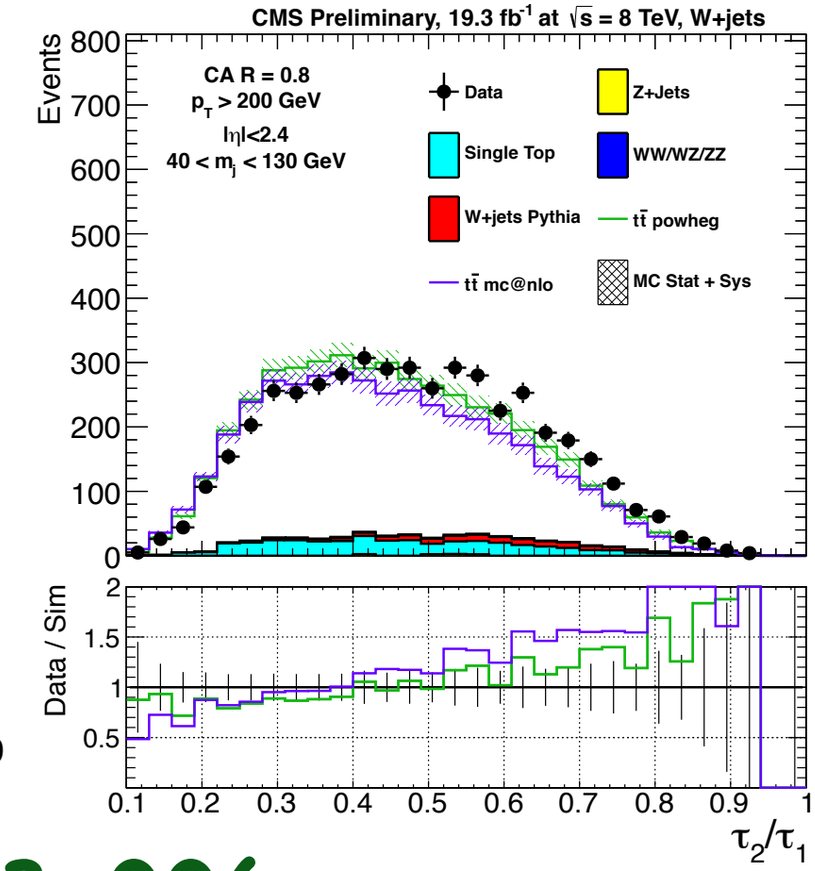


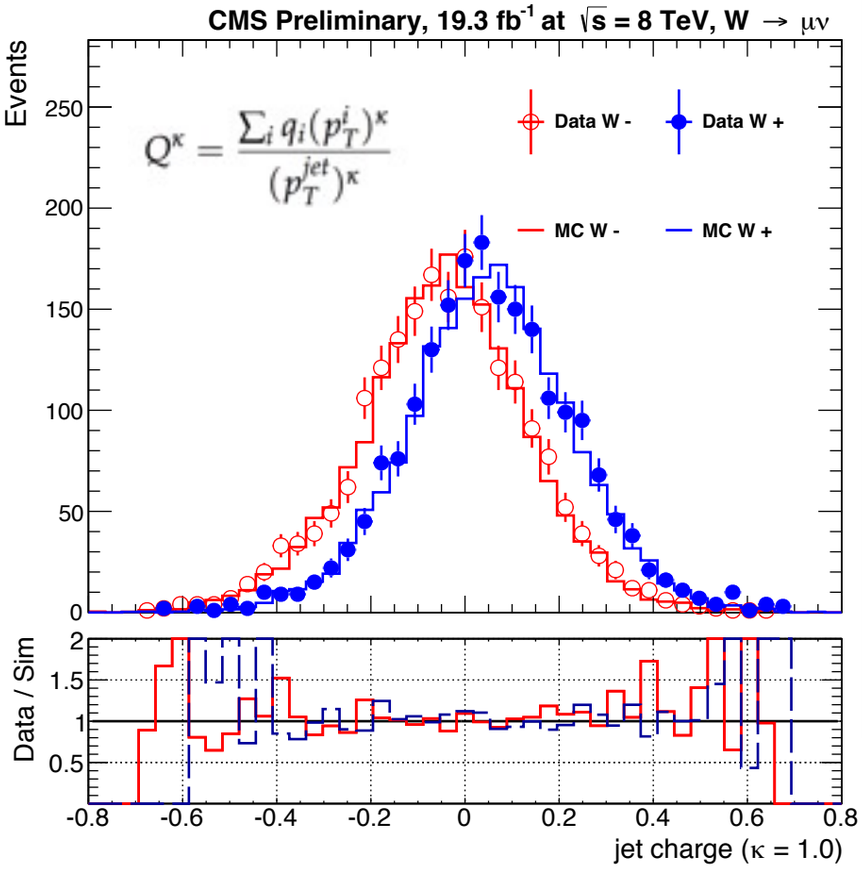


Validation With Data

- Select a sample of boosted tt as one bjet + 1 lepton recoiling against one bjet and one jet (the W candidate)
- Fit the W mass to separate signal from bkg
- Study substructure variables data vs MC
- Derive a signal systematic from data-MC agreement



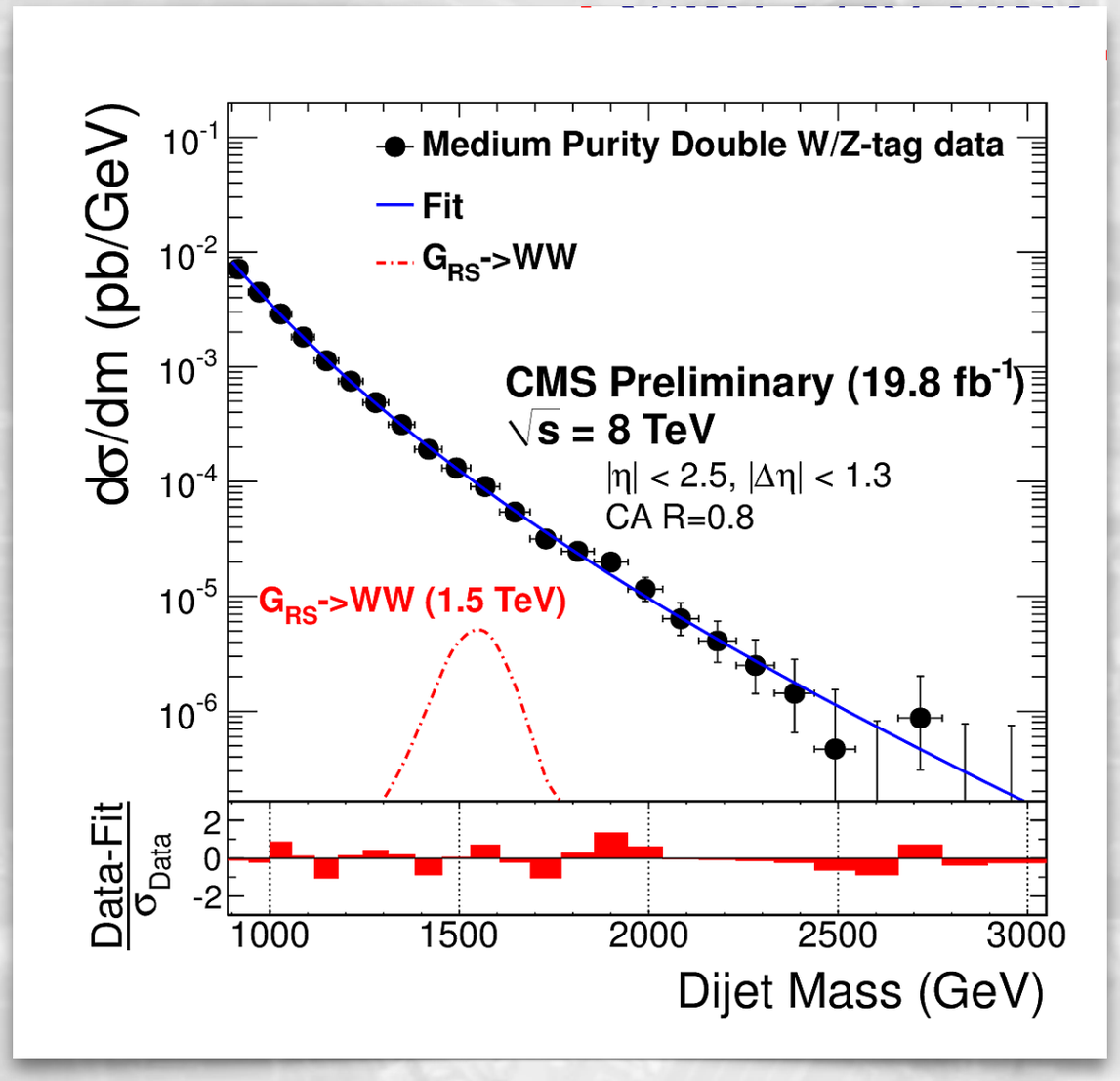


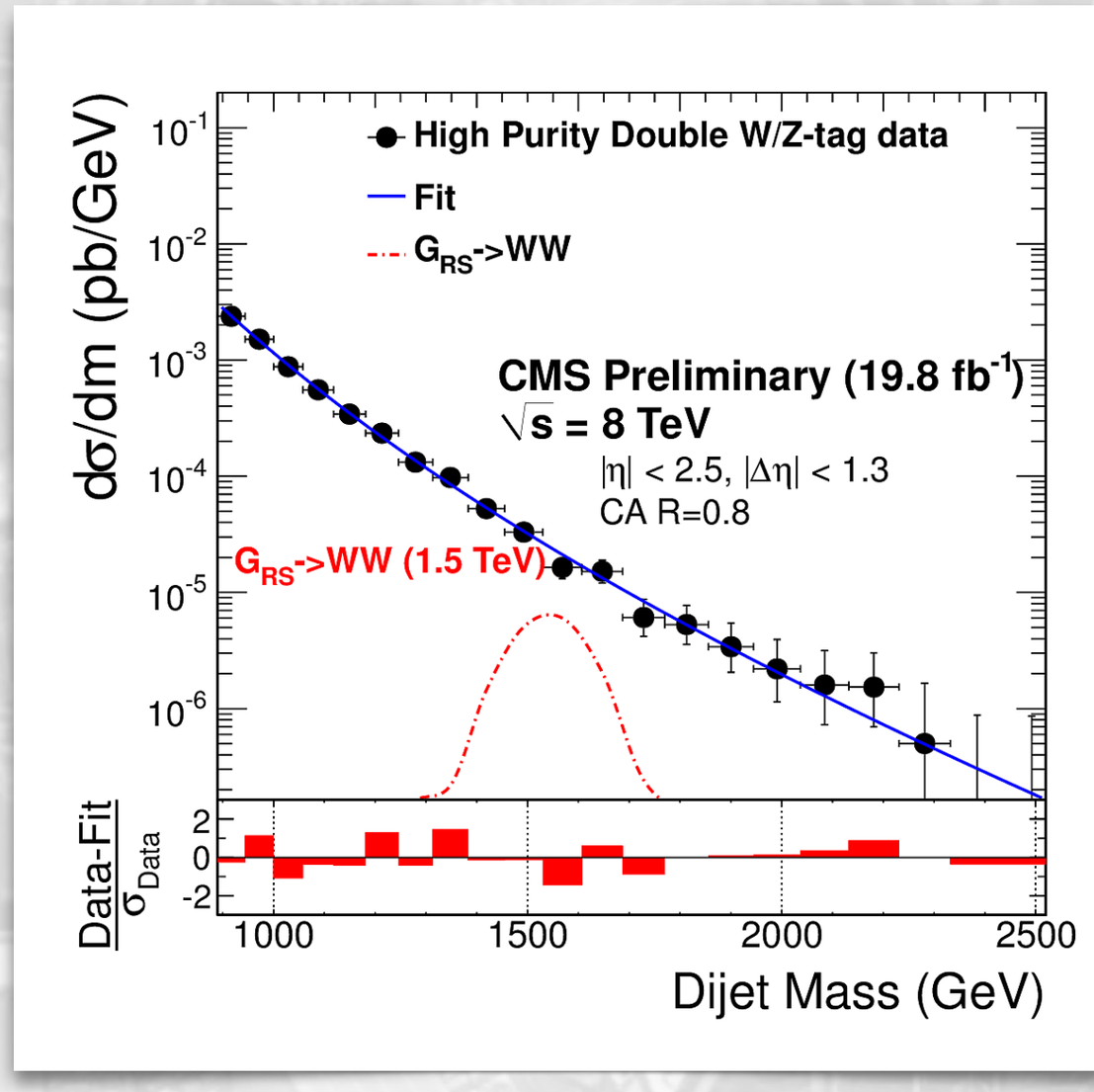


CMS-PAS-JME-13-006

Fully-hadronic X→VV

- Bump hunt in dijet final state after applying V tagging
- 2V tags reduce QCD by a factor 200 retaining 10%

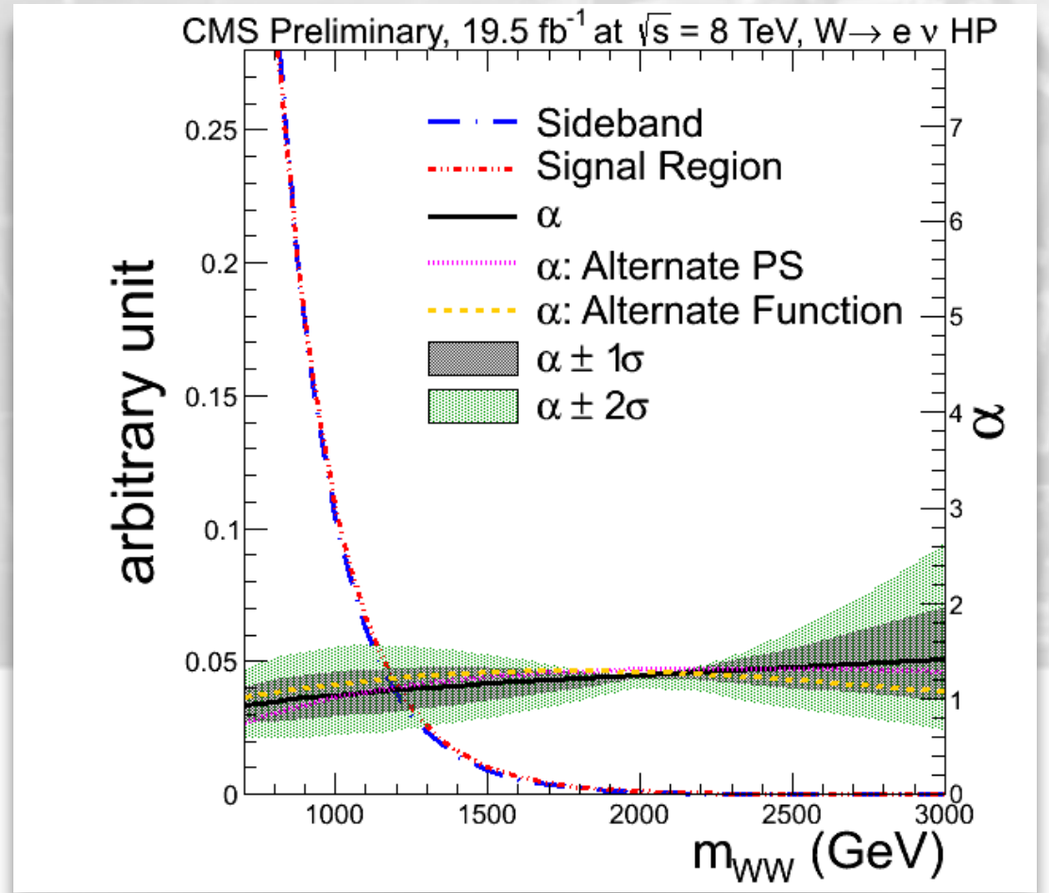




CMS-PAS-EXO-12-024

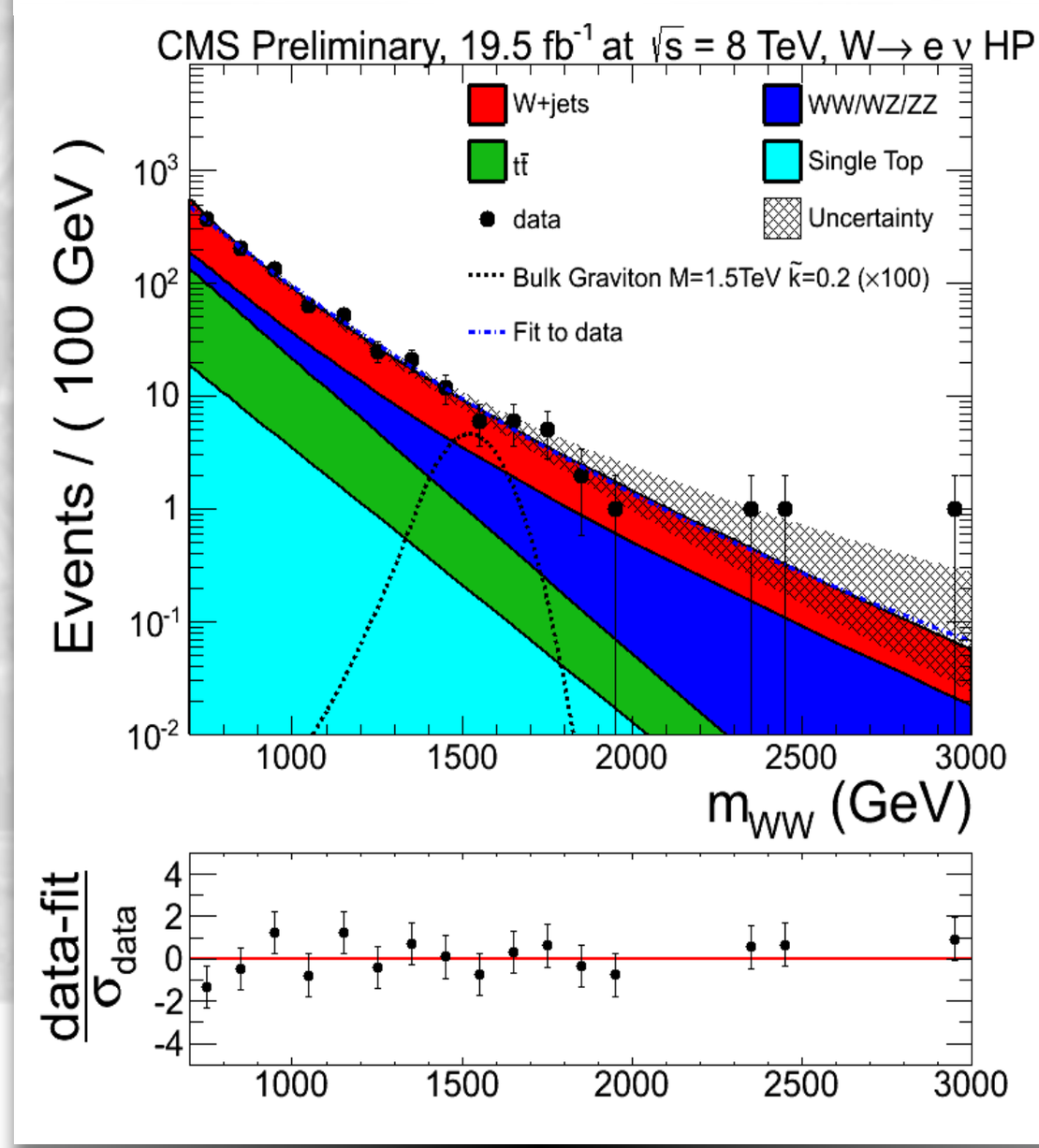
$X \rightarrow W(\ell\nu)V$

- Predict the bkg from the jet mass sideband
- Use the MC to transfer the mWW mass distribution from the sideband to the signal region



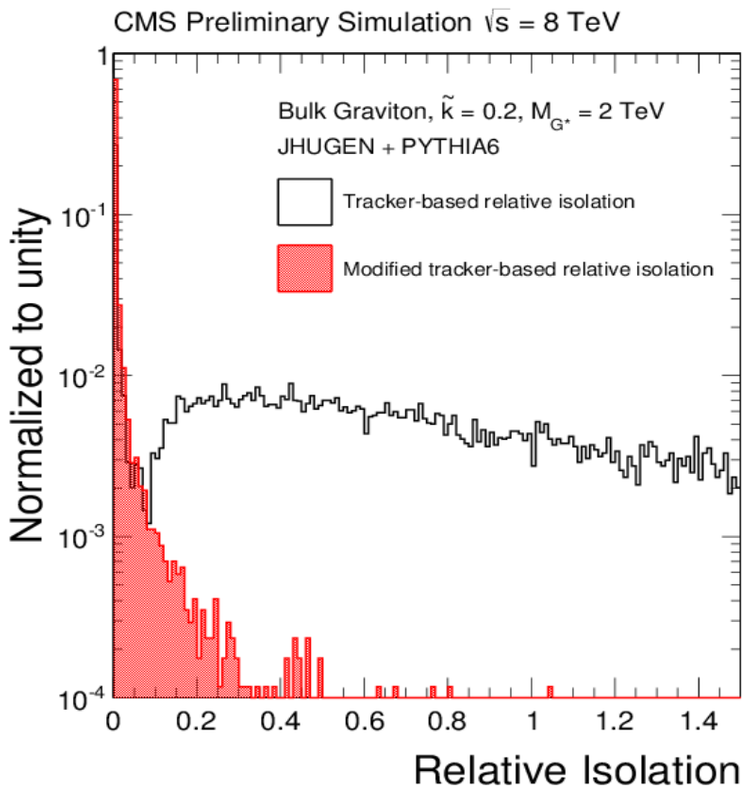
CMS-PAS-EXO-12-021

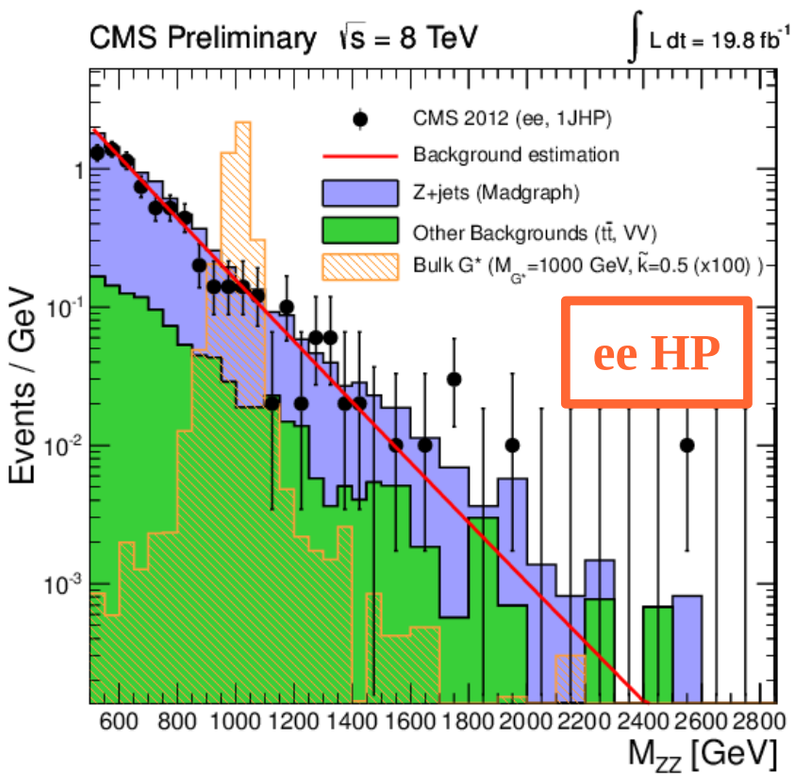
$$\alpha = \frac{N^{MC-Bkg}_{Sig-Reg}}{N^{MC}_{SB-Reg}} \qquad N^{DATA-Bkg}_{Sig-Reg} = \alpha \bullet N^{DATA}_{SB-Reg}$$

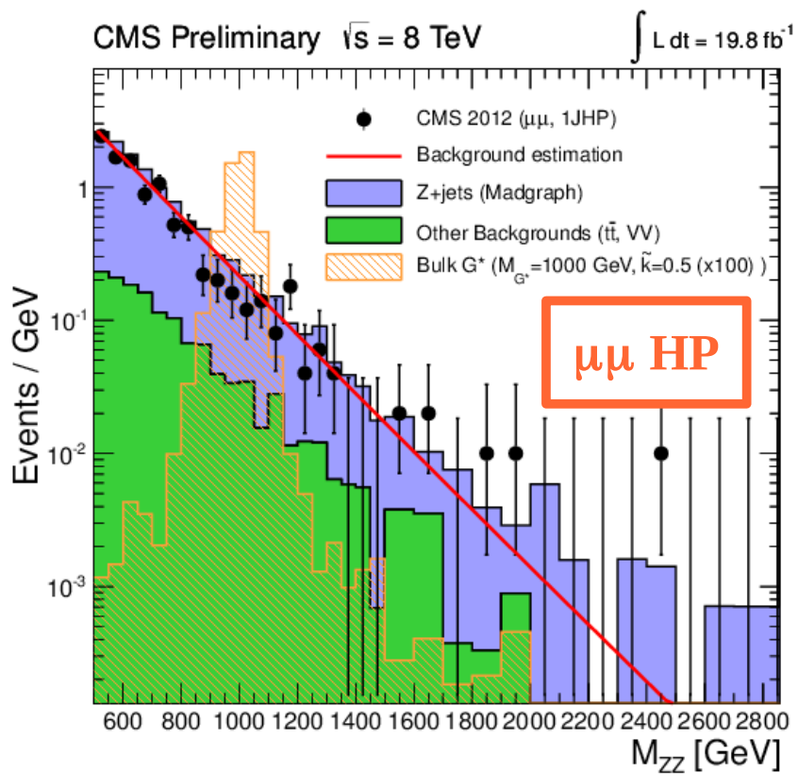


$X \rightarrow Z(\ell\ell)V$

- Similar to the other analyses, with the complication of two collinear leptons
- Standard lepton ID changed to account for the "interference" between the isolation



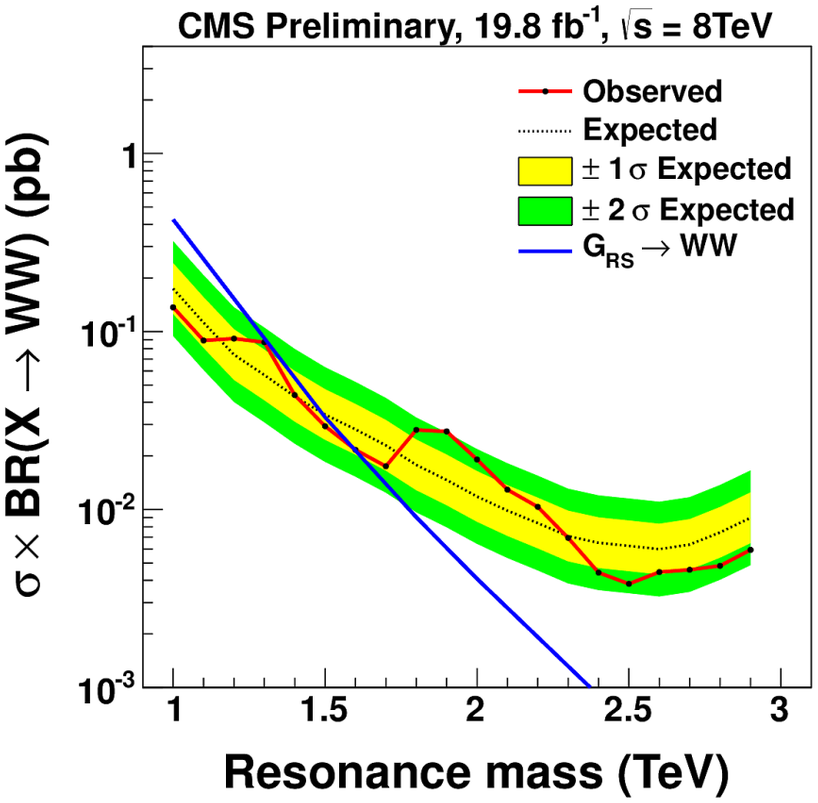




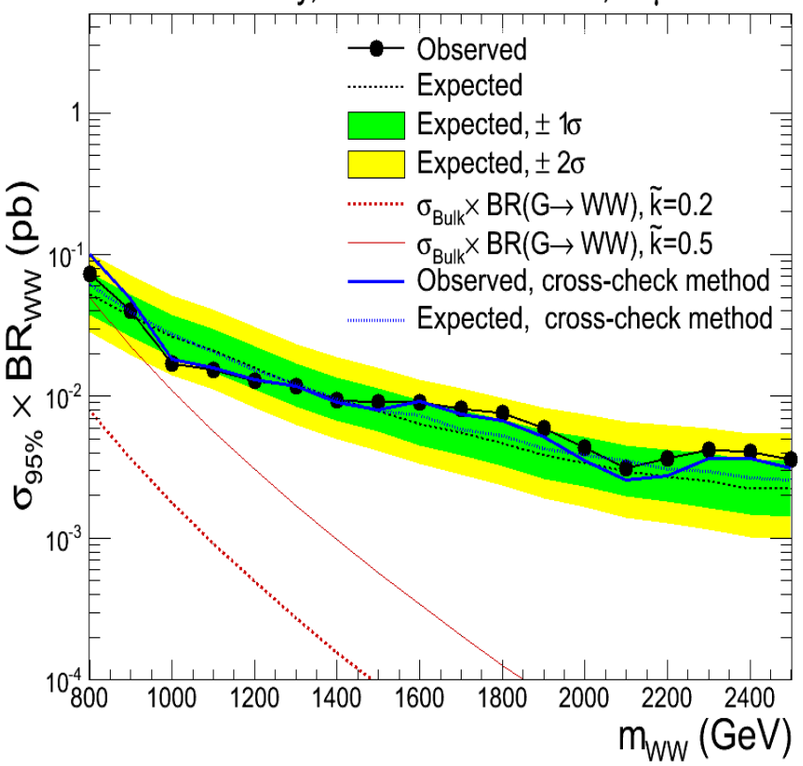
CMS-PAS-EXO-12-022

Limits Compared

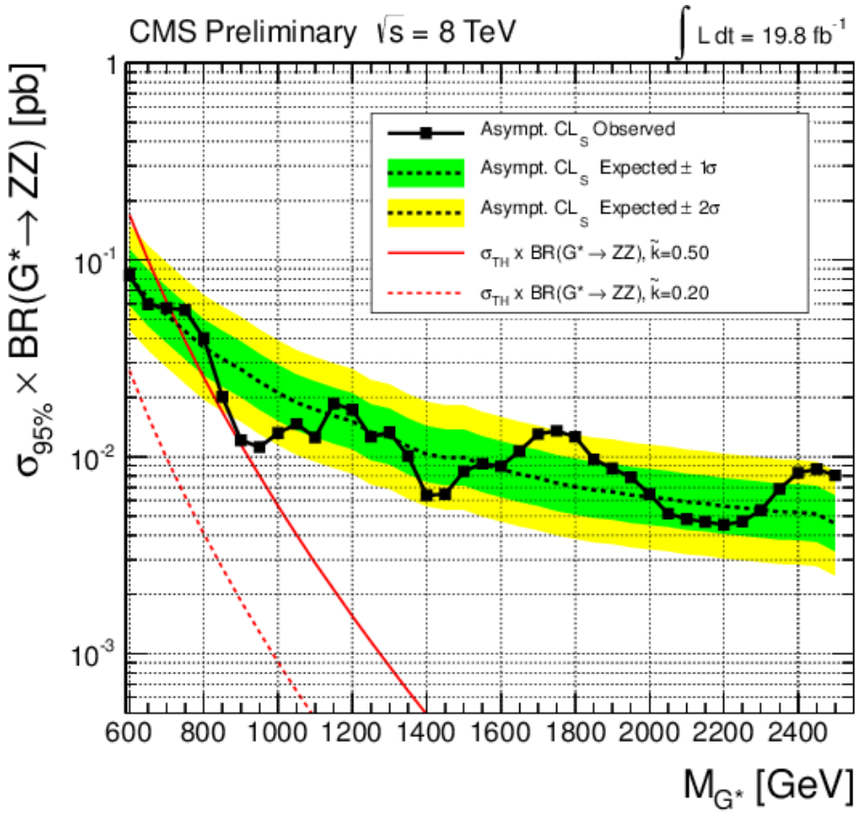
VV → fully hadronic



WV → semileptonic



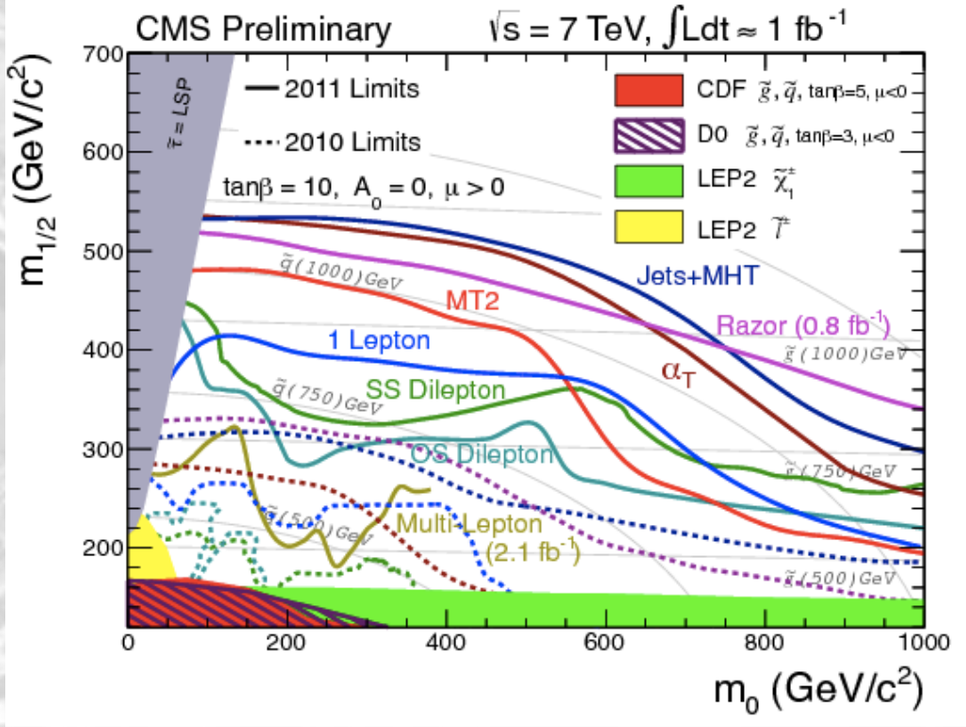
ZV → semileptonic



Search for natural SUSY with razor

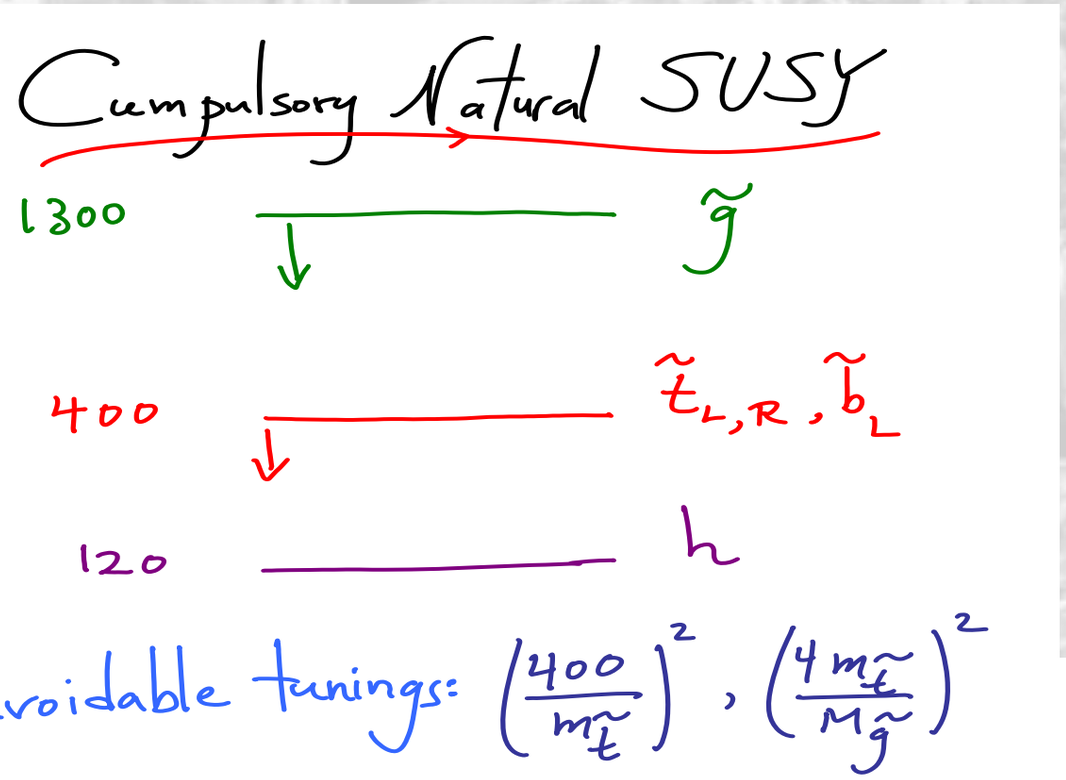
Searching For SUSY@LHC

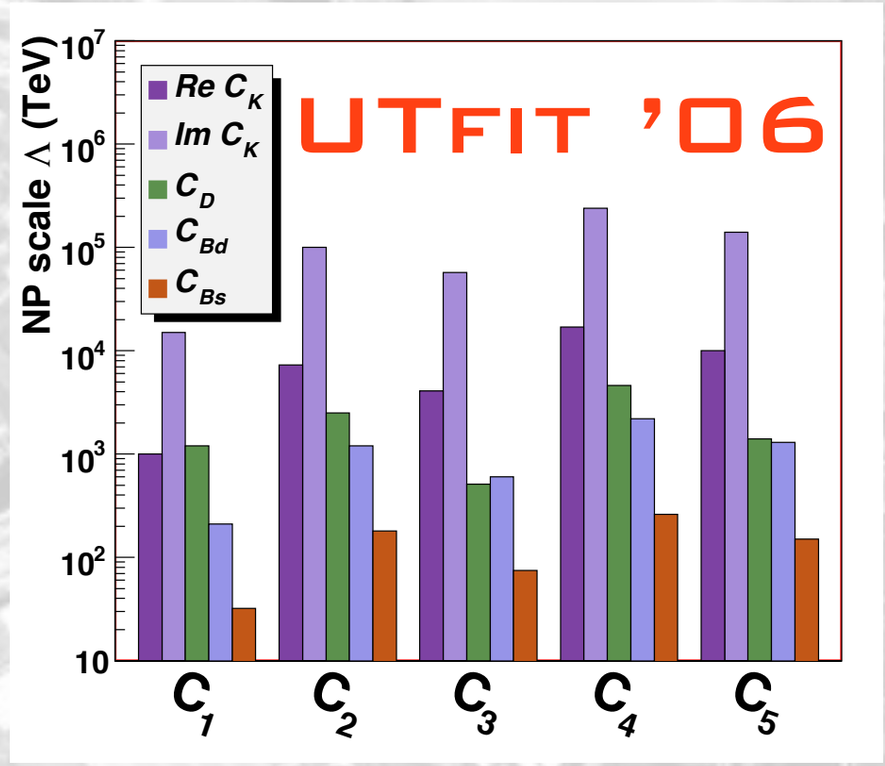
- With the 7TeV run we learned that, If there, SUSY is not as simple as we imagined
- And this was not even unexpected (e.g. EW precision, flavor)



<u>Rethinking the search</u>

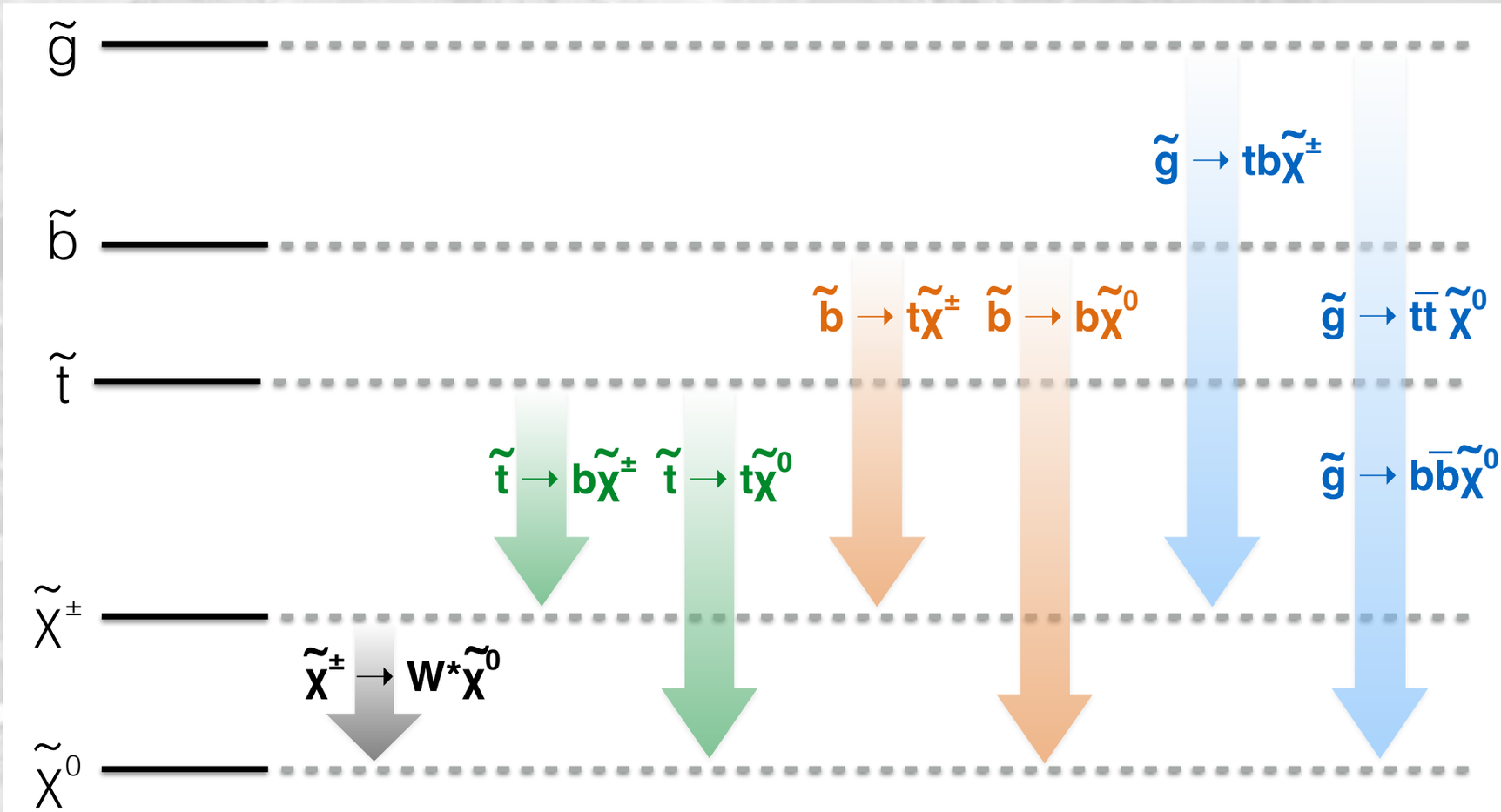
- new motivations (what do we search for)
- an analysis strategy (how?)





One scenario, many signatures

- Multijet final states with many b quarks (4t, 3t1b, 2t2b, 1t3b, 4b) from gluinos
- High-pT leptons from W decays
- Same/opposite charge lepton pairs, with same or different flavour



- Rather than focusing on one signature, we decided to design an inclusive search
- Rather than focusing on the tail of some kinematic distribution, we decided to use a loose selection: more signal, but also more background

The Razor

- The prototype process is squark-squark -> jj + 2 LSP

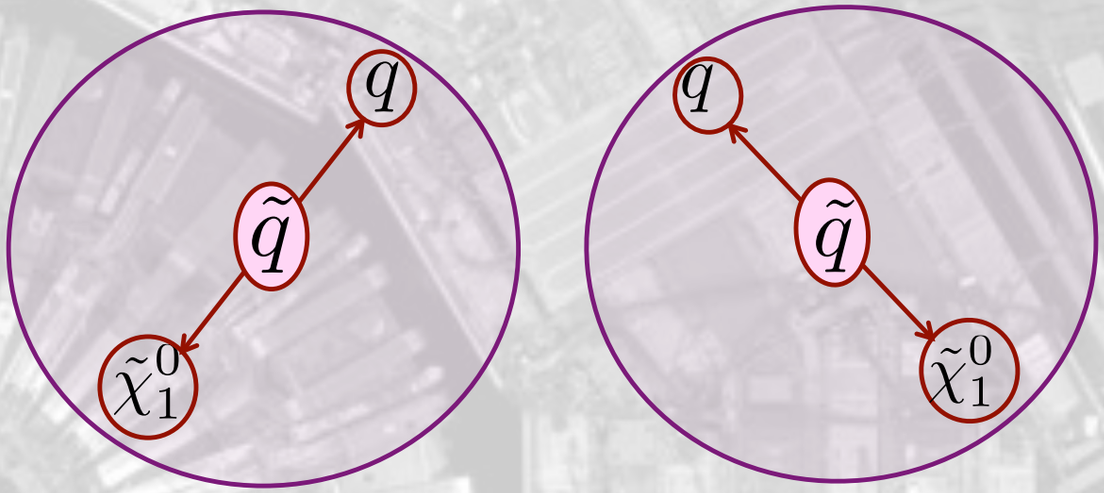
- If we could put the squarks in their rest frames, we would see two jest with same |p|

$$M_\Delta \equiv \frac{M_{\tilde{q}}^2 - M_{\tilde{\chi}}^2}{M_{\tilde{q}}} = 2M_{\tilde{\chi}}\gamma_\Delta\beta_\Delta$$

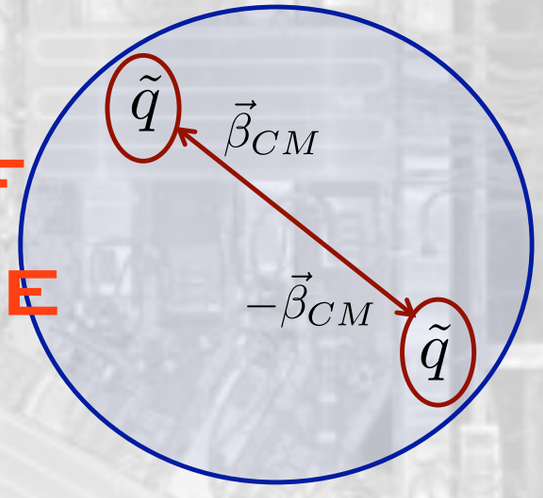
- We observe the jets in the lab frame, boosted by relative squark momentum and partons boost

- We would like to undo the two boosts

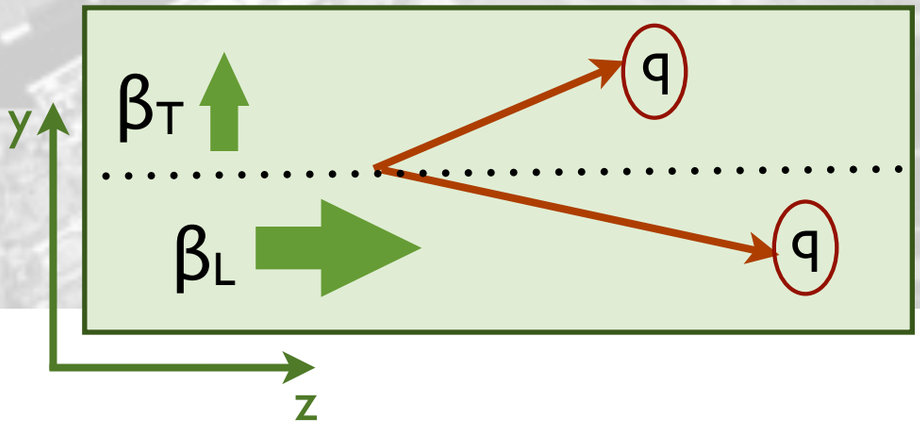
Squark rest frames



Center of mass frame



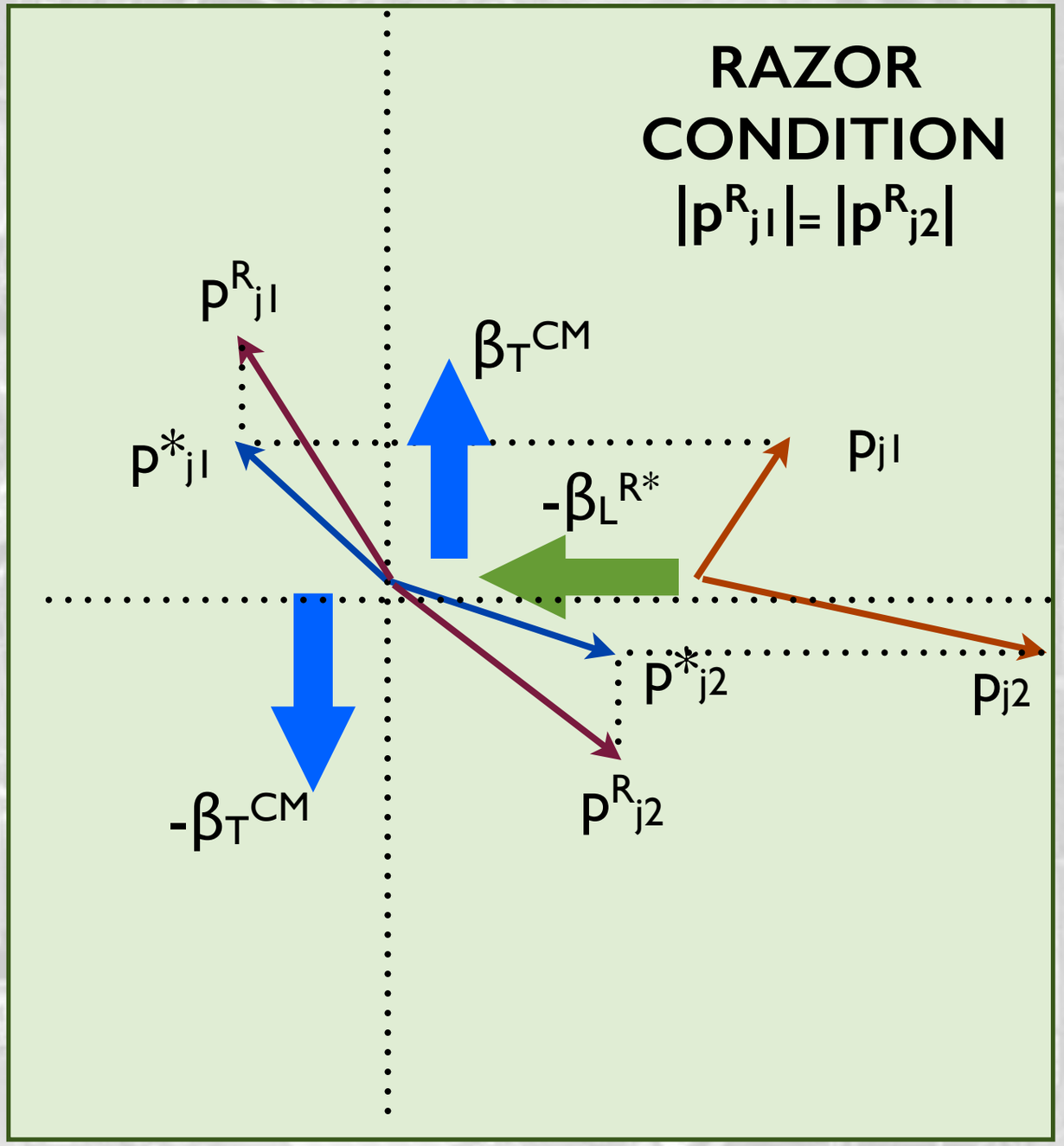
Lab frame



The Razor

- In reality, the best we can do is to compensate the missing degrees of freedom with assumptions on the boost direction

- The parton boost is forced to be longitudinal
- The squark boost in the CM frame is assumed to be transverse



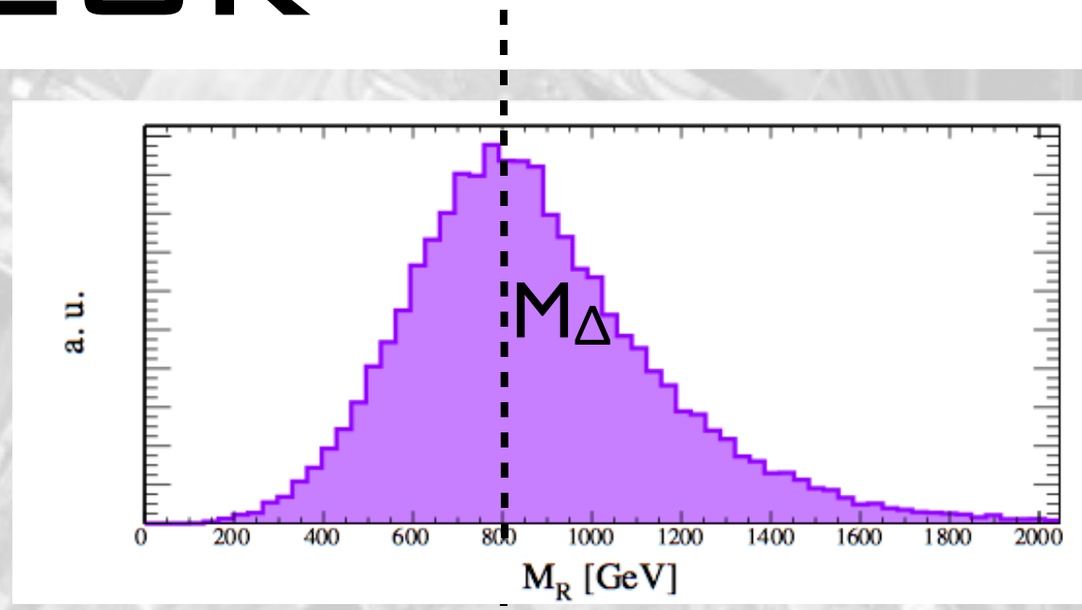
- We require that the two jets have the same momentum after the transformation, and we solve for the boost

- The transformed momentum defines the M_R variable

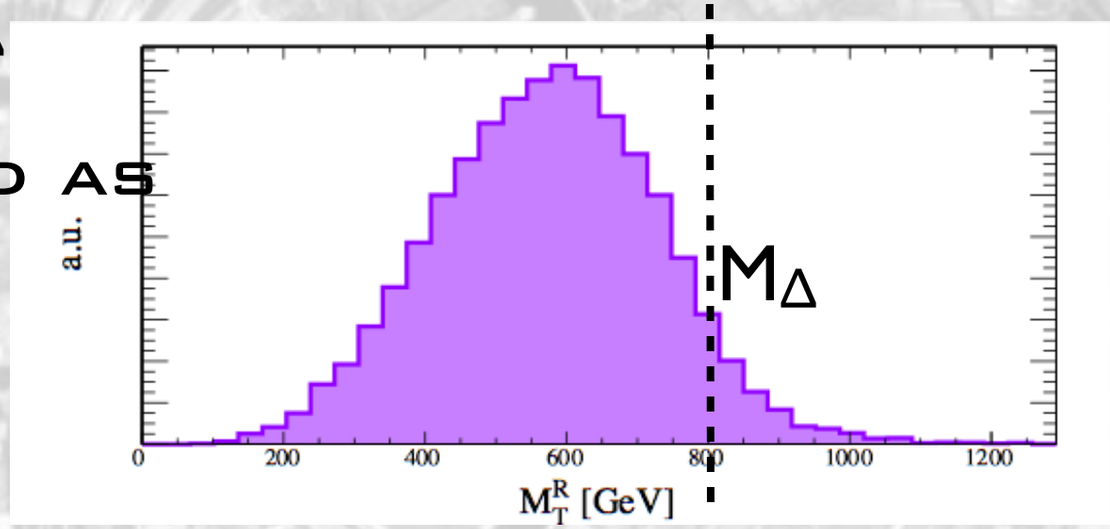
$$M_R \equiv \sqrt{(E_{j_1} + E_{j_2})^2 - (p_z^{j_1} + p_z^{j_2})^2}$$

The Razor

- M_R is boost invariant, even if defined from 3D momenta
- No information on the MET is used
- The peak of the M_R distribution provides an estimate of M_Δ
- M_Δ could be also estimated as the "edge" of M_T^R



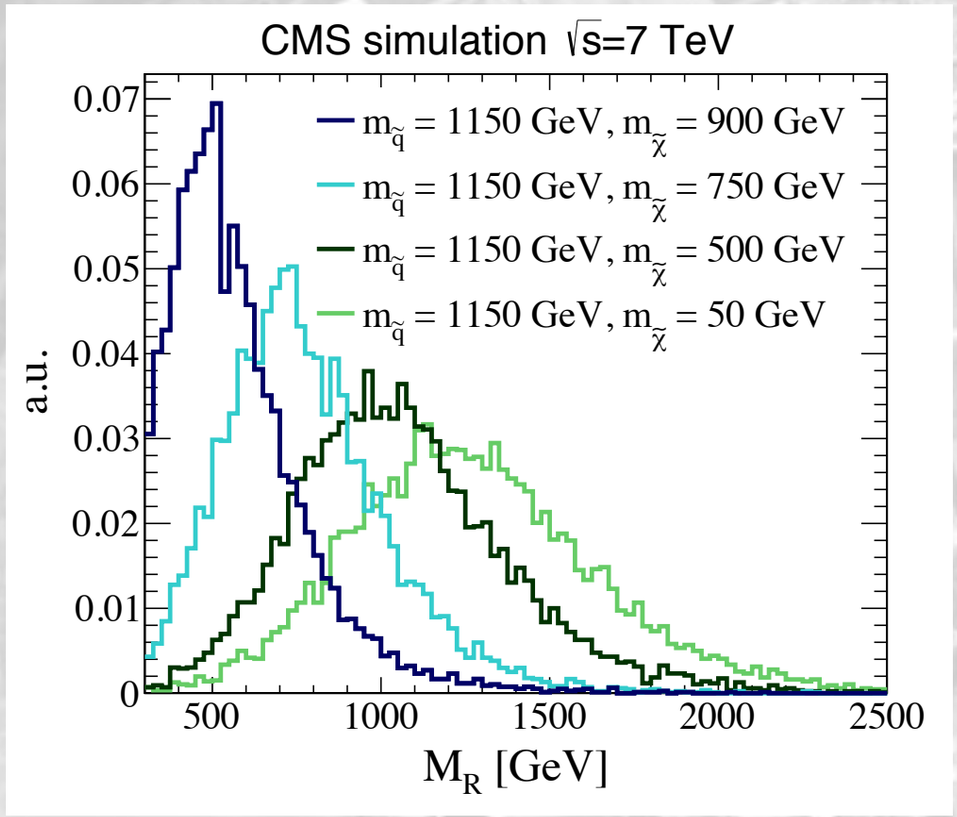
$$M_T^R \equiv \sqrt{\frac{E_T^{miss}(p_T^{j1}+p_T^{j2}) - \vec{E}_T^{miss}\cdot(\vec{p}_T^{\,j1}+\vec{p}_T^{\,j2})}{2}}$$

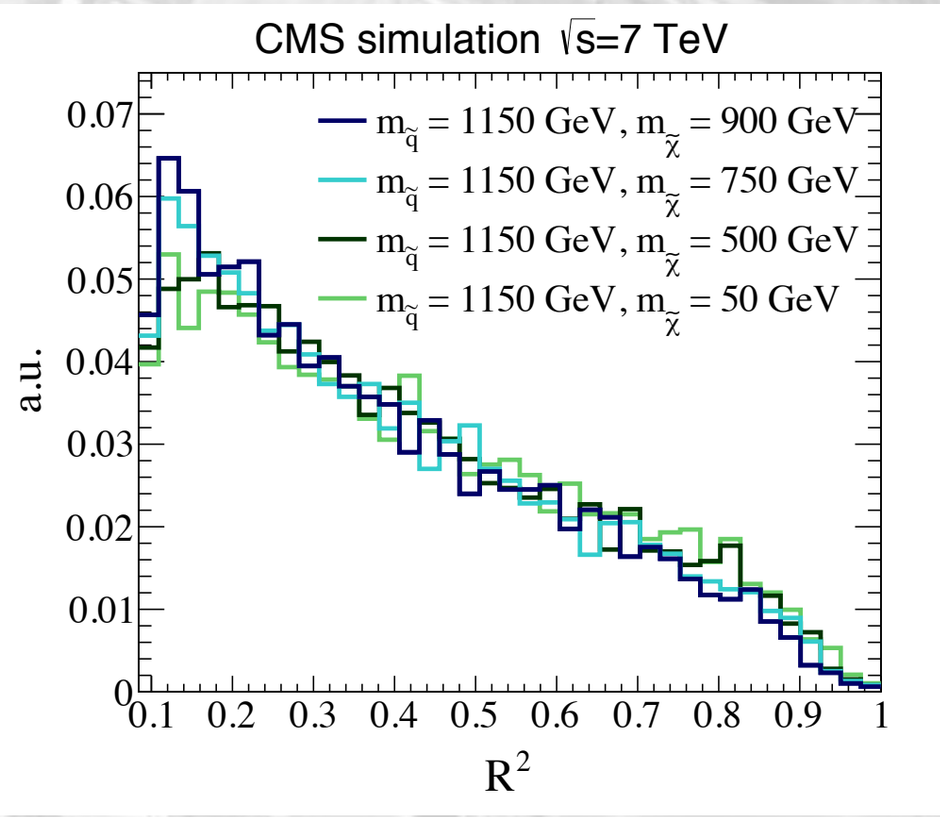


- M_T^R is defined using transverse quantities and it is MET-related
- The Razor (aka R) is defined as the ratio of the two variables

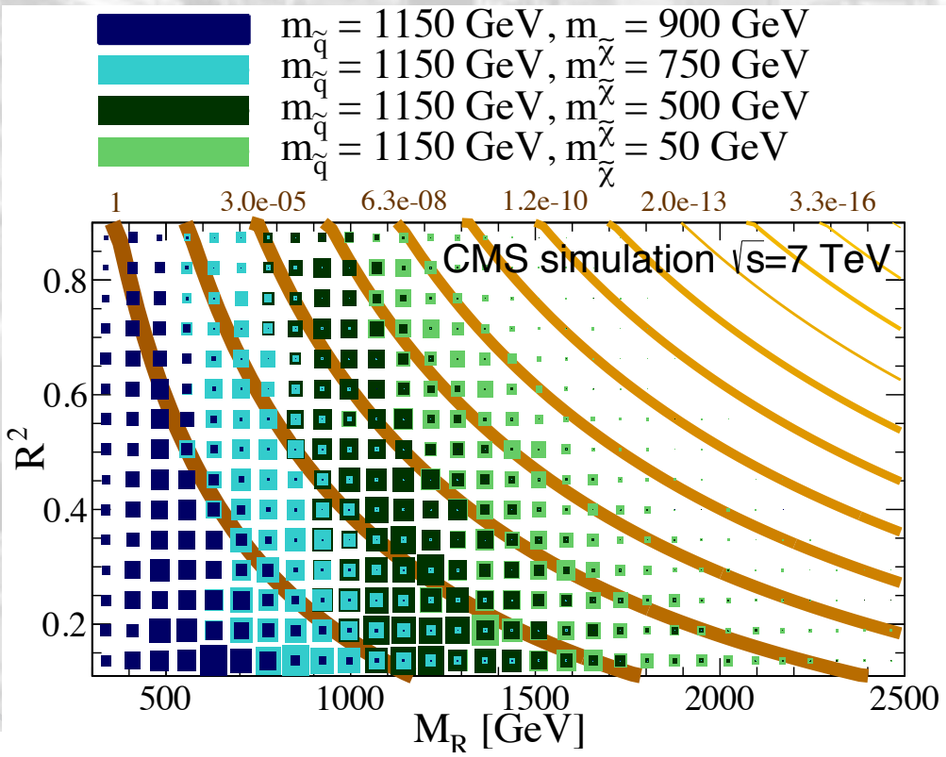
$$R \equiv \frac{M_T^R}{M_R}$$

SUSY Search As a Bump Hunting





- Peaking signal at $M_R \sim M_\Delta$ (discovery and characterisation)
- R^2 has a mild dependence on topology
- correlation allows to tune the amount of background vs signal in the analysis

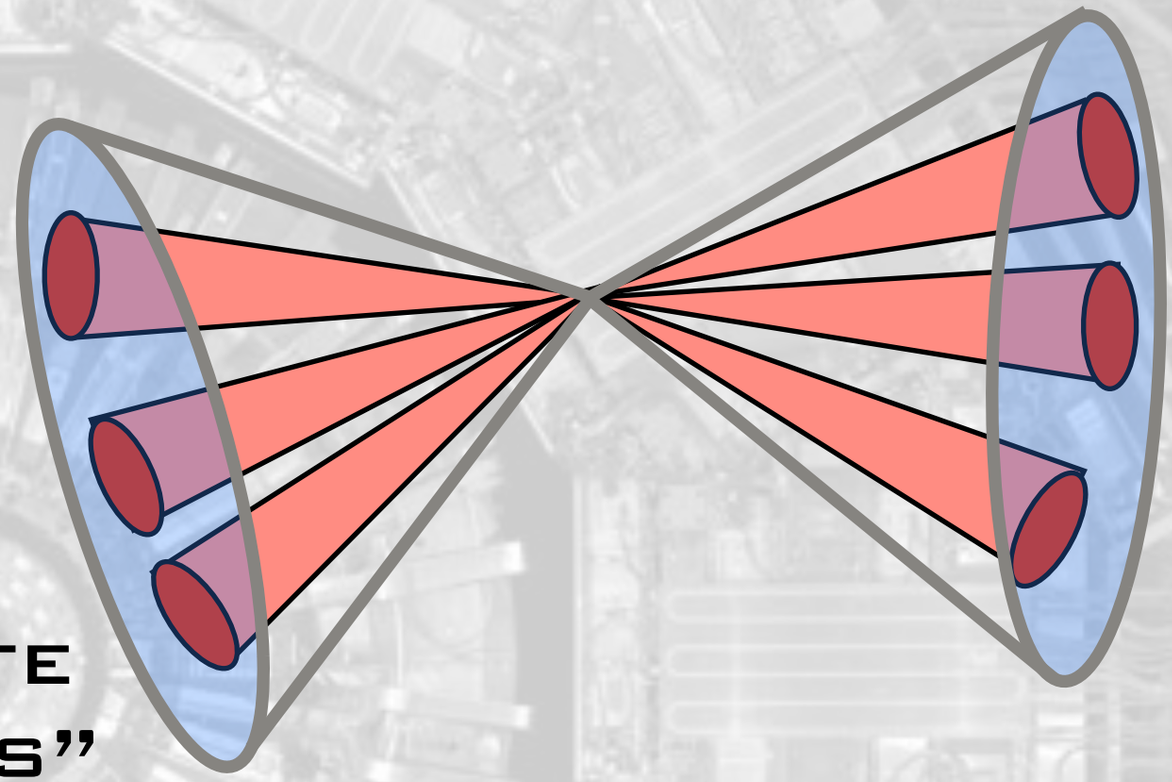


Inclusive search: The Boxes

We target a multijet +MET topology

We built a framework for dijet+MET decays

We consider all the possible ways to separate the jets in two "megajets"



We compute the four-momentum of one megajet as the sum of the four-momenta of jets it contains

We take the configuration for which the megajets give the smallest sum of inv. masses

Inclusive search: The Boxes

Requirements				
Box	lepton	b-tag	kinematic	jet
Two-Lepton Boxes				
MuEle	$\geq$ 1 tight electron and $\geq$ 1 loose muon	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	$\geq$ 2 jets
MuMu	$\geq$ 1 tight muon and $\geq$ 1 loose muon	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	$\geq$ 2 jets
EleEle	$\geq$ 1 tight electron and $\geq$ 1 loose electron	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	$\geq$ 2 jets
Single Lepton Boxes				
MuMultiJet	$\geq$ 1 tight muon	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	$\geq$ 4 jets
MuJet	$\geq$ 1 tight muon	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	2 or 3 jets
EleMultiJet	$\geq$ 1 tight electron	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	$\geq$ 4 jets
EleJet	$\geq$ 1 tight electron	$\geq$ 1 b-tag	($M_R > 300$ GeV and $R^2 > 0.15$) and ($M_R > 450$ GeV or $R^2 > 0.2$)	2 or 3 jets
Hadronic Boxes				
MultiJet	none	$\geq$ 1 b-tag	($M_R > 400$ GeV and $R^2 > 0.25$) and ($M_R > 550$ GeV or $R^2 > 0.3$)	$\geq$ 4 jets
2b-tagged jet	none	$\geq$ 2 b-tag	($M_R > 400$ GeV and $R^2 > 0.25$) and ($M_R > 550$ GeV or $R^2 > 0.3$)	2 or 3 jets

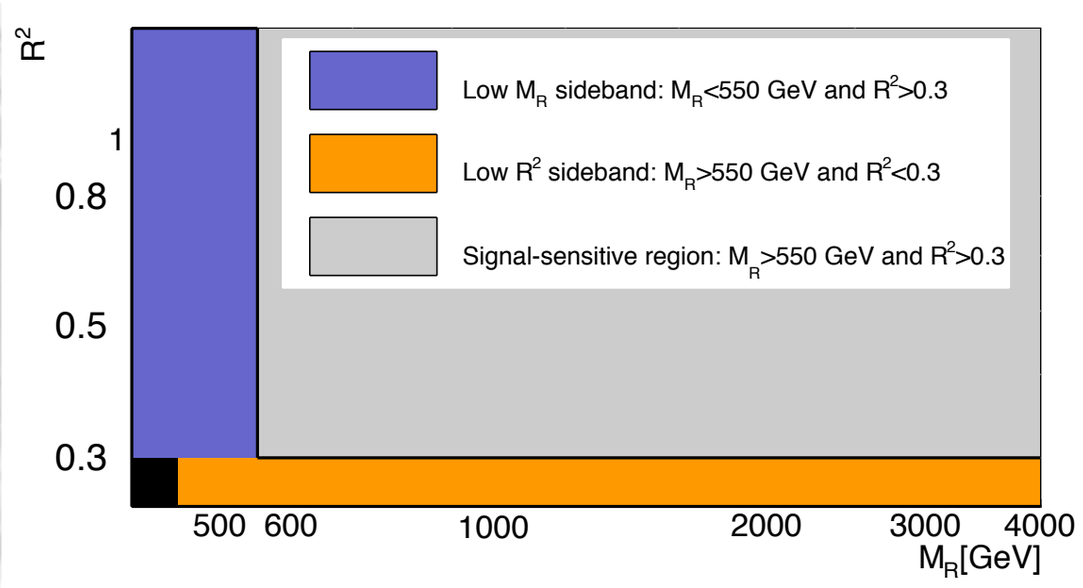
Binned in btag (1b, 2b, $\geq$3b)

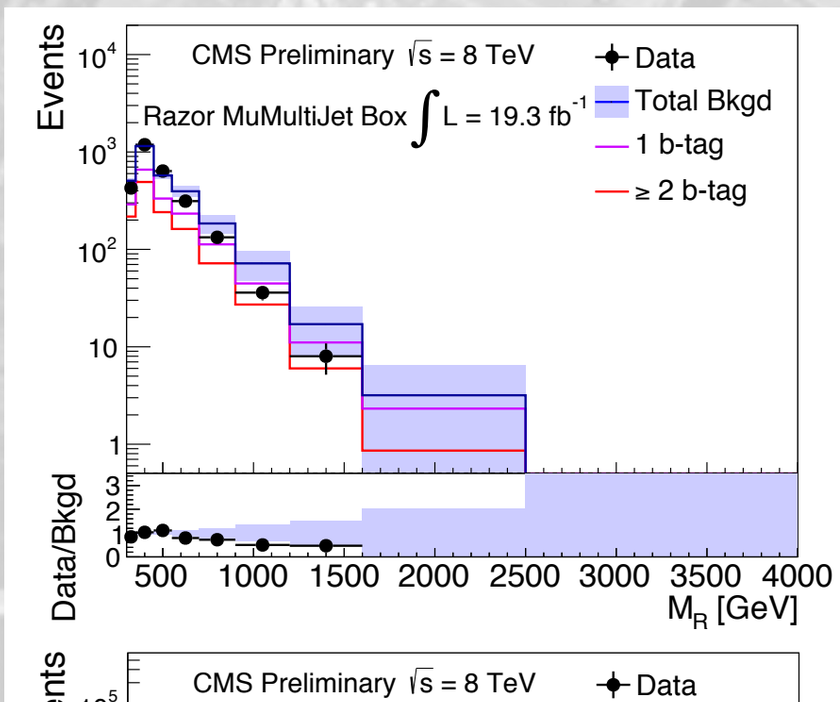
Fit and prediction

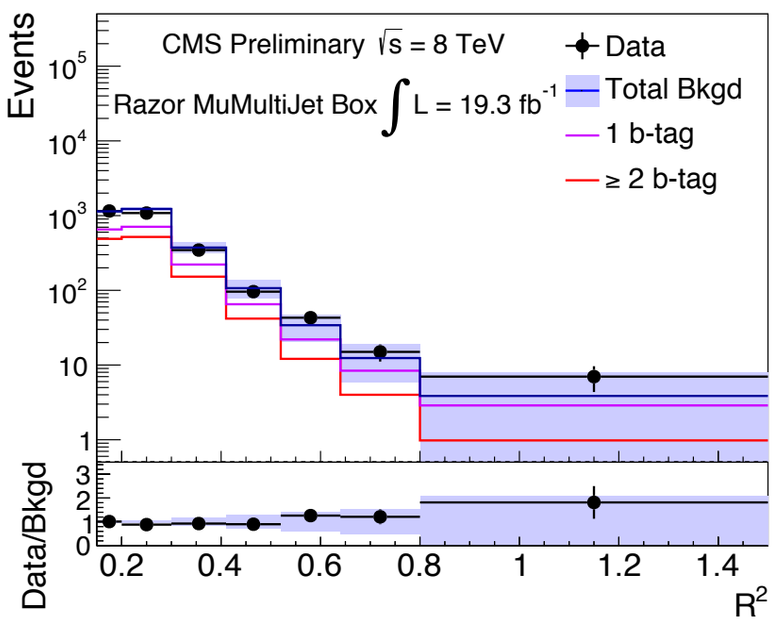
Fit the bkg in the sideband

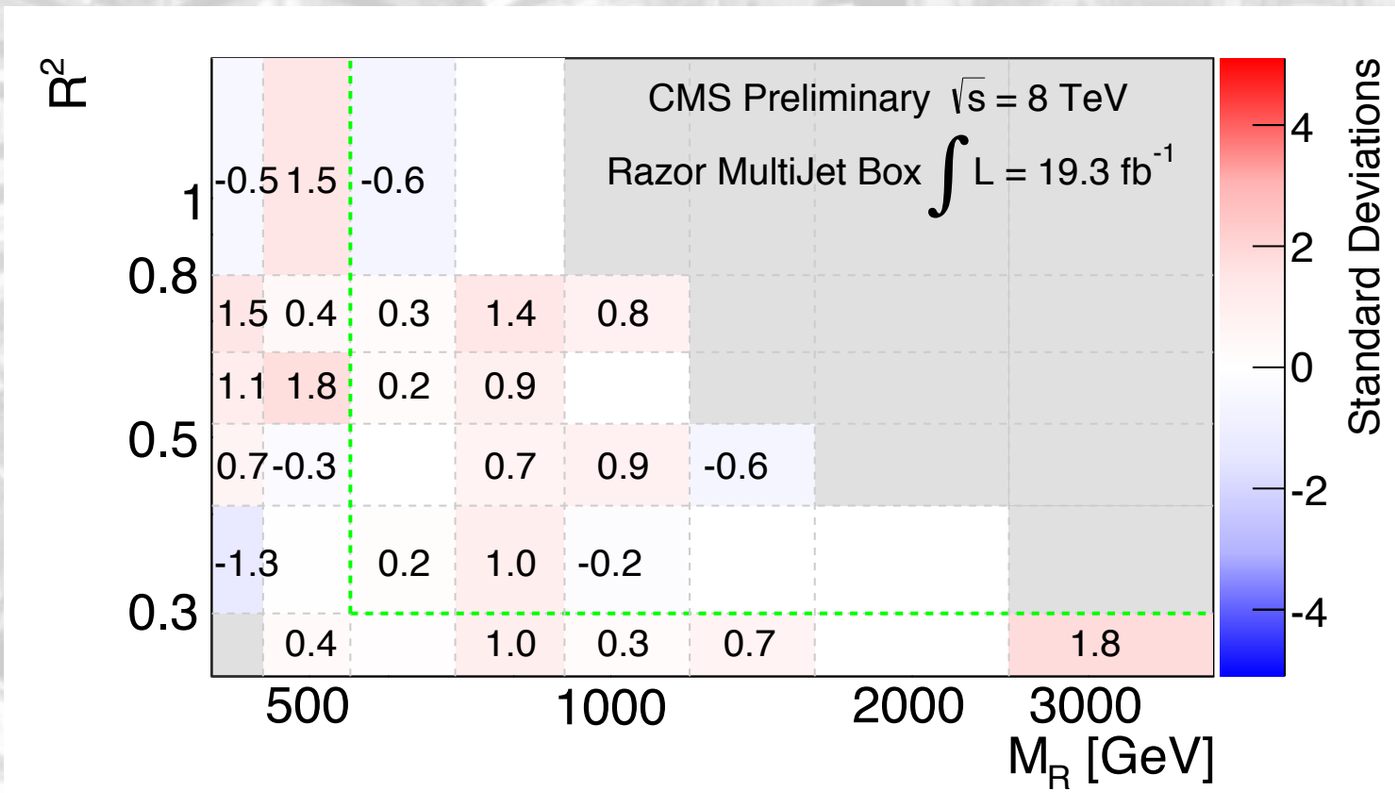
$$F_j(\mathrm{M_R}, \mathrm{R}^2) = \left[k_j(\mathrm{M_R} - M^0_{\mathrm{R},j})(\mathrm{R}^2 - \mathrm{R}^2_{0,j}) - 1\right] \times e^{-k_j(\mathrm{M_R} - M^0_{\mathrm{R},j})(\mathrm{R}^2 - \mathrm{R}^2_{0,j})}$$

Extrapolate to the signal-sensitive region



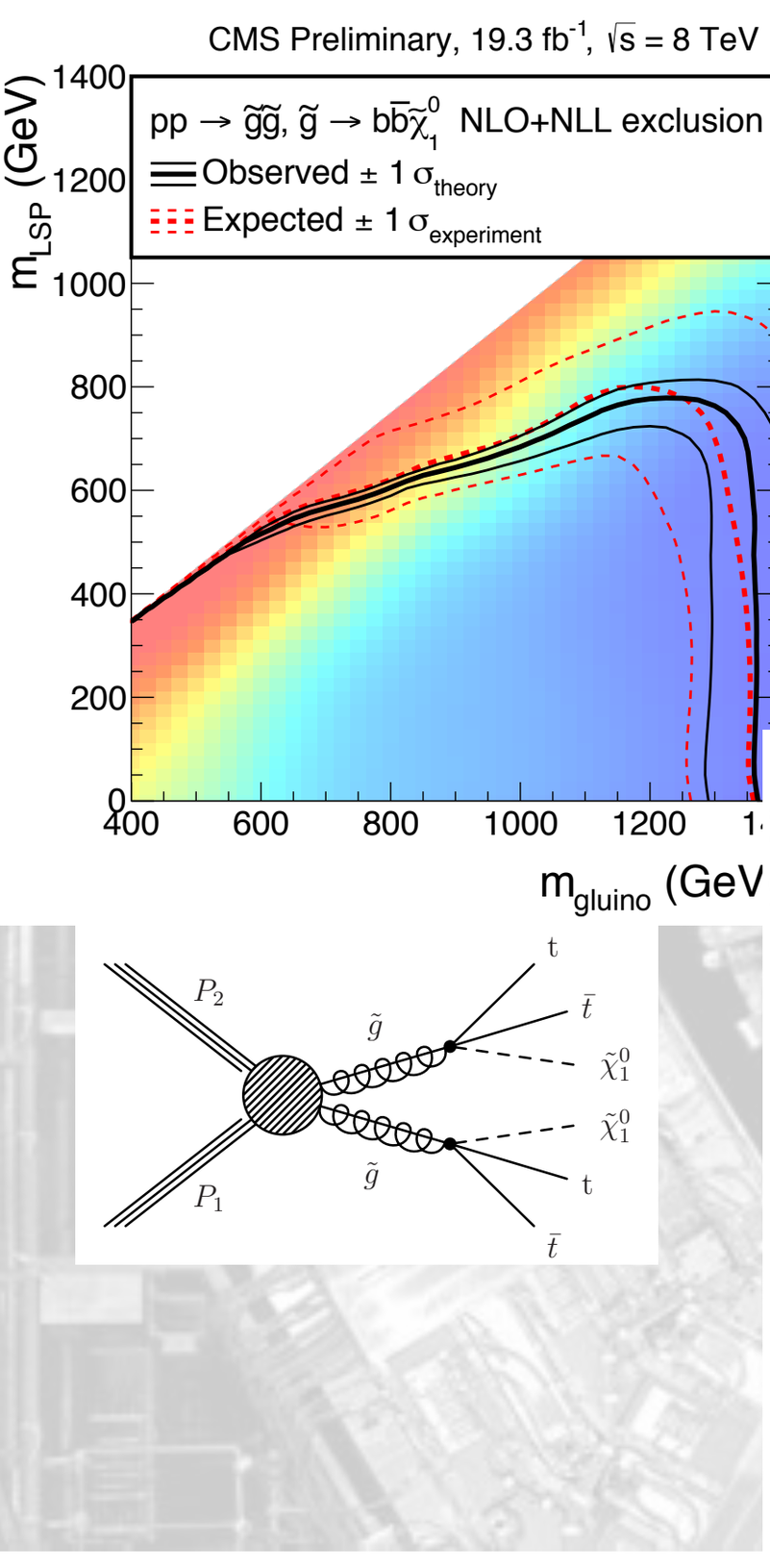


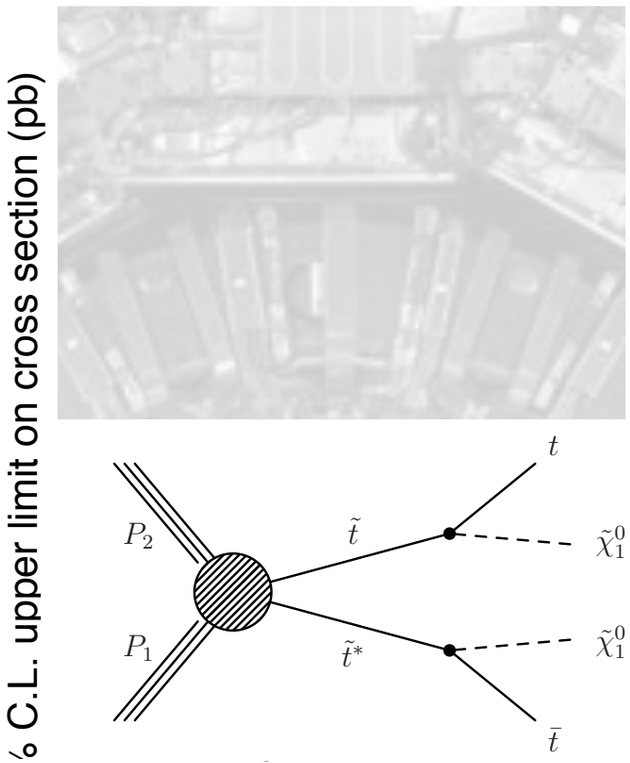


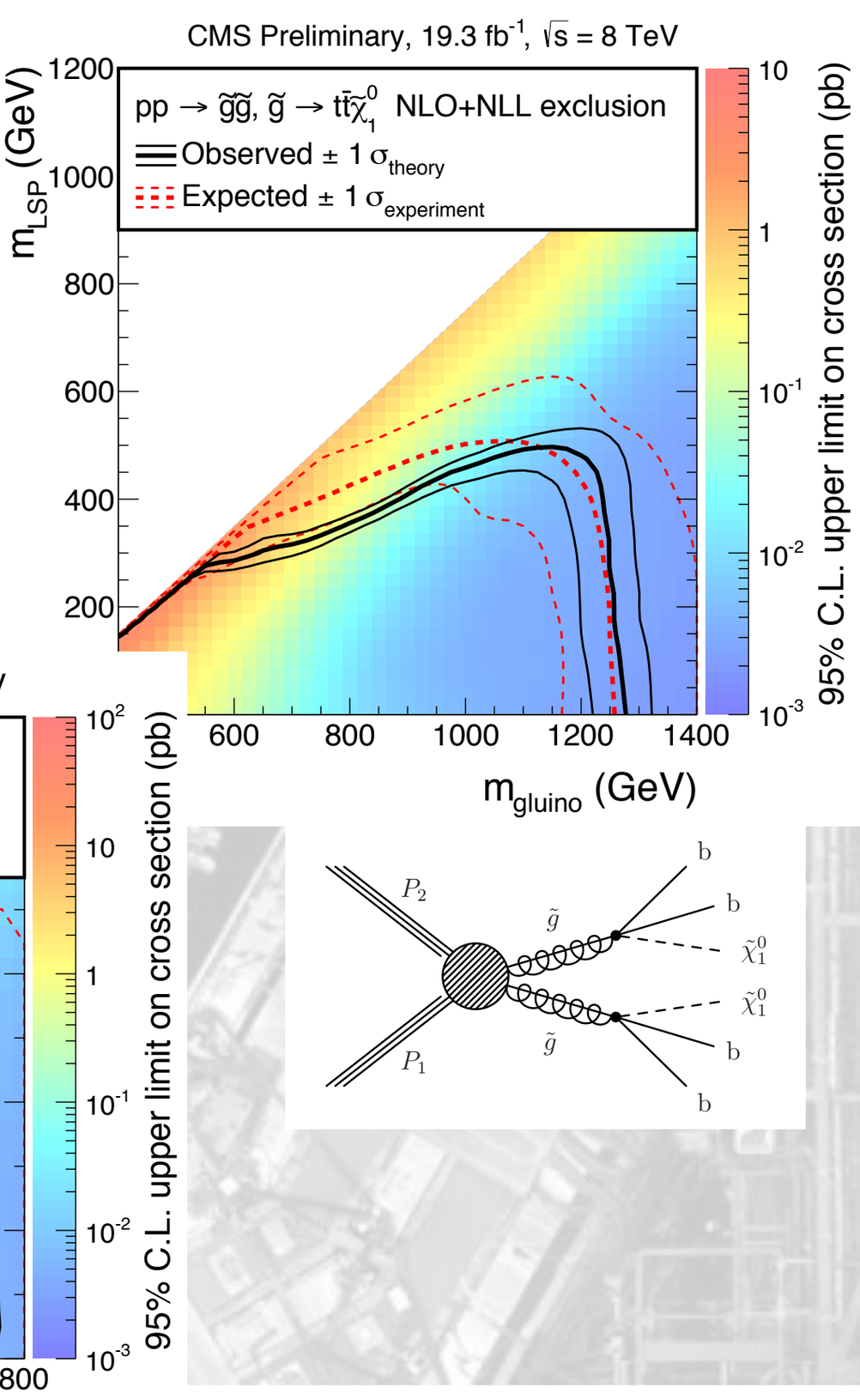


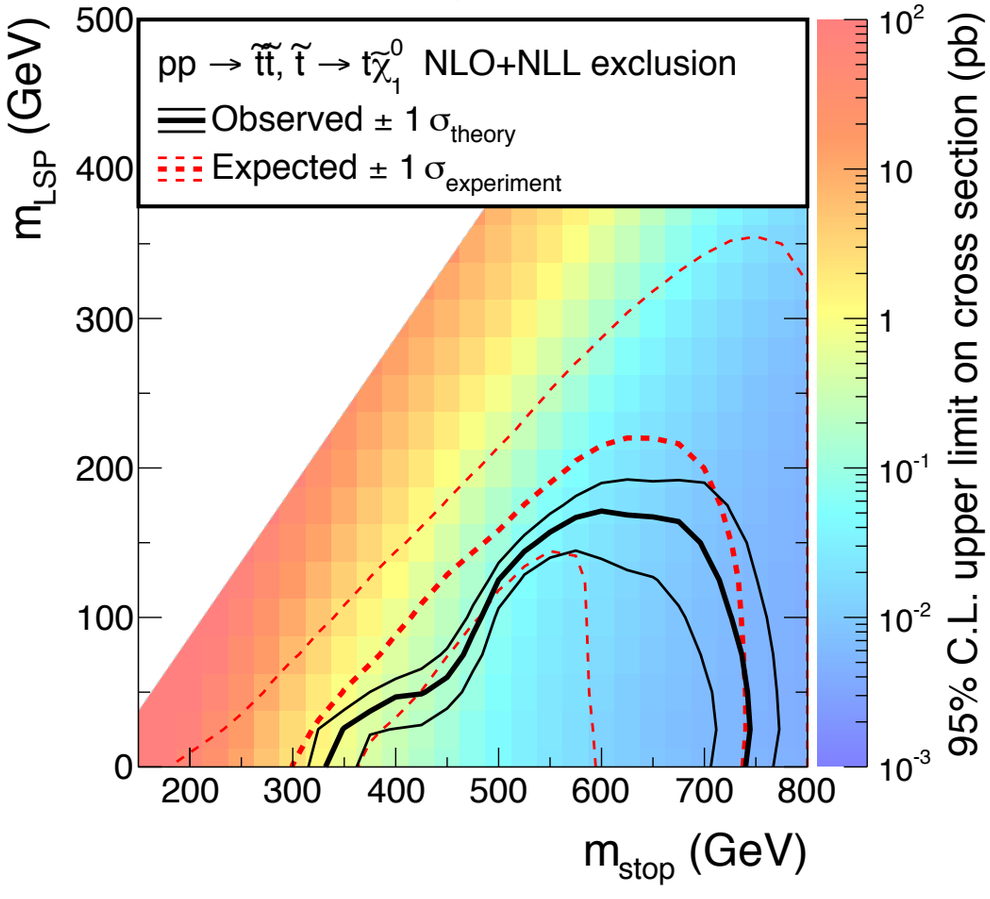
Fit + extrapolation vs data agreement

8 TeV Result


CMS Preliminary, 19.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
pp → g̃g̃, g̃ → bb̄χ̃$_1^0$ NLO+NLL exclusion
Observed ± 1 σ$_{theory}$
Expected ± 1 σ$_{experiment}$
m$_{LSP}$ (GeV)
m$_{gluino}$ (GeV)
95% C.L. upper limit on cross section (pb)
P_2
P_1
$\tilde{g}$
t
$\bar{t}$
$\tilde{\chi}_1^0$


P_2
P_1
$\tilde{t}$
$\tilde{t}^*$
t
$\bar{t}$
$\tilde{\chi}_1^0$
$\tilde{\chi}_1^0$


CMS Preliminary, 19.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
pp → g̃g̃, g̃ → tt̄χ̃$_1^0$ NLO+NLL exclusion
Observed ± 1 σ$_{theory}$
Expected ± 1 σ$_{experiment}$
m$_{LSP}$ (GeV)
m$_{gluino}$ (GeV)
95% C.L. upper limit on cross section (pb)
P_2
P_1
$\tilde{g}$
b
$\tilde{\chi}_1^0$


CMS Preliminary, 19.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
pp → t̃t̃, t̃ → tχ̃$_1^0$ NLO+NLL exclusion
Observed ± 1 σ$_{theory}$
Expected ± 1 σ$_{experiment}$
m$_{LSP}$ (GeV)
m$_{stop}$ (GeV)
95% C.L. upper limit on cross section (pb)

Conclusions

- searches with jets@LHC are challenging but rich. Only a few examples here. Many more out there...

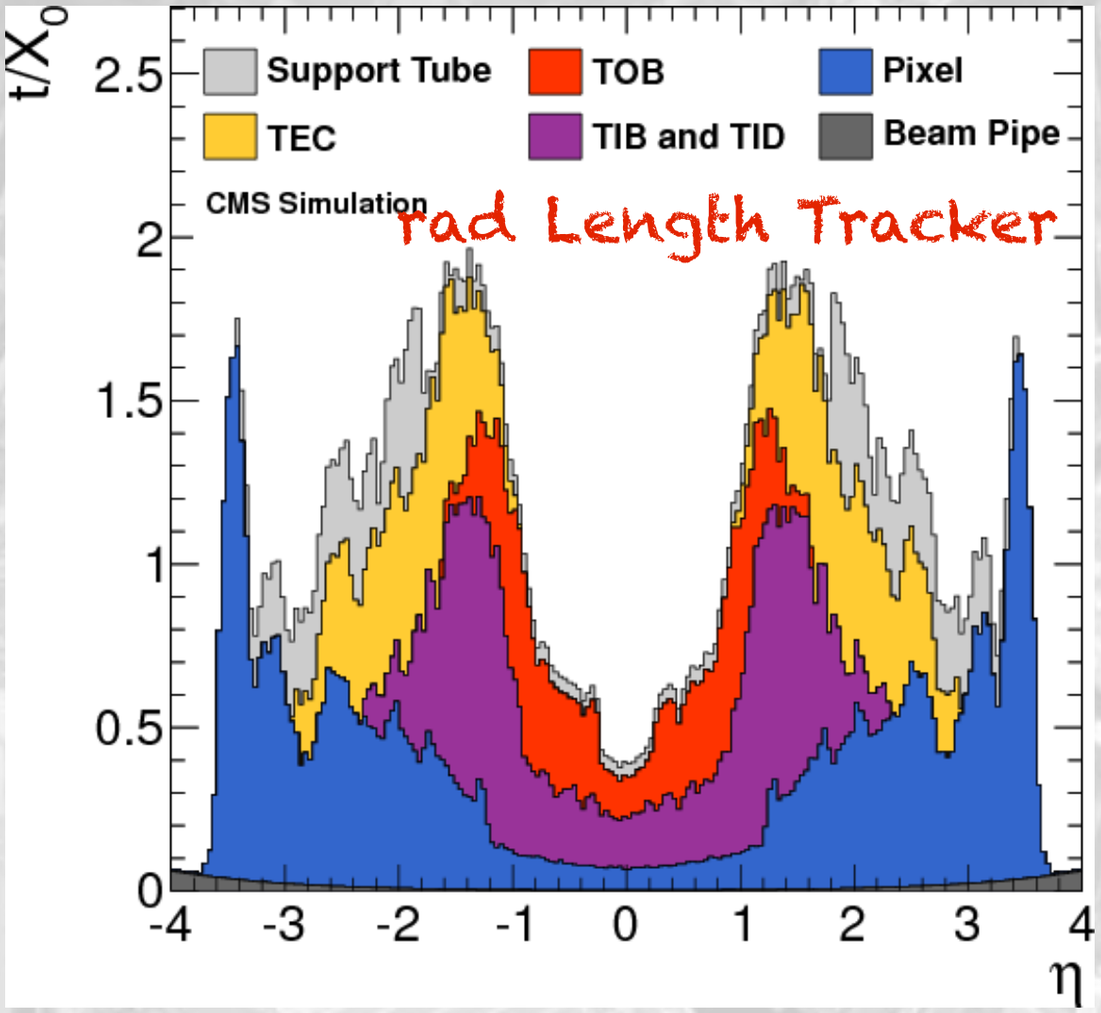
- The environment is challenging, but we can achieve unexpected results when thinking out of the box, twisting the rules and pushing the detectors beyond its limits

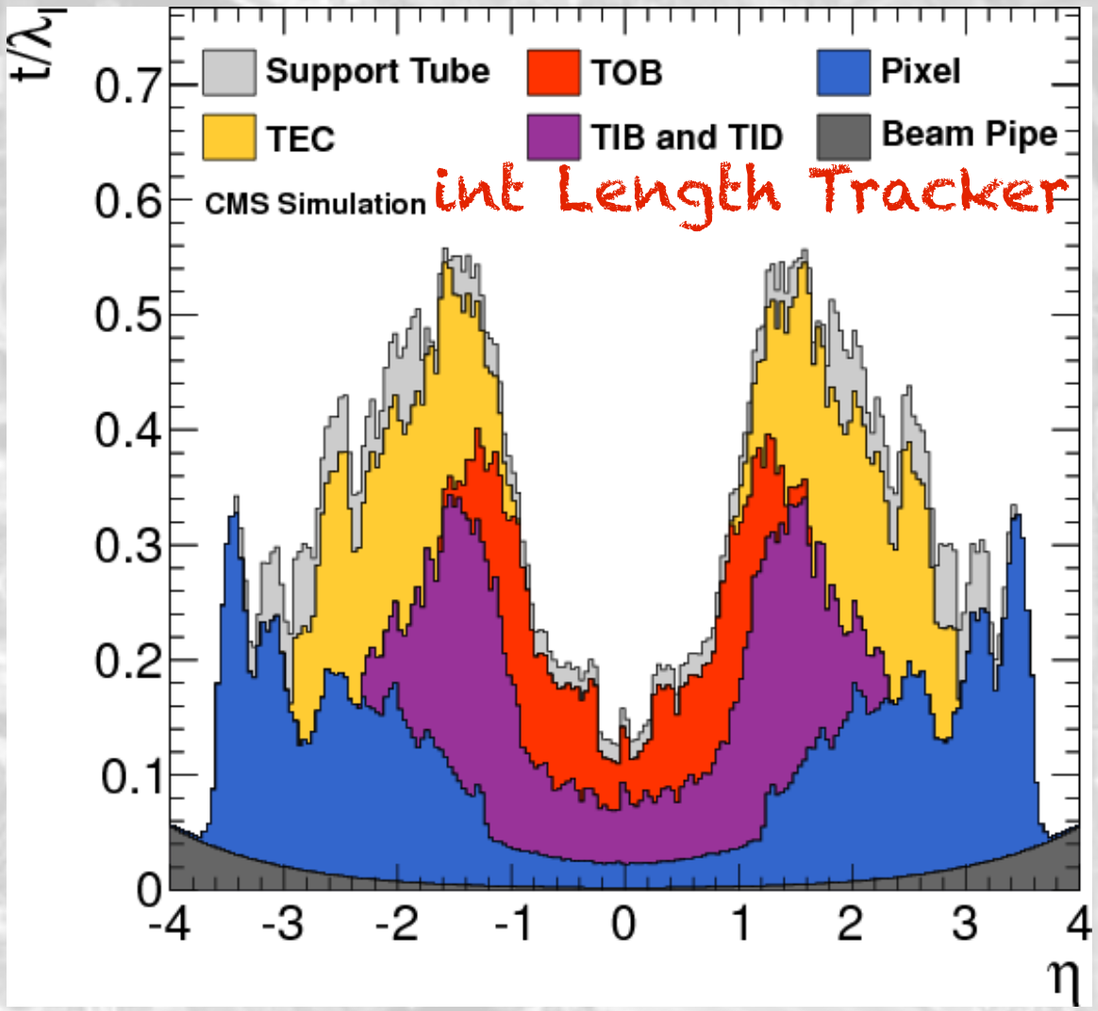
- We did not find anything striking so far, but we are not done yet...

- And we have data @13 TeV next year: unexplored territory ahead and (hopefully) big surprises...

Backup

Material Budget

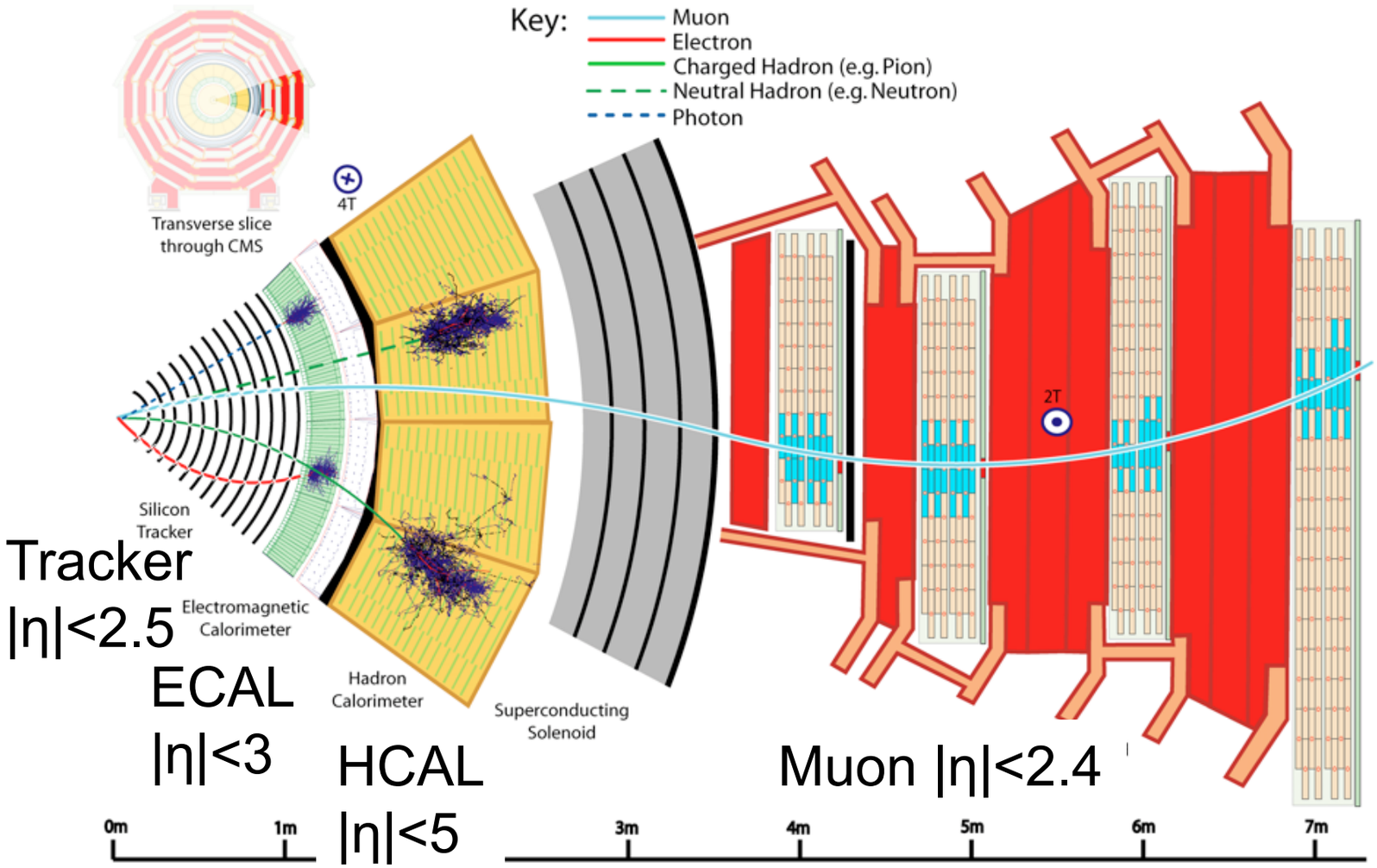


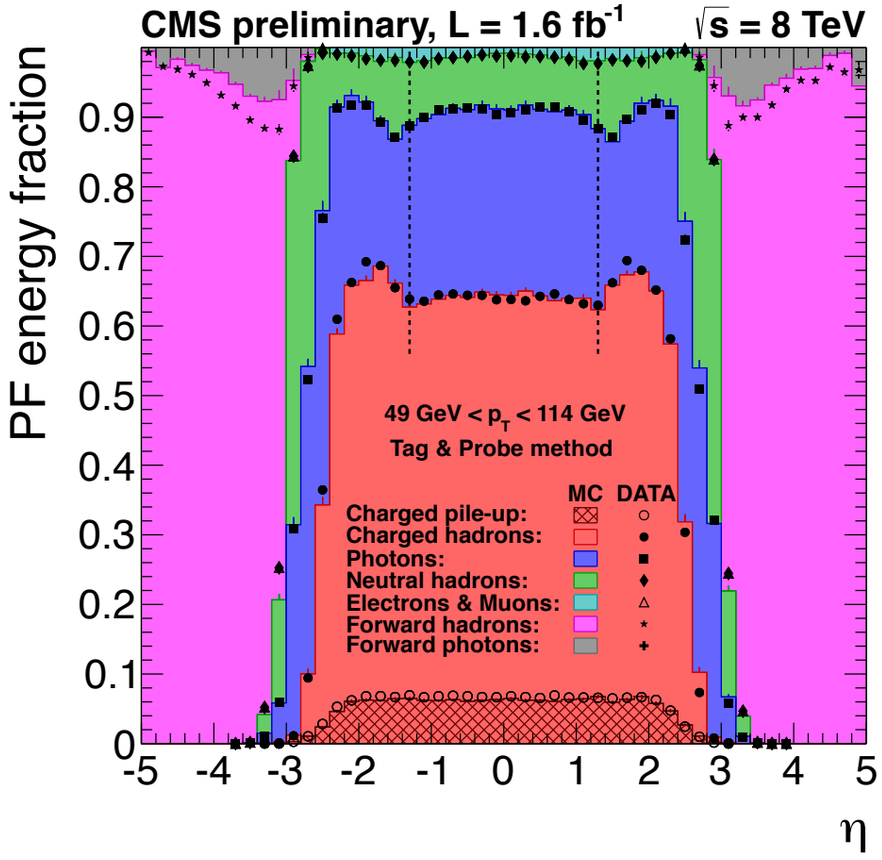


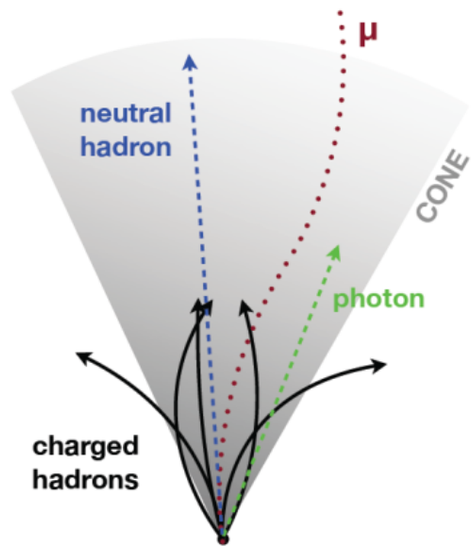
HCAL: At eta=0 there is less than 6 interaction lengths of HCAL. ECAL provides an additional interaction length

ECAL ~25 radiation length (23 cm crystals, 0.9 cm rad length)

Jet reconstruction in CMS





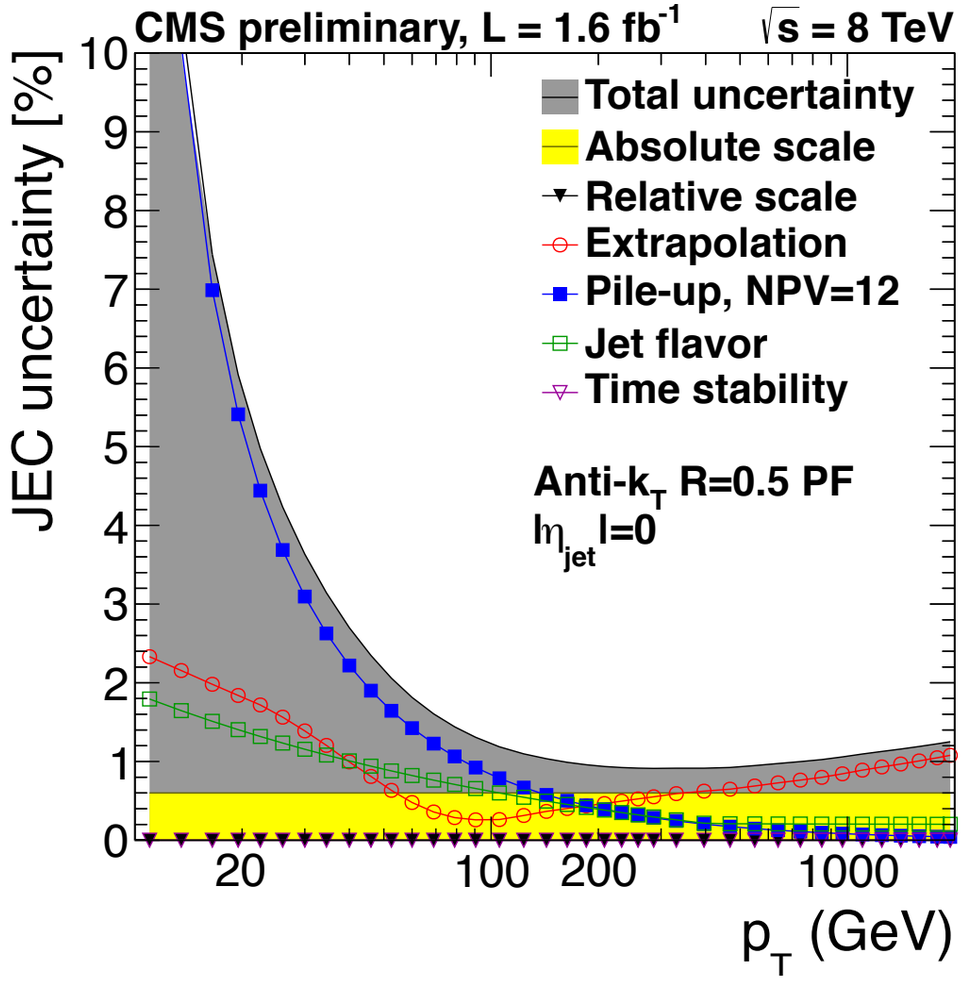


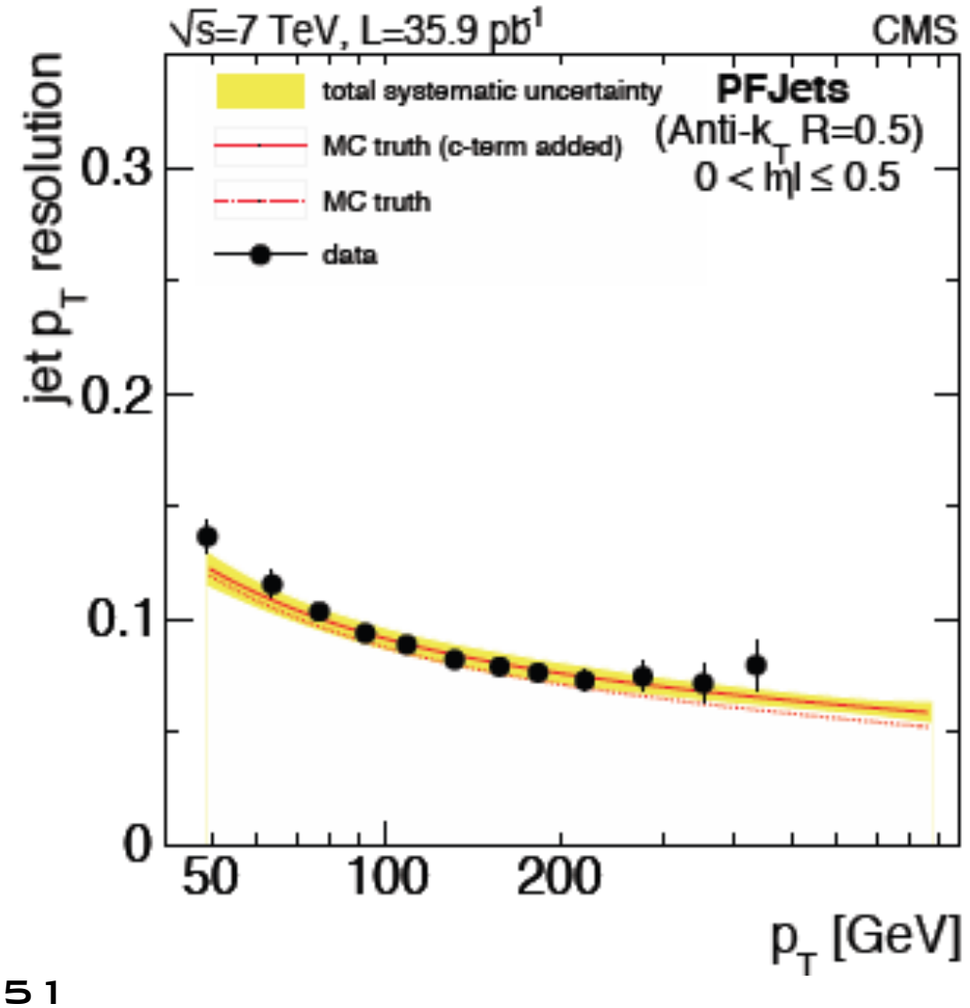
- **Particle Flow Jets** clustered from identified particles reconstructed using all detector components

- Default jet clustering algorithms for p+p collisions: **Anti-K_T** with R=0.5 (and 0.7)

Jet performance in CMS

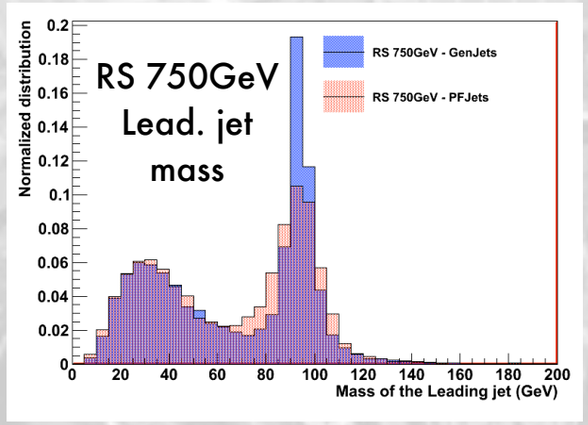
- Jet energy scale uncertainty: 1% for p_T > 150 GeV
- Jet calibration vs. η better than 1% per unit of pseudorapidity for |η|<2.5
- Jet energy resolution: 10% @ p_T = 100 GeV
- Jet position resolution in φ and η: ~0.01 @ p_T = 100 GeV
- Jet trigger efficiency: >99.9% above p_T threshold
- Event-by-event correction for pileup based on jet area [PLB659(2008)119]

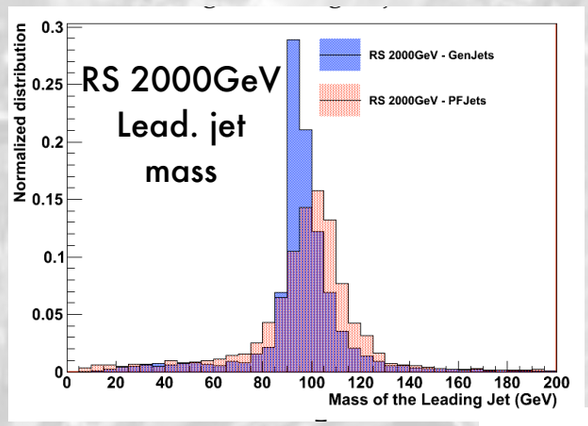


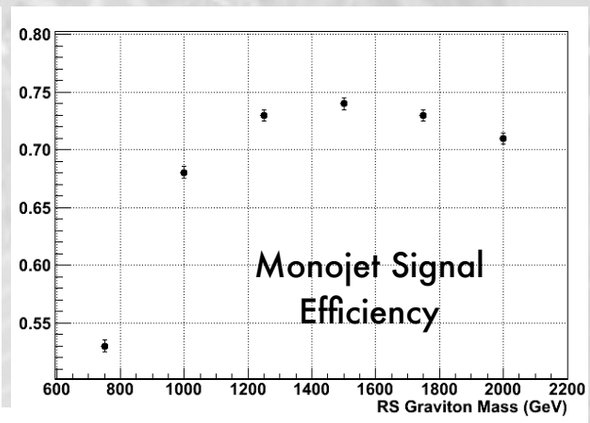


Hadronic V with Fat Jets

looking for one massive jet is the minimal requirement. High efficient for $\sqrt{s}$>1000 GeV







It allows to suppress the background from QCD by several orders of magnitude. More effective if on both jets

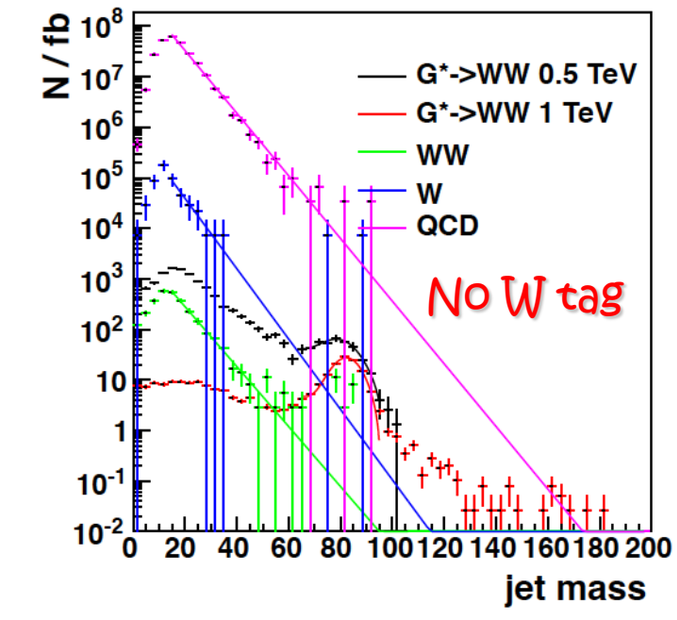
One can do better digging in the jet substructure

- Start with the objects (e.g. towers) forming a jet found with a recombination algorithm
- Rerun the algorithm, but at each recombination test whether:
  - $z < z_{cut}$ and $\Delta R_{ij} > D_{cut}$ (θ_J is angle at final recombination in original found jet)

CA: z_{cut} = 0.1 and $D_{cut} = \theta_J/2$

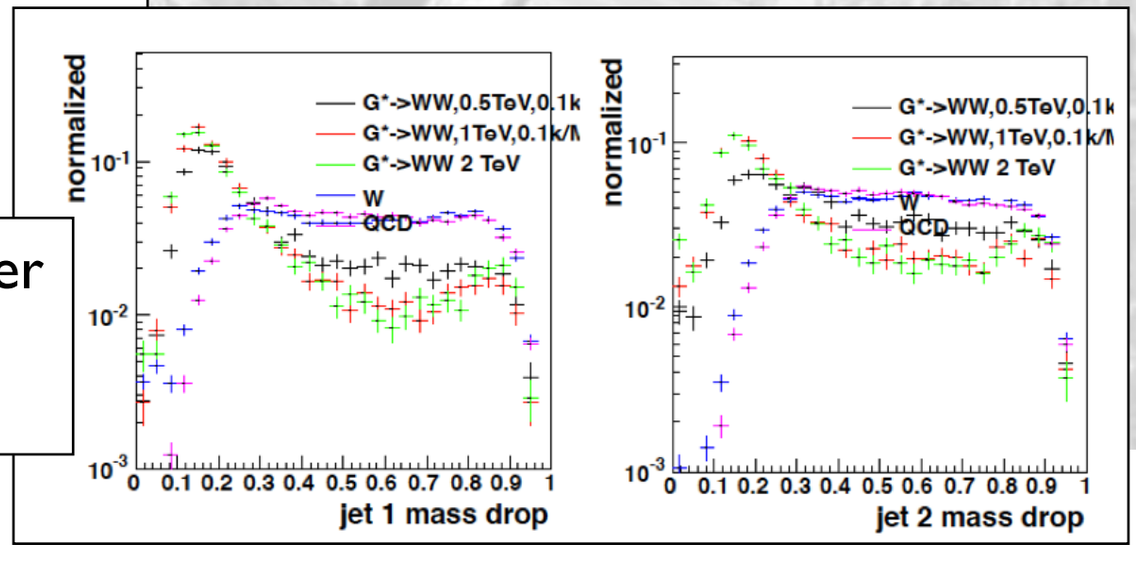
kT: z_{cut} = 0.15 and $D_{cut} = \theta_J/2$

- If true (a soft, large angle recombination), prune the softer branch by NOT doing the recombination and discarding the softer branch
- Proceed with the algorithm



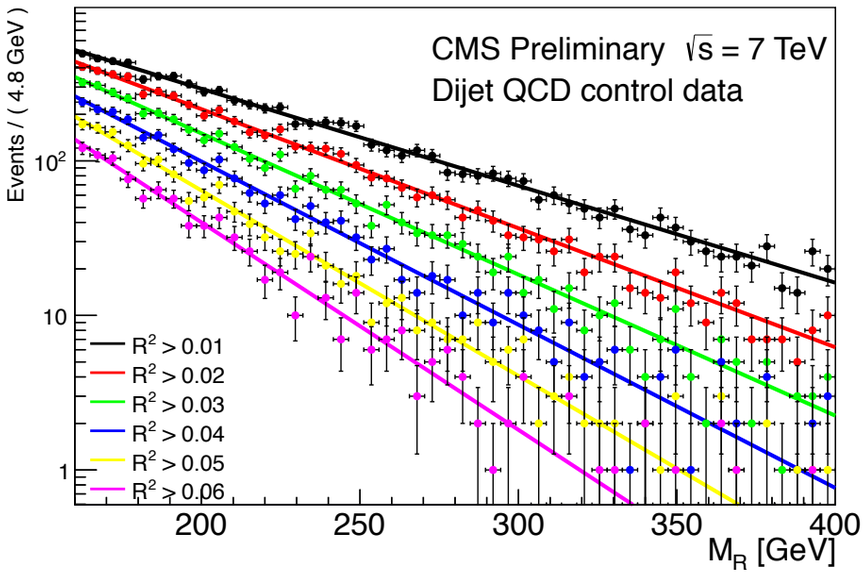
Use Mass Drop as further discriminating variable

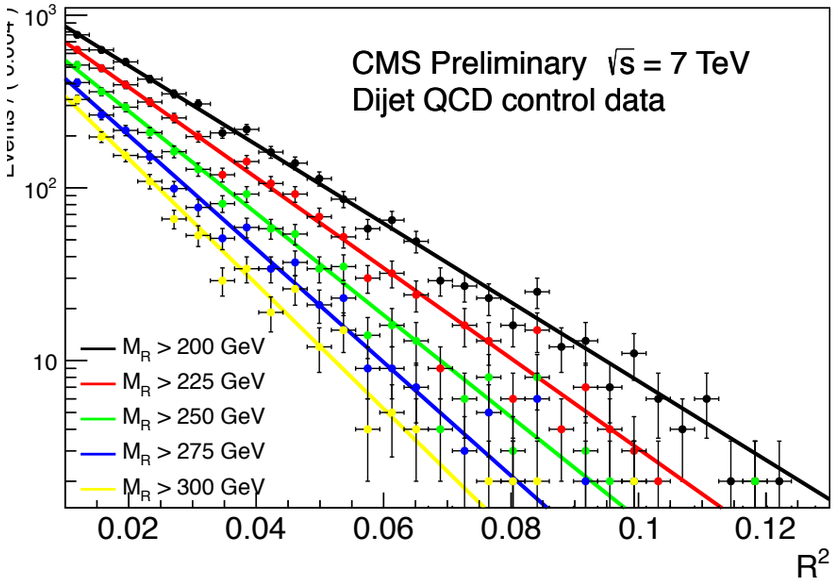
max(mj1+mj2)/mjet

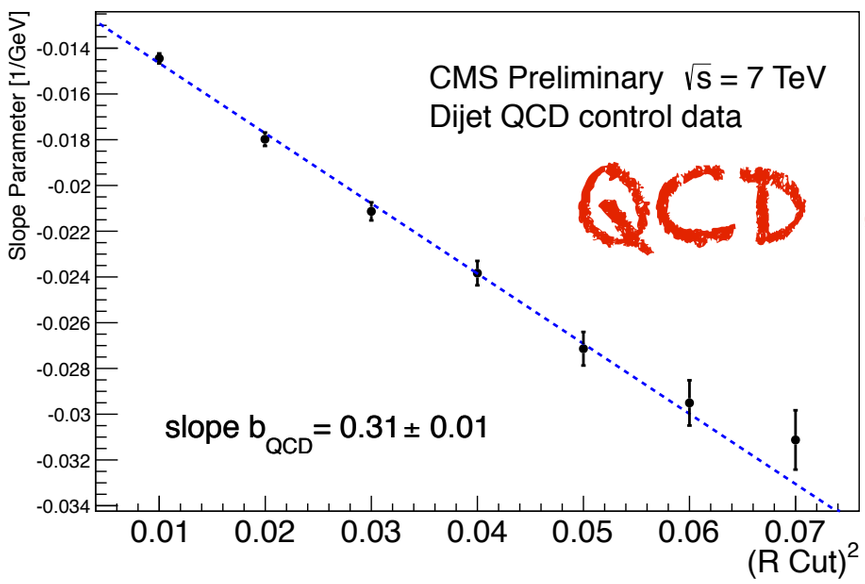


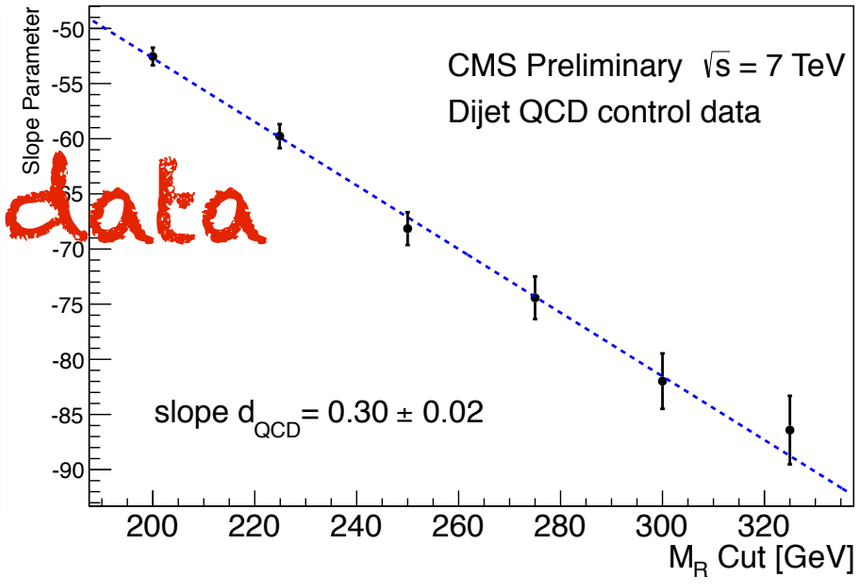
1D Background Model

$$f(M_R) \sim e^{-k M_R} \quad k = a + b R^2_{cut}$$



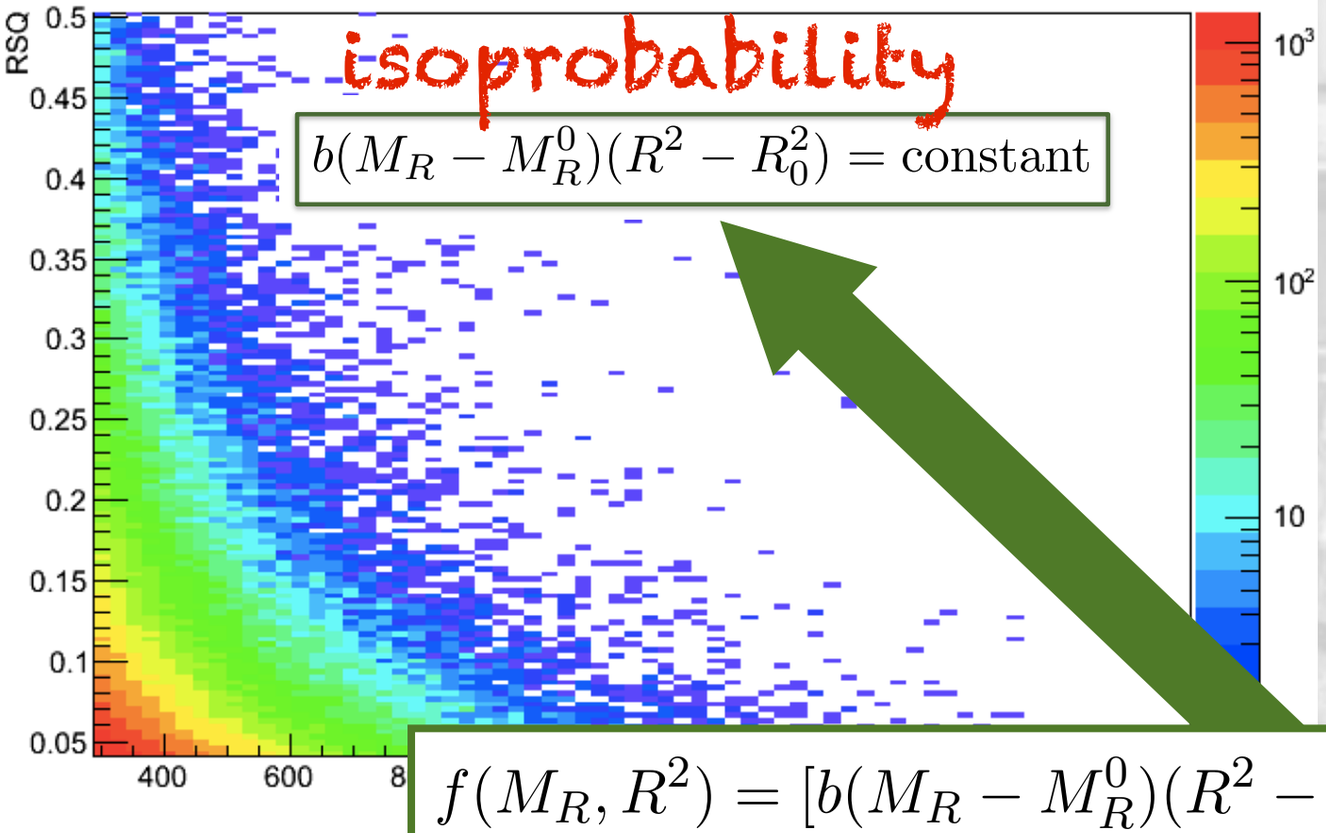






$$f(R^2) \sim e^{-k R^2} \quad k = c + b M_R^{cut}$$

From 1D to 2D

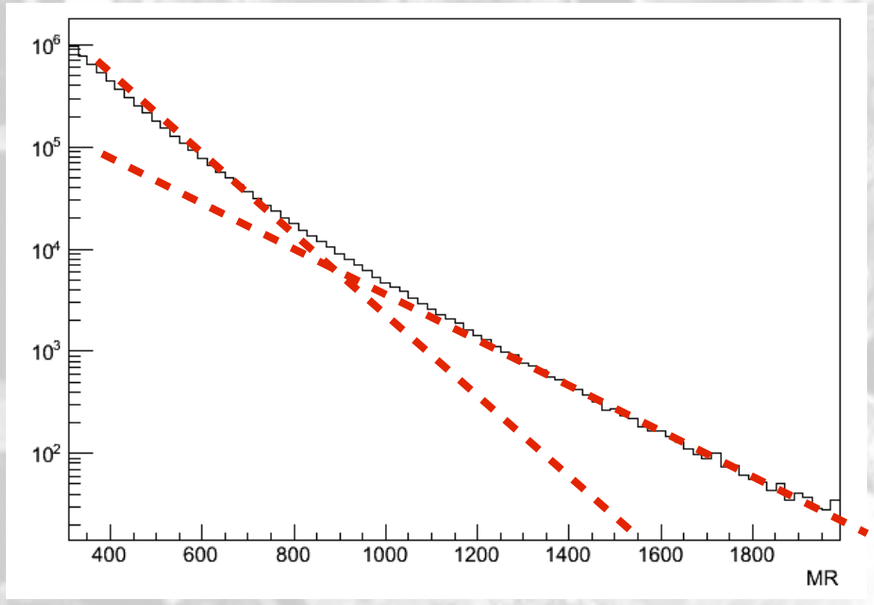


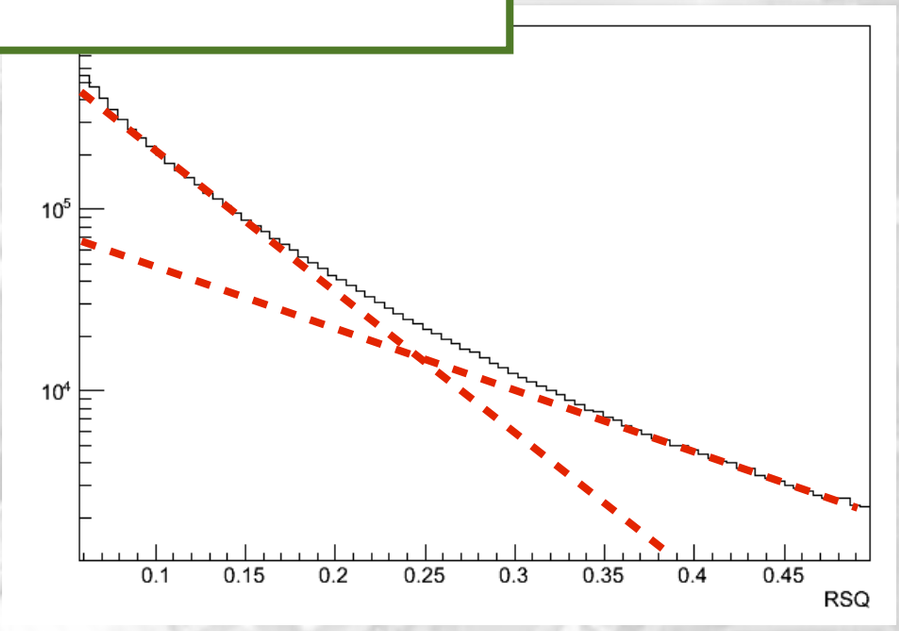
isoprobability

$$b(M_R - M_R^0)(R^2 - R_0^2) = \text{constant}$$

Each Bkg components (Z+jets, W+jets, tt +jets) well described by the sum of two of these pdfs

$$f(M_R, R^2) = [b(M_R - M_R^0)(R^2 - R_0^2) - 1]e^{-b(M_R - M_R^0)(R^2 - R_0^2)}$$





$$\int_{R^2_{cut}}^{+\infty} f(M_R, R^2) dR^2 \sim e^{-(a+bR^2_{cut})M_R}$$

$$\int_{M_R^{cut}}^{+\infty} f(M_R, R^2) dM_R \sim e^{-(c+bM_R^{cut})R^2}$$

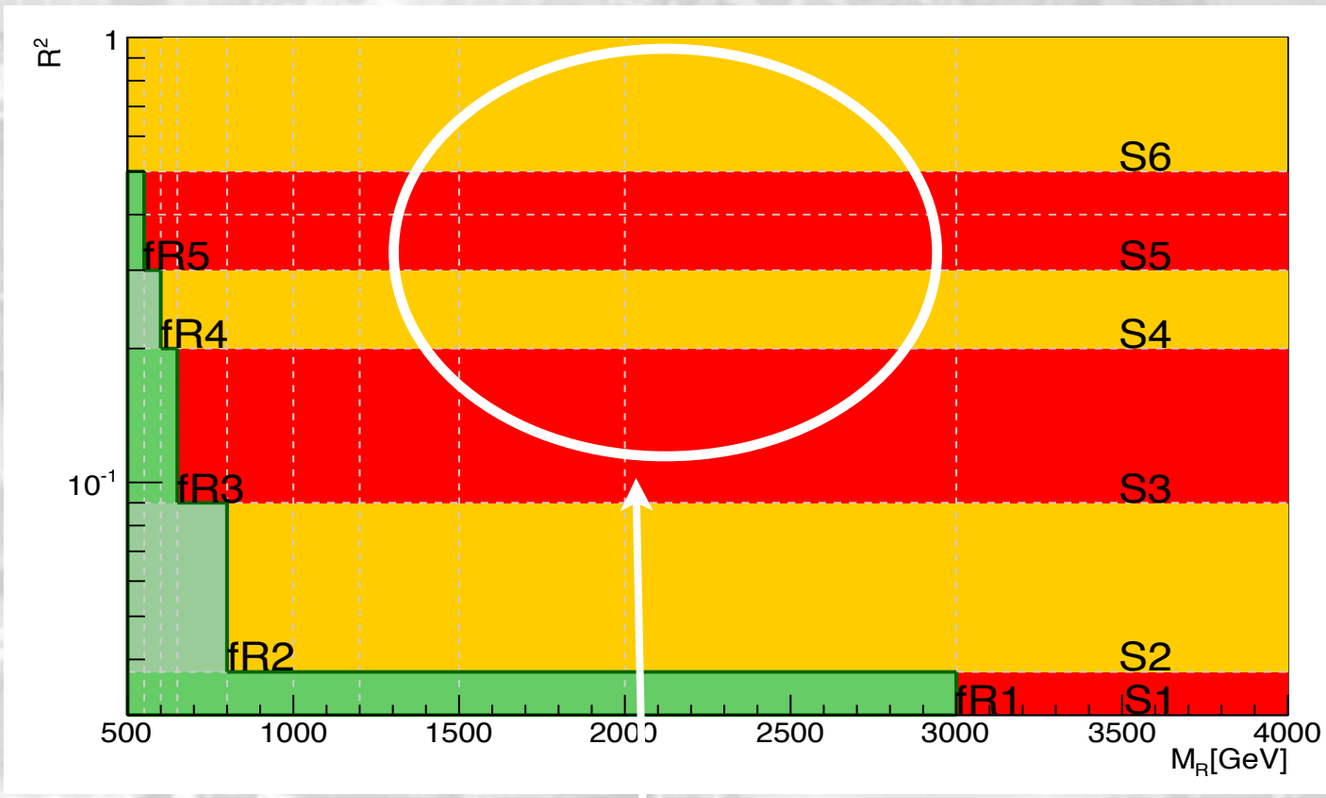
The Background Model

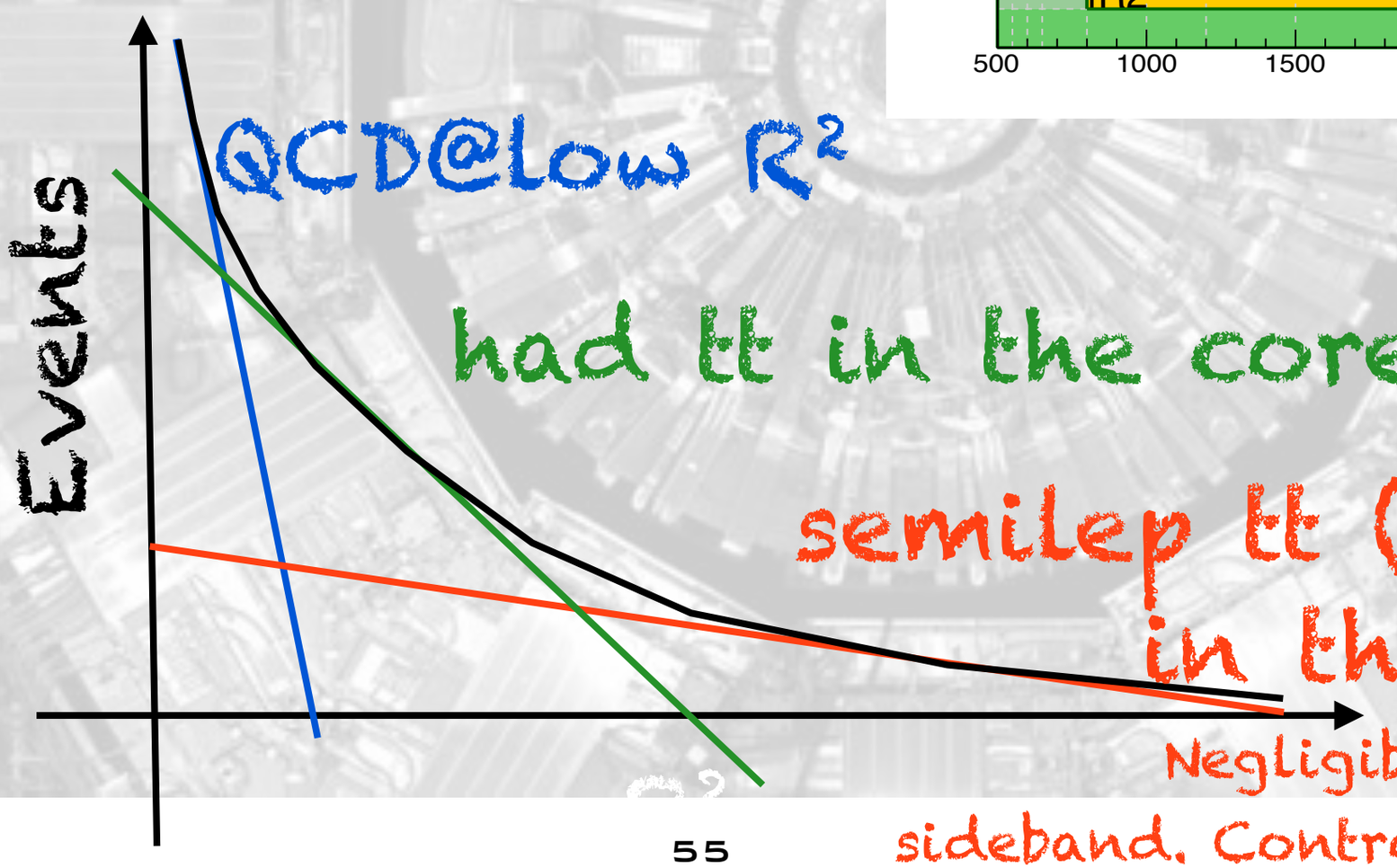
Determine the shape with a ML fit

Fit performed in the sideband

Shape extrapolated outside

Bkg prediction (integral) compared to data to look for an excess





The Event Selection

Events collected with set of multijet triggers (4j,5j, 6j)

At least 6 jets with pT>30 GeV

MR>500 GeV & R^2>0.03

Three subsamples are considered, depending on lepton multiplicity and btag

QCD control sample

no loose btag & no lepton

1 medium btag & no lepton

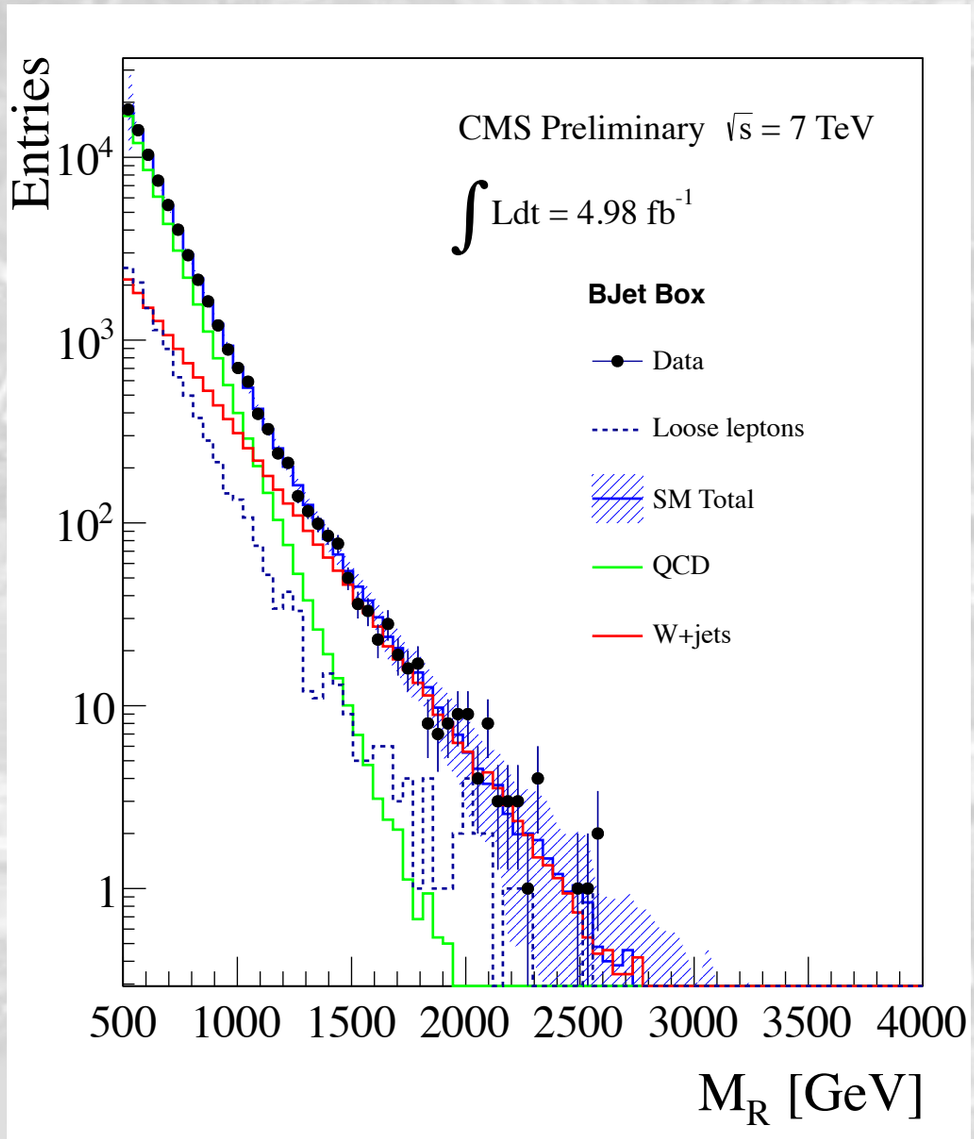
Search sample

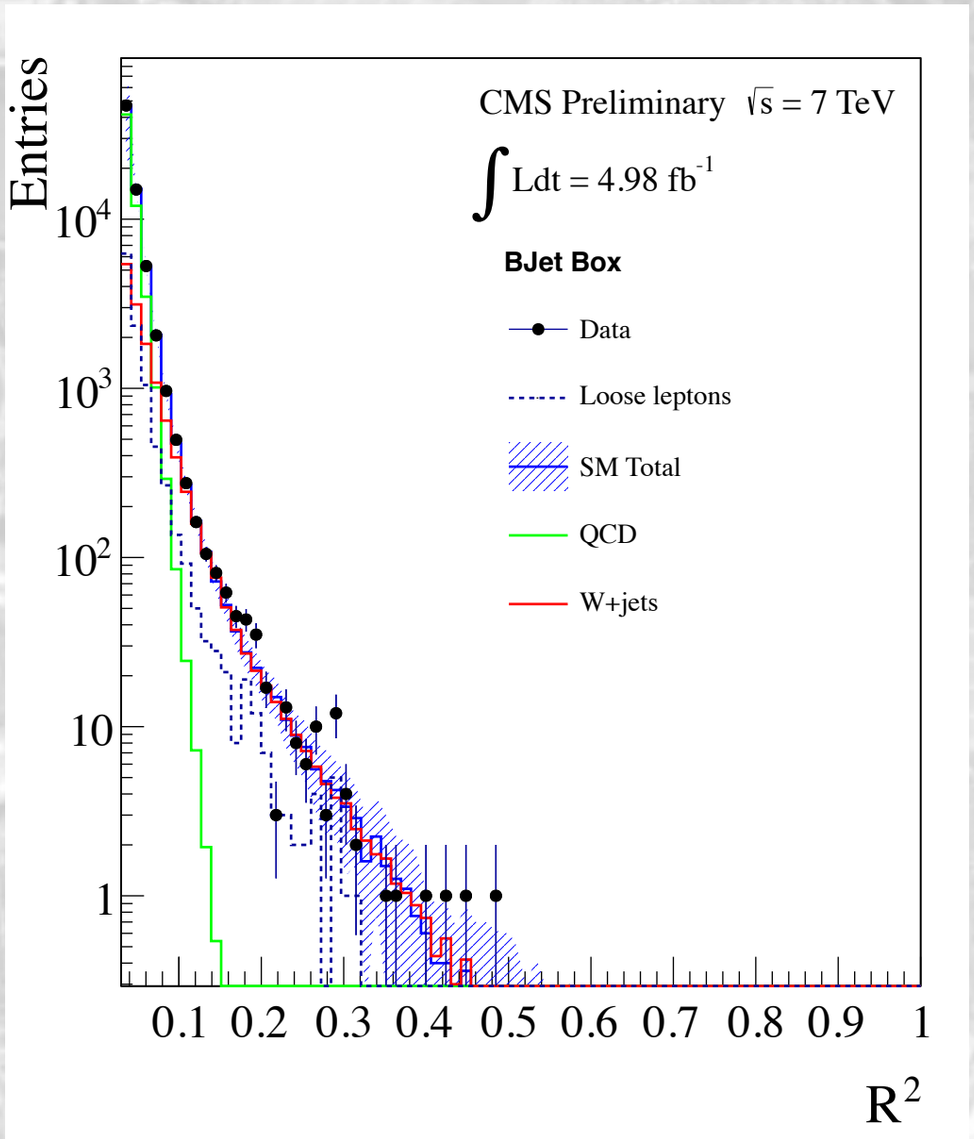
"Lost lepton" control sample

no loose btag & one lepton

Control samples to study dedicated BKGs

The Background Model





No significant excess seen
Result translated in a limit