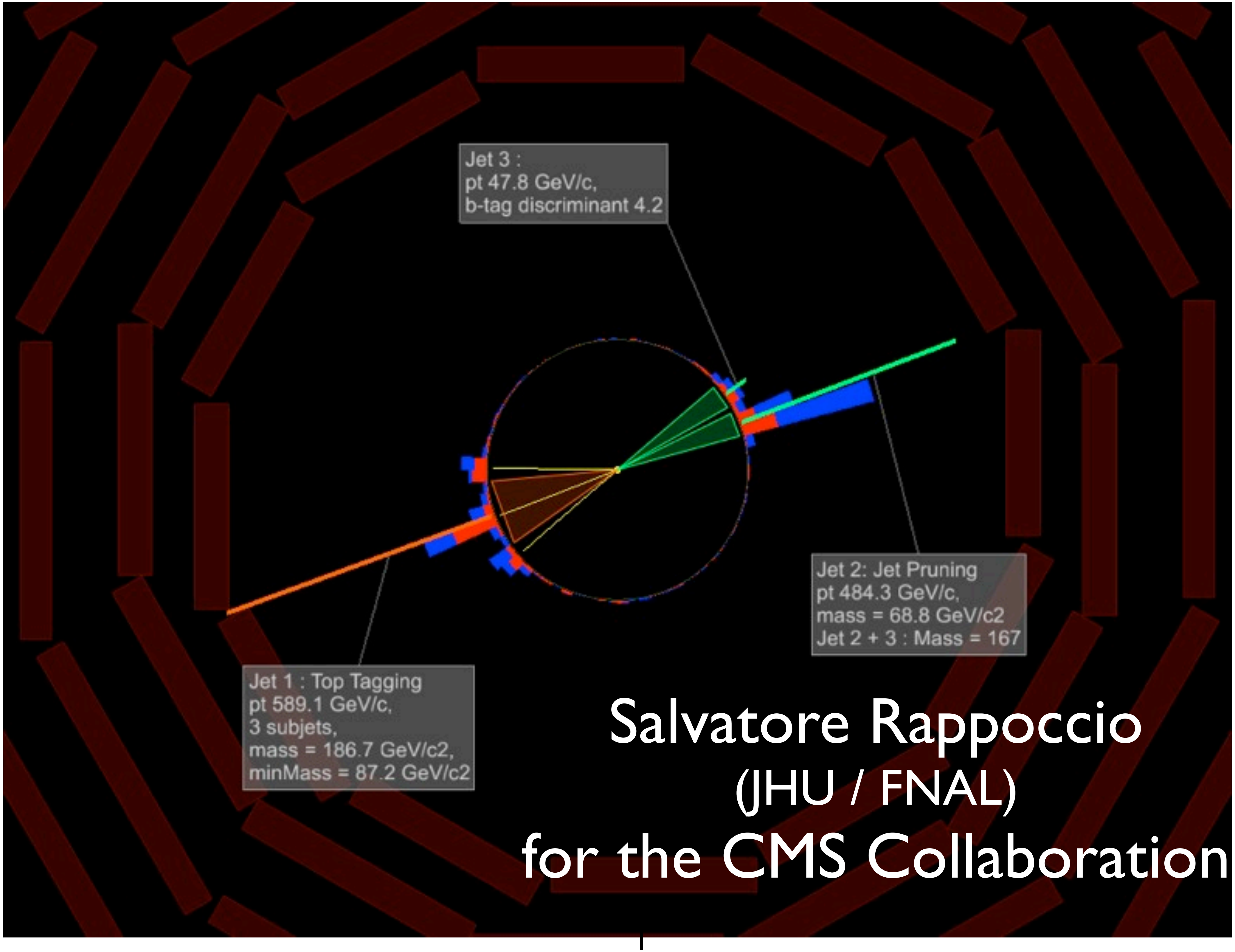
CMS studies and application of jet substructure



Salvatore Rappoccio
(JHU / FNAL)
for the CMS Collaboration

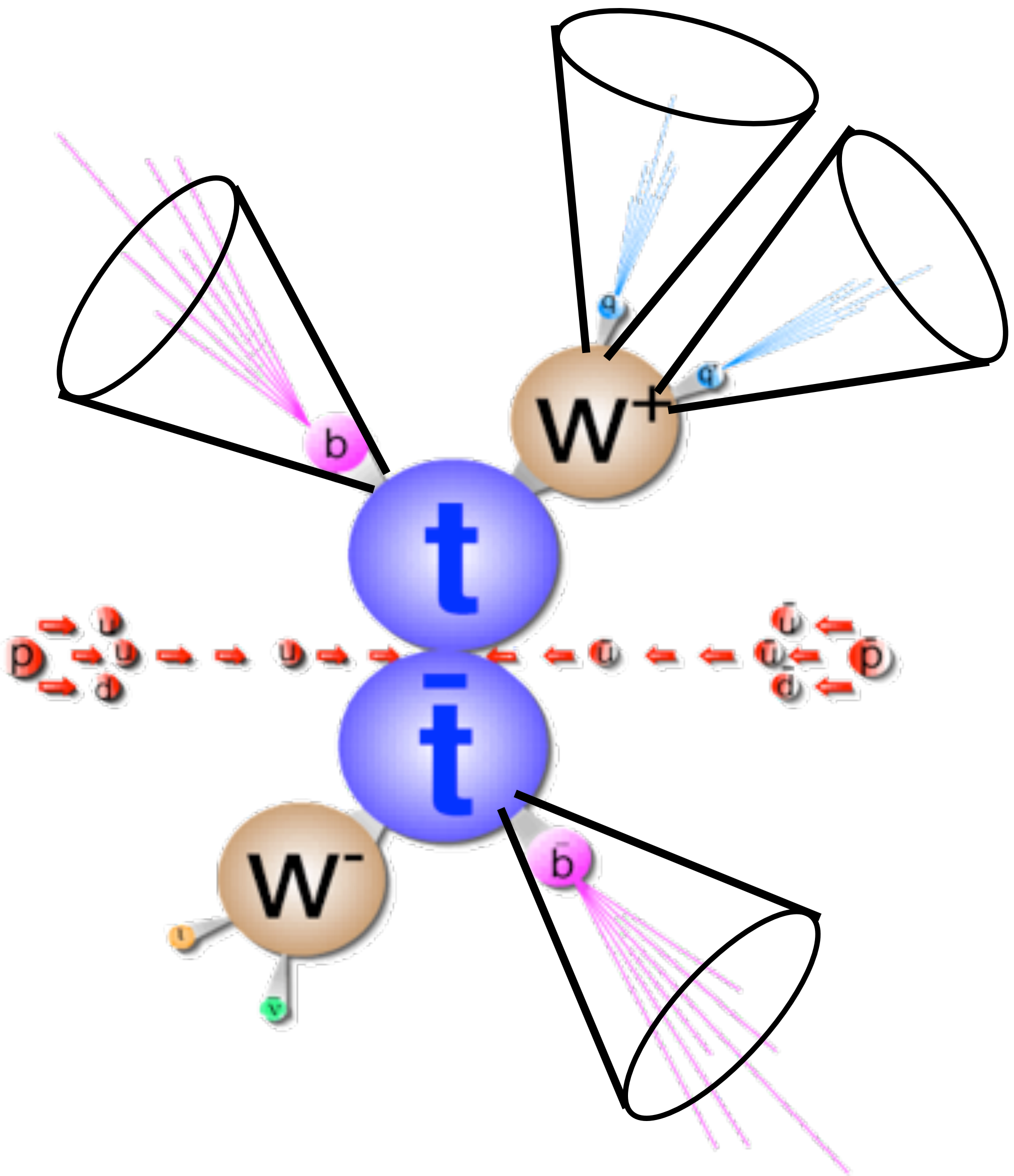
Documentation

- Jet substructure studies technical report
  - JME-10-013
- Boosted ttbar search
  - EXO-11-006, http://arxiv.org/abs/1204.2488, submitted to JHEP
- Boosted V + MET search
  - EXO-11-061
- Boosted V + ll search
  - Hot off the presses! EXO-11-081

Joint publication in preparation

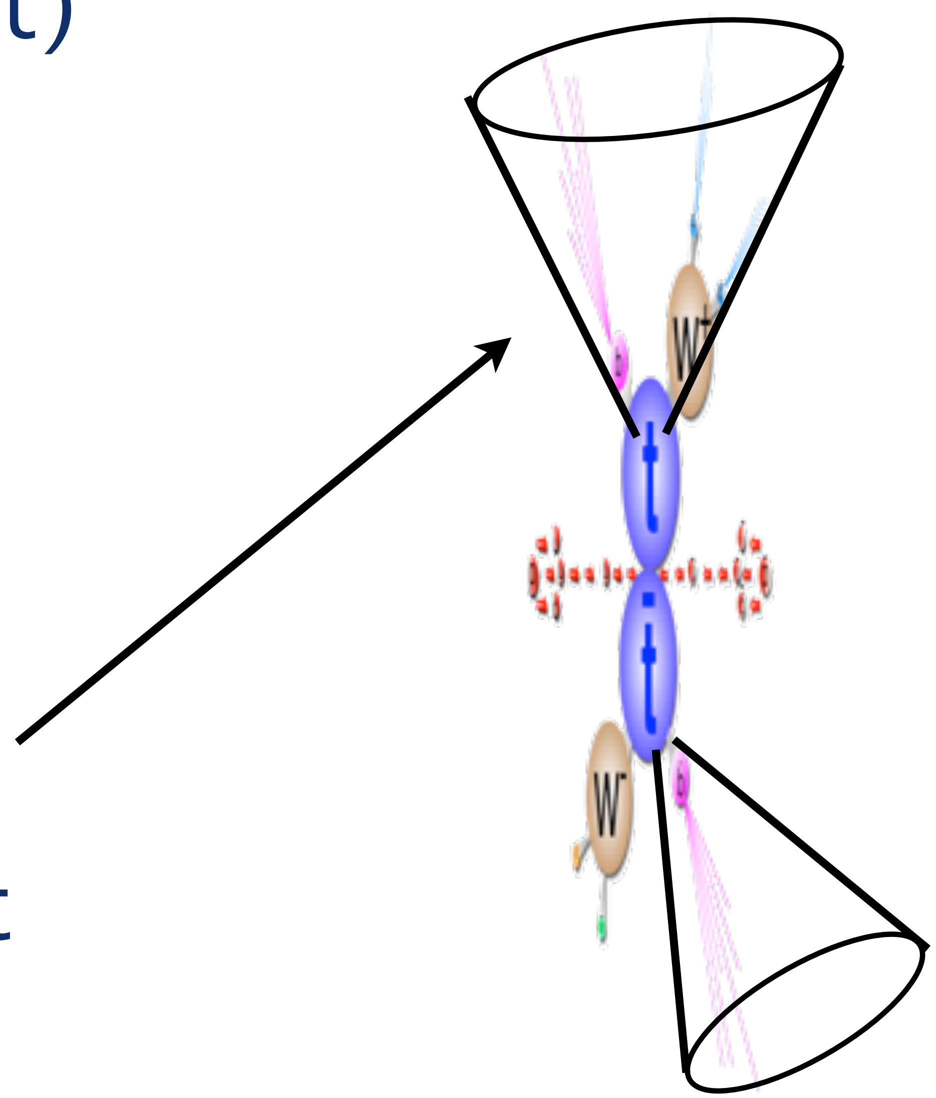
- H->bb
  - HIG-11-031, http://arxiv.org/abs/1202.4195, submitted to PLB
  - Cross-check analysis uses boosted topology but no substructure tools yet
- Dijet analysis
  - EXO-11-015, http://arxiv.org/abs/1107.4771, Phys. Lett. B 704 (2011) 123
  - “Fat jet” approach motivated by boosted techniques

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on traditional methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



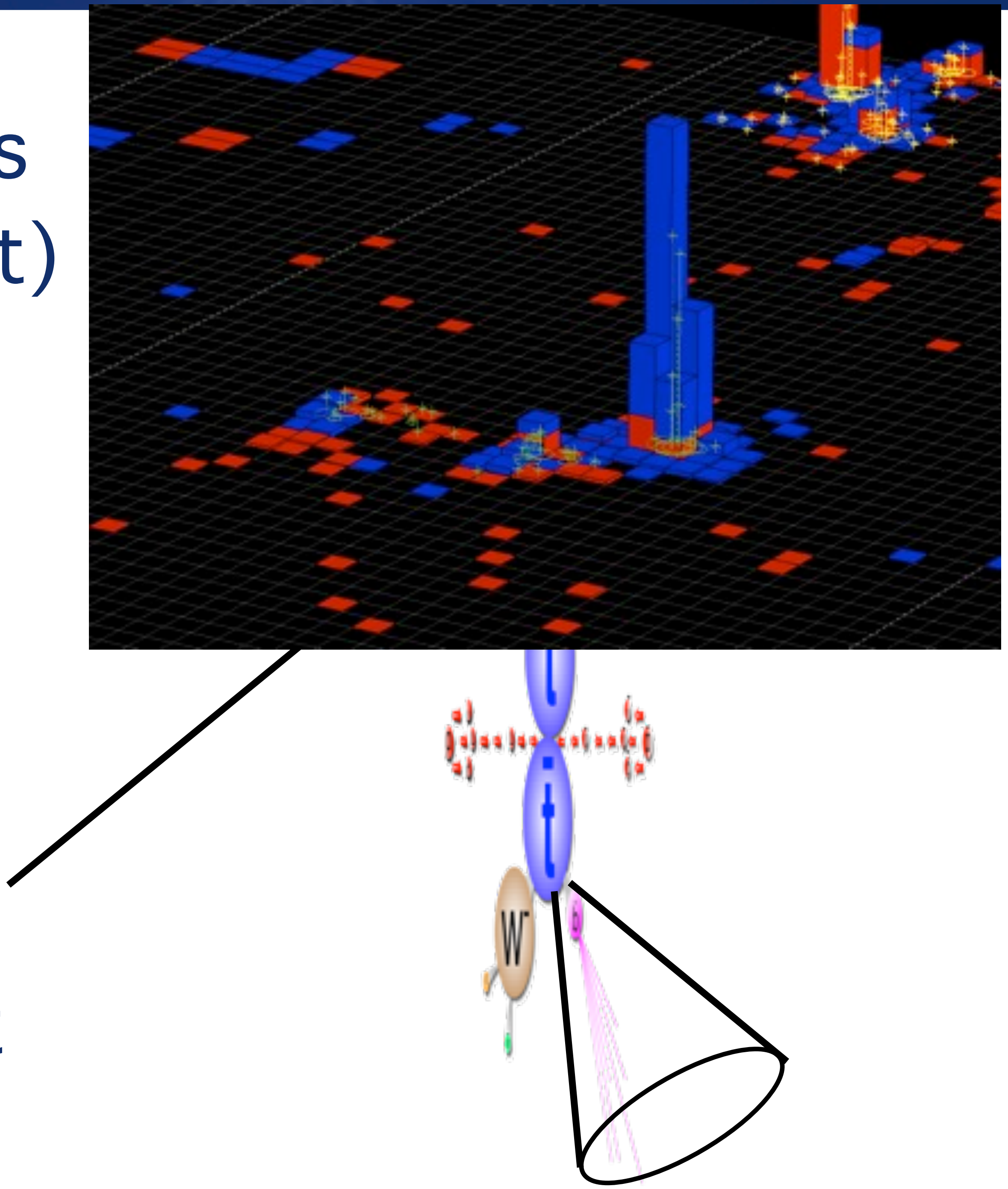
Motivation

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on traditional methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



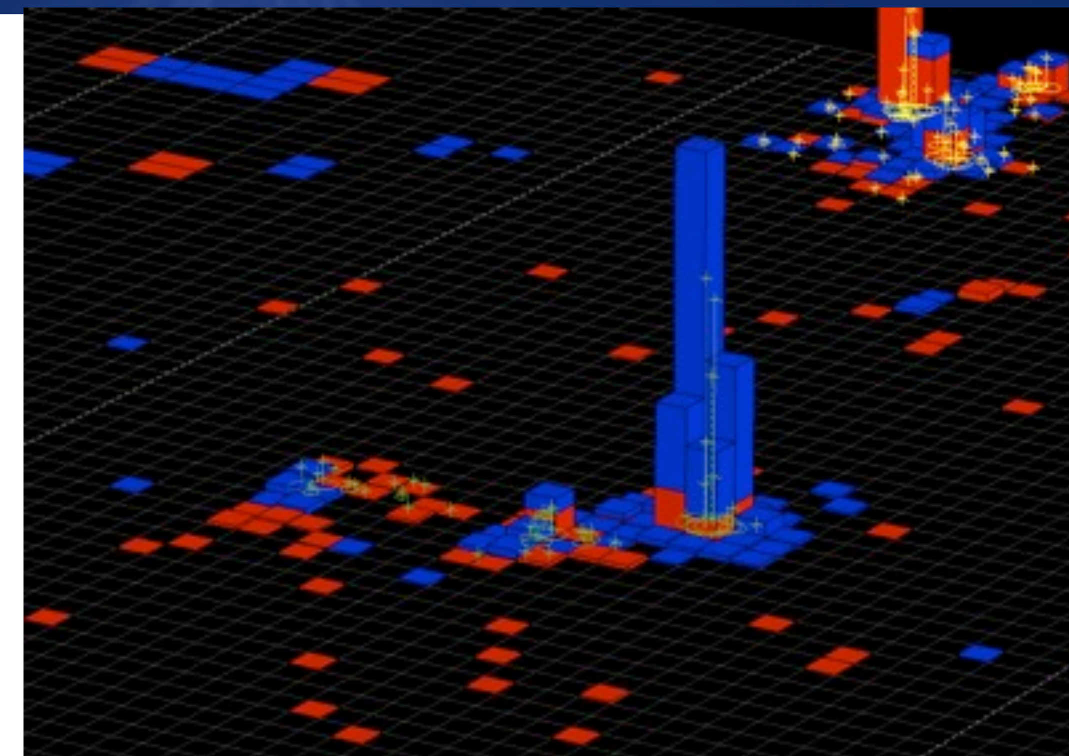
Motivation

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on traditional methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



Motivation

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on traditional methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



"These ain't your daddy's jets!"
(Joey Huston)

Or if you prefer…

Boosted jet,
massive jet,
fat jet,
hulk jet

Jet

Or if you prefer...

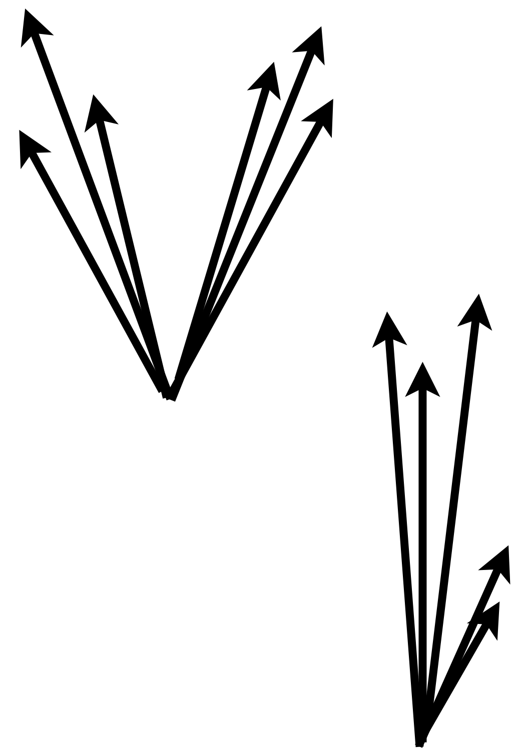
Boosted jet,
massive jet,
fat jet,
hulk jet

(OK, I made
the last one up)

Jet

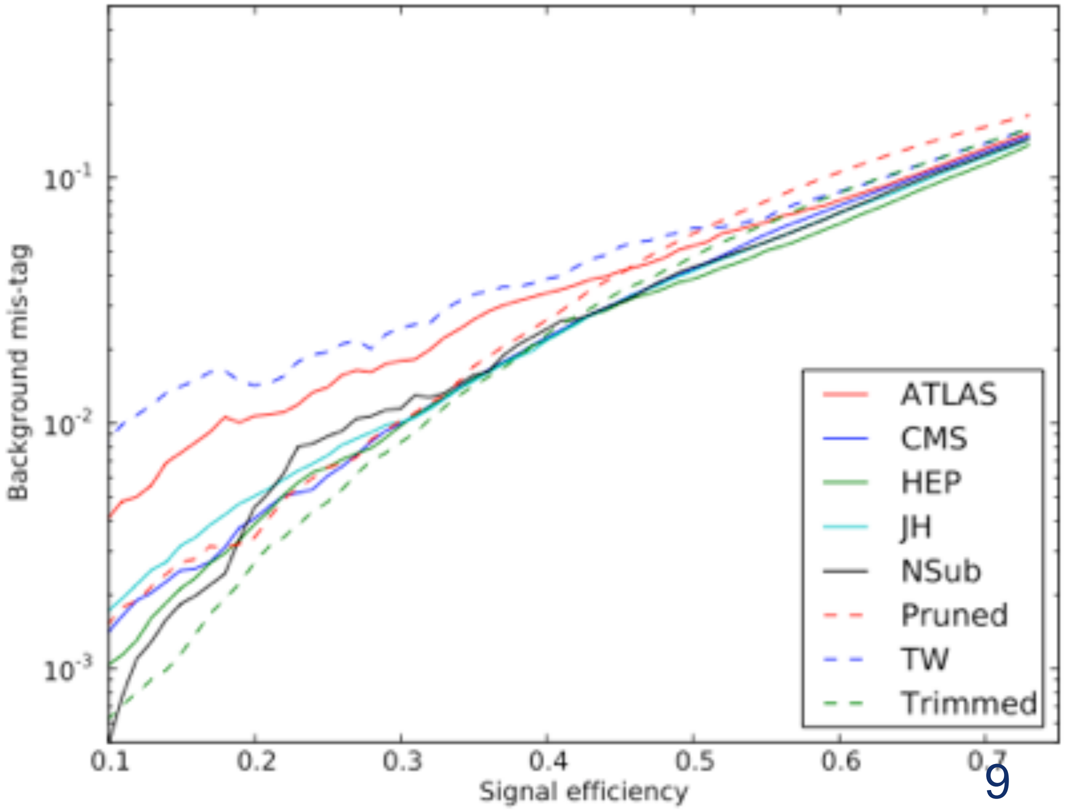
Jet Substructure

- Massive particles:
  - Massive 4-vector sum of daughters
  - 1-2 wide-angle splittings
  - Symmetric splittings



- QCD :
  - Low-mass 4-vector sum of daughters
  - Many low-angle splittings
  - Asymmetric splittings

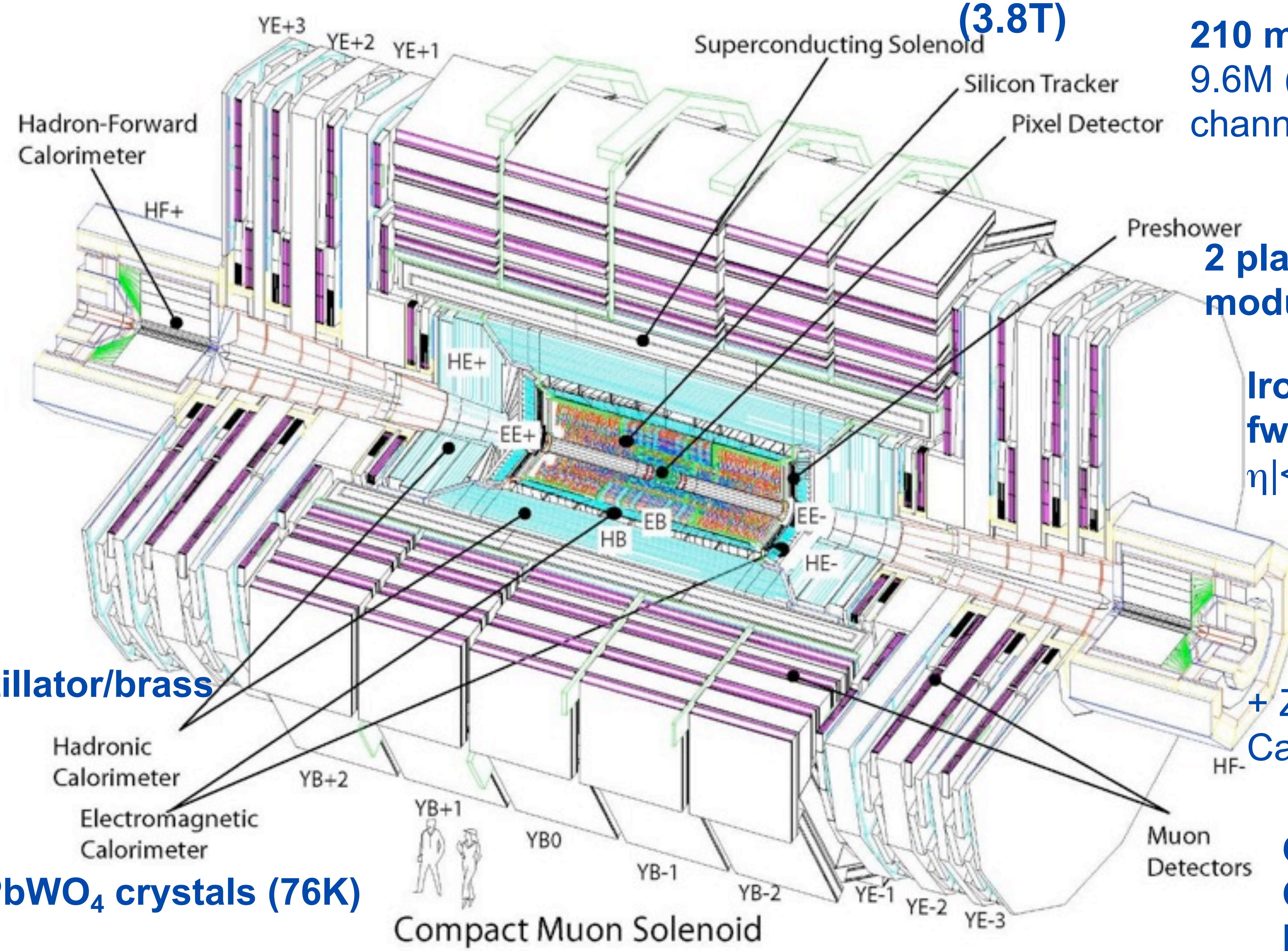
- Recent explosion of tools to exploit this!
  - Jet filtering : Butterworth, Davison, Rubin, Salam
  - JHU Top tagger: Kaplan, Reherman, Tweedie, Schwartz
  - Jet pruning : Ellis, Vermillion, Walsh
  - Jet trimming: Thaler, Wang
  - N-subjettiness : Thaler, Van Tilburg
  - HEP top tagger : Plehn, Salam, Spannowski
  - Many, many more!
  - For an overview check out the BOOST proceedings :
    - arXiv:1012.5412v2 [hep-ph]
    - arXiv:1201.0008v1 [hep-ph]



Outline

- Motivation
- Experimental and algorithmic overview
- Applications

Experimental Overview



(3.8T)

210 m² of silicon sensors: 9.6M (Str) & 66M (Pix) channels

2 planes of silicon modules for ECAL

Iron / Quartz fiber fwd calorimeter, $3<|\eta|<5$;

+ Castor, $5<|\eta|<6.55$

+ Zero Degree Calorimeter

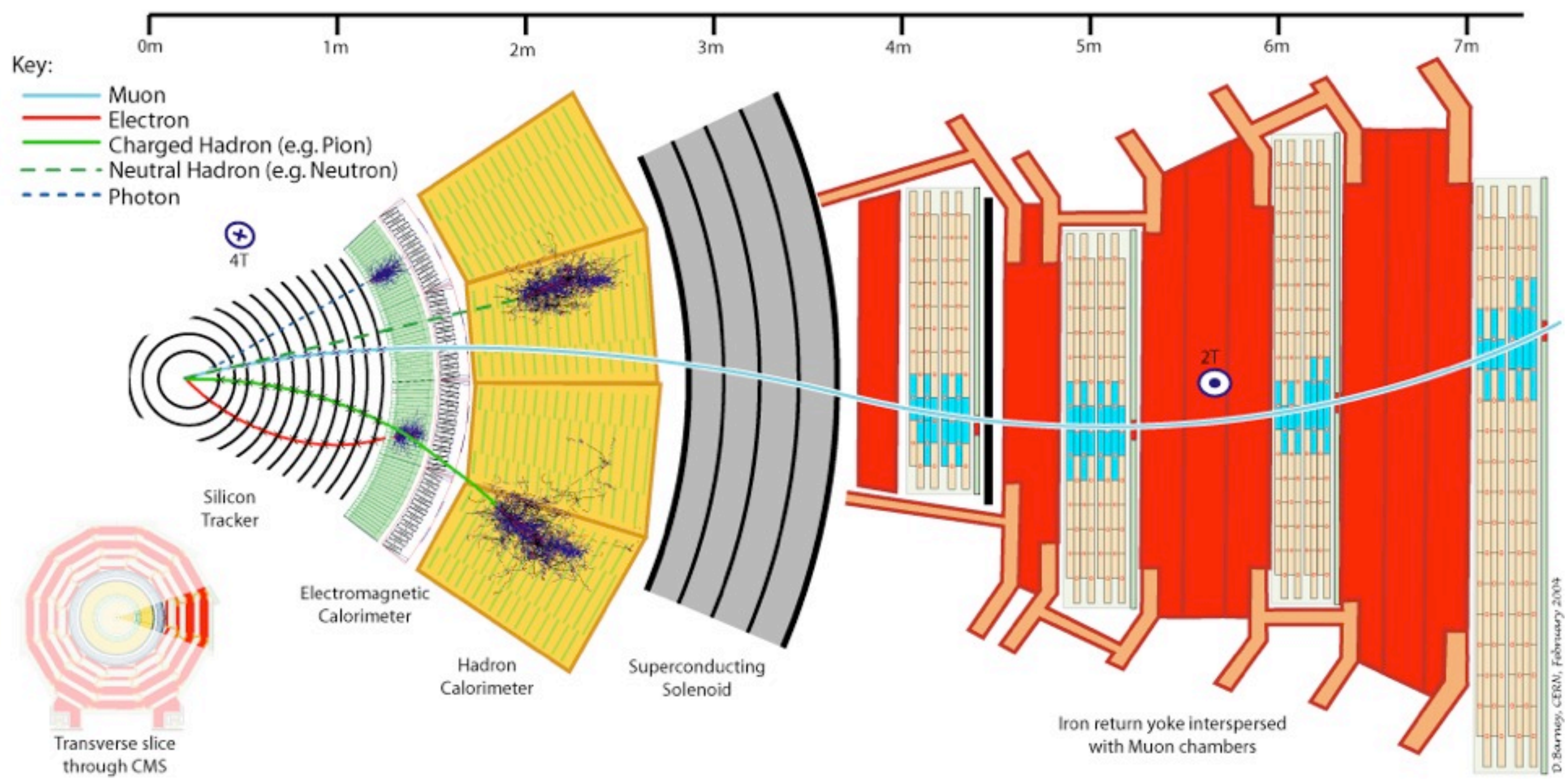
Scintillator/brass

$PbWO_4$ crystals (76K)

Cathode Strip Chambers, Drift Tubes, Resistive Plates

Experimental Overview

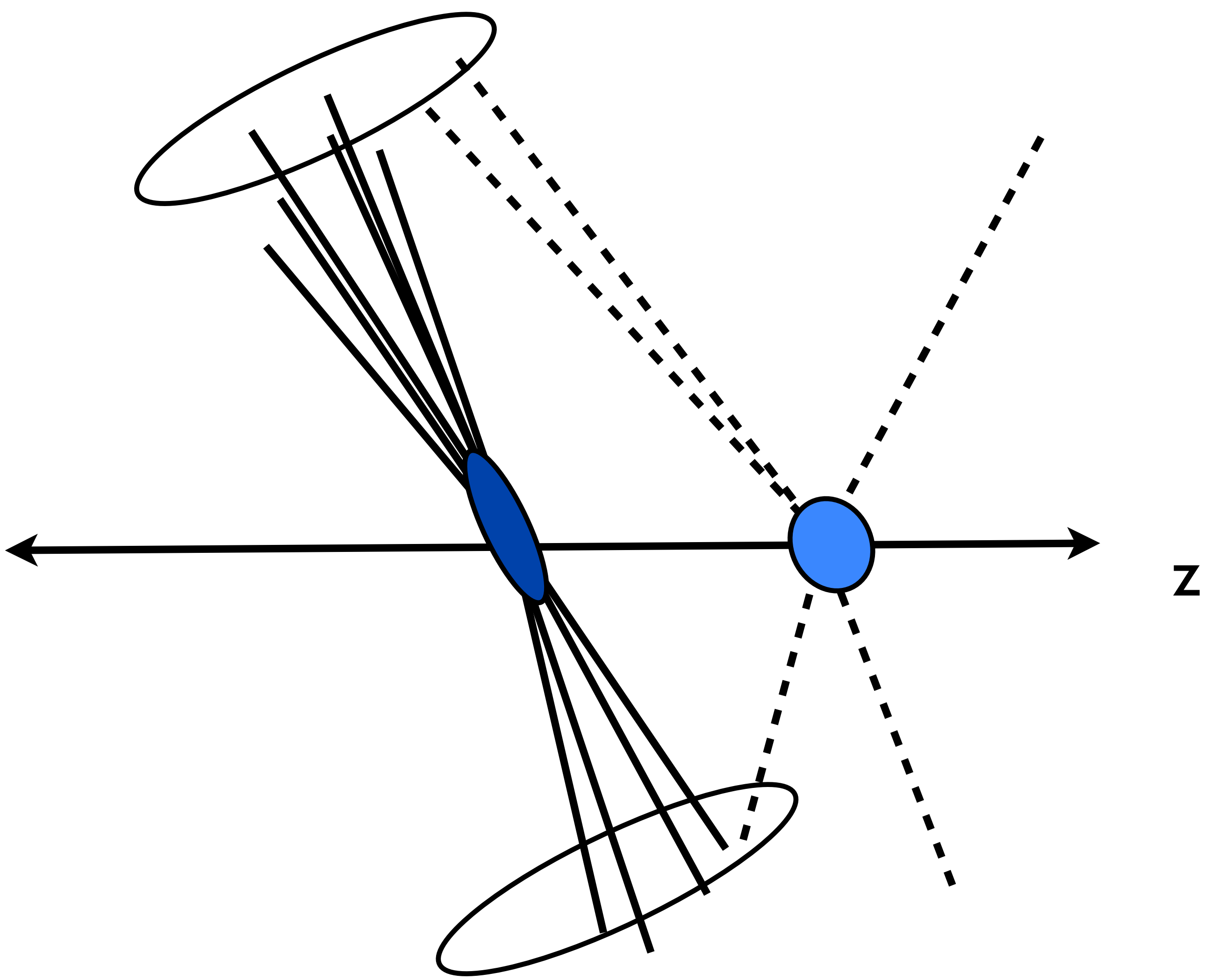
Classify objects into 5 categories



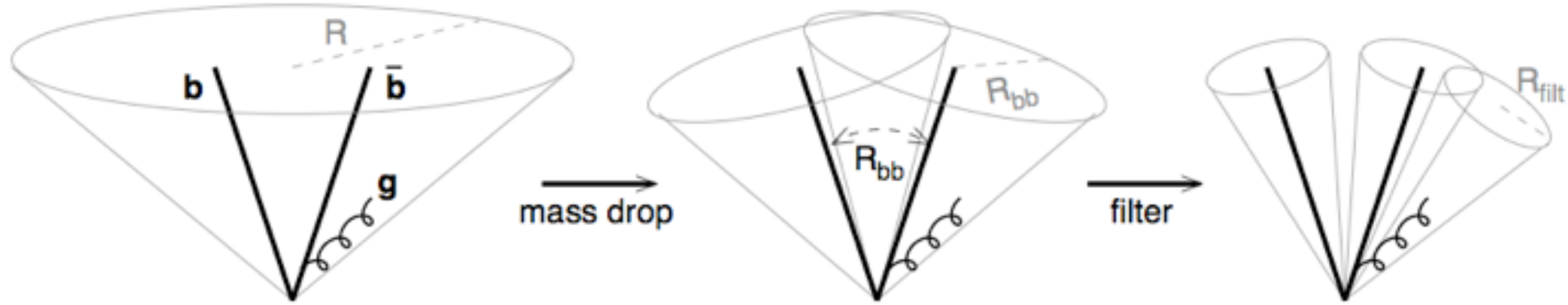
"Holistic" approach to reconstruction at CMS: Particle flow!

Experimental Overview

- Advantages of PF reconstruction for substructure:
  - Excellent subjet angular and energy resolution (<10% where interesting)
  - Subjet energy scales are very close to typical QCD energy scales
  - Can explicitly remove pileup from non-leading primary vertices



Jet Filtering Details



- From Butterworth, et al (arXiv:0802.2470 [hep-ph])
- Undo clustering sequence until a sufficiently symmetric mass drop is obtained

$$Y_{ij} = \frac{\min(p_{ti}^2, p_{ti}^2)}{p_t^2} \Delta R_{ij}^2 > Y_{cut} \qquad \mu = \frac{m_{j1}}{m_j}$$

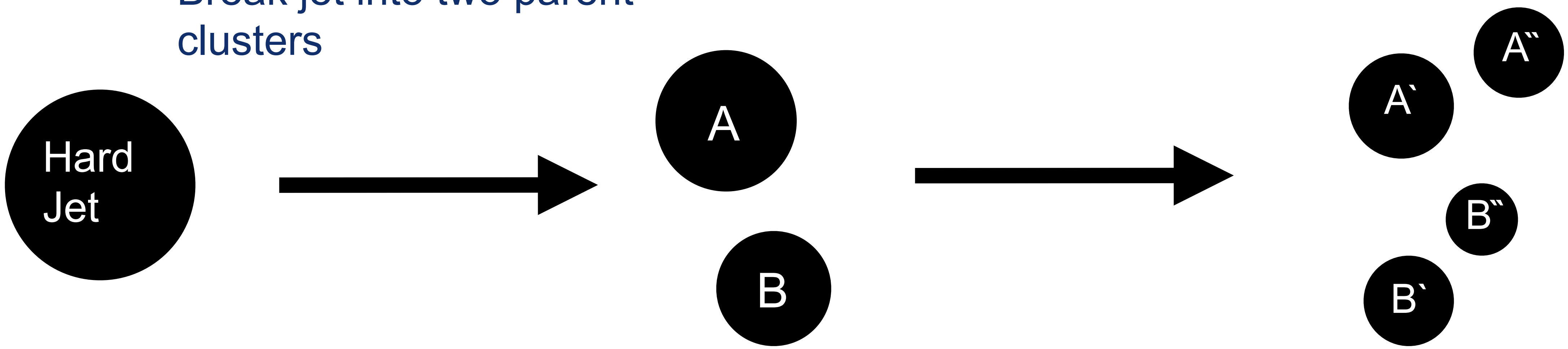
- Filter constituents by using a smaller distance parameter to remove "noise" components

Top Tagging Details

- Based on Kaplan et al. (arXiv:0806.0848)
- Cluster particle flow candidates using Cambridge Aachen
- Reverse the clustering sequence in order to find substructure
- Subjets must satisfy two requirements
  - Momentum fraction criterion: $p_{T}^{subjet} > 0.05 \times p_{T}^{hard\ jet}$ ← Removes soft subjets
  - Adjacency criterion: $\Delta R(A, B) > 0.4 - 0.0004 \times p_T$ ← Removes adjacent subjets
- Iterative process - throw out objects that fail and try to decluster again
- Stops when >=3 distinct, sufficiently hard subjets emerge

Primary Decomposition:
Break jet into two parent clusters

Secondary Decomposition
Repeat on parent clusters

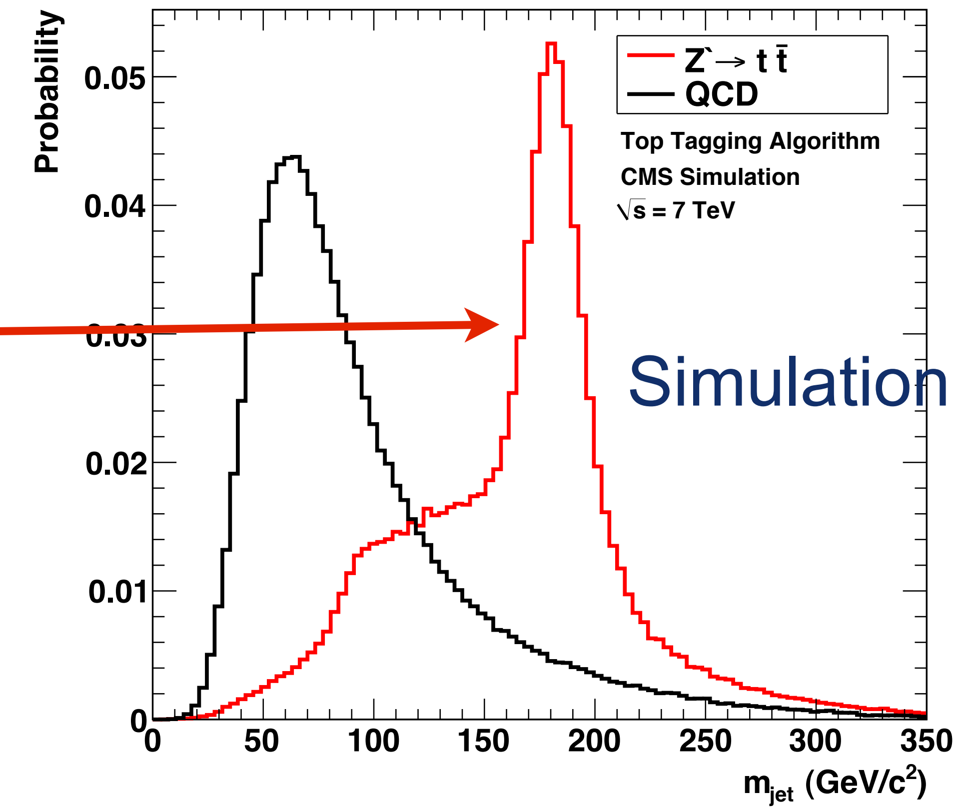


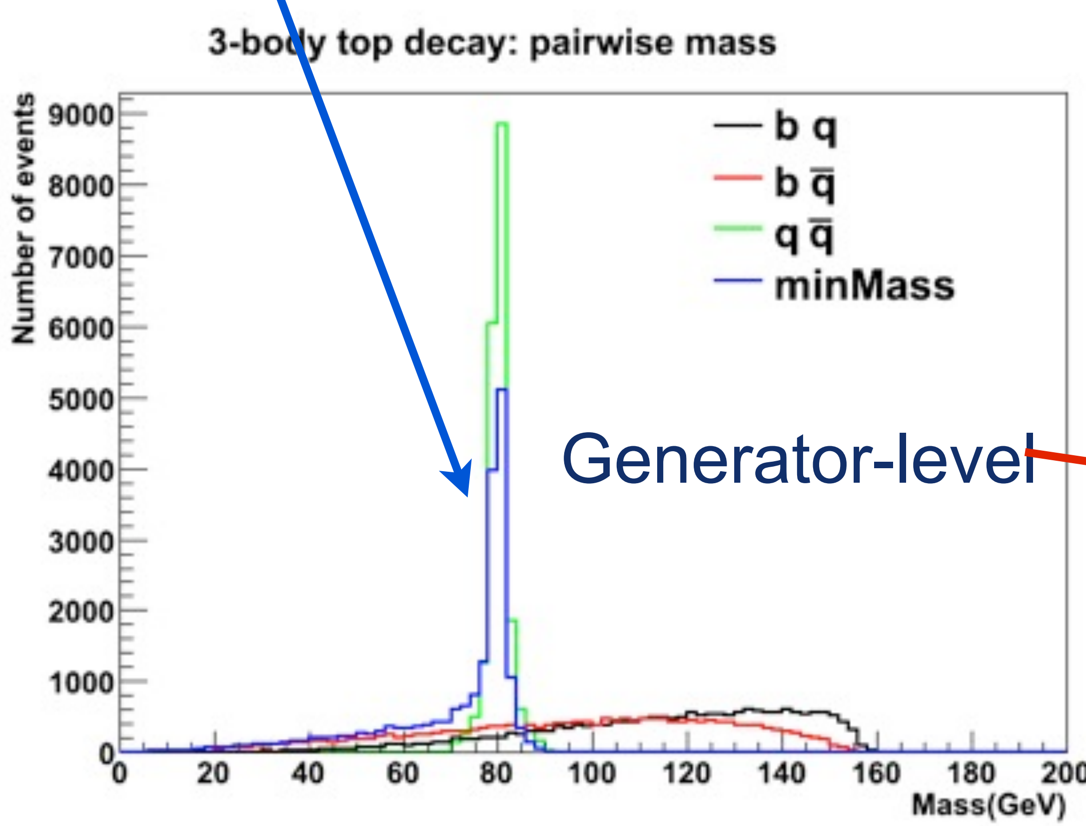
Require >= 3

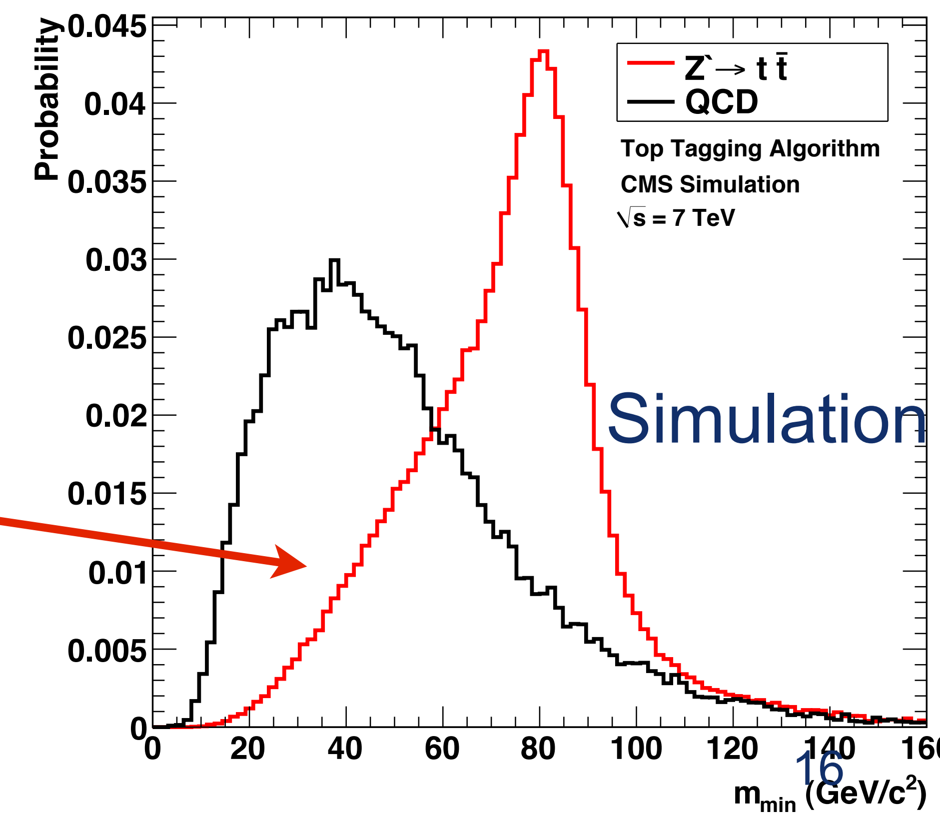
Top Tagging Details

- Discriminating variables:
  - Number of subjets: 3 or 4
  - Top Mass: Approximated by jet mass
    - Mass in 100-250 GeV/c^2
  - W Mass: Approximated by min pairwise mass
    - Min mass > 50 GeV/c^2

$$m_{\min} = \min[m_{12}, m_{13}, m_{23}]$$







Jet Pruning Details

- Ellis et al. (arXiv:0903.5081)
- Attempts to isolate subjet showers by removing soft, large angle particles from each subjet
- The "interesting" recombination occurs at wide angles
- Recluster each jet, requiring that each recombination satisfy the following:

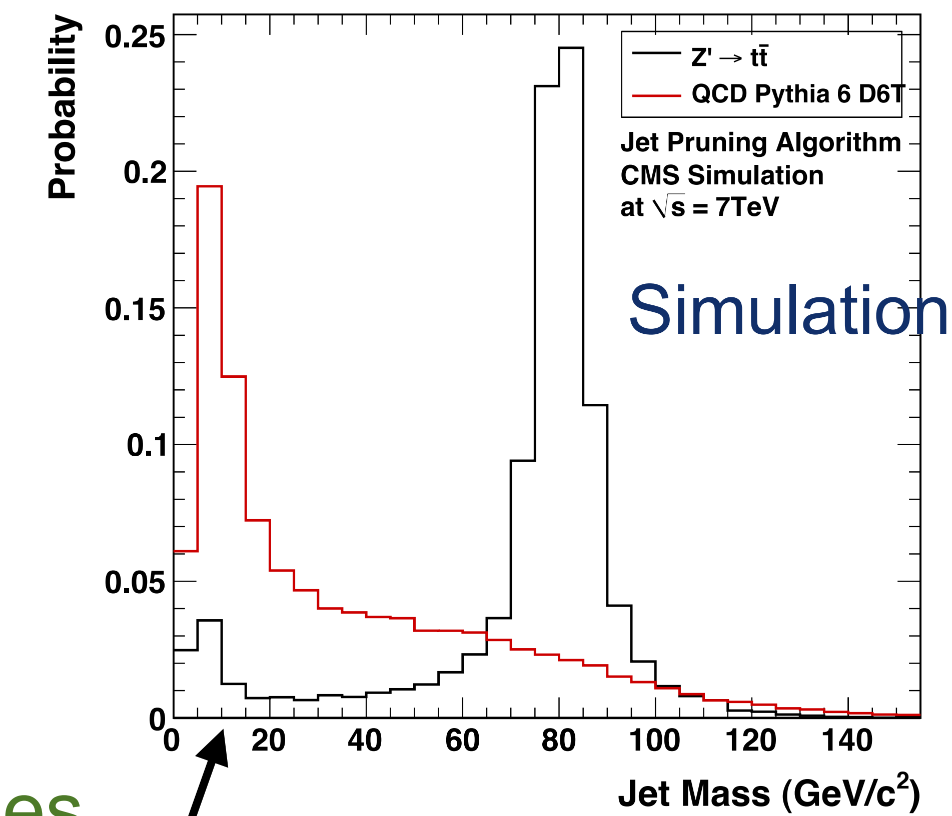
$$\frac{\min(p_{T1}, p_{T2})}{p_{Tp}} > 0.1$$

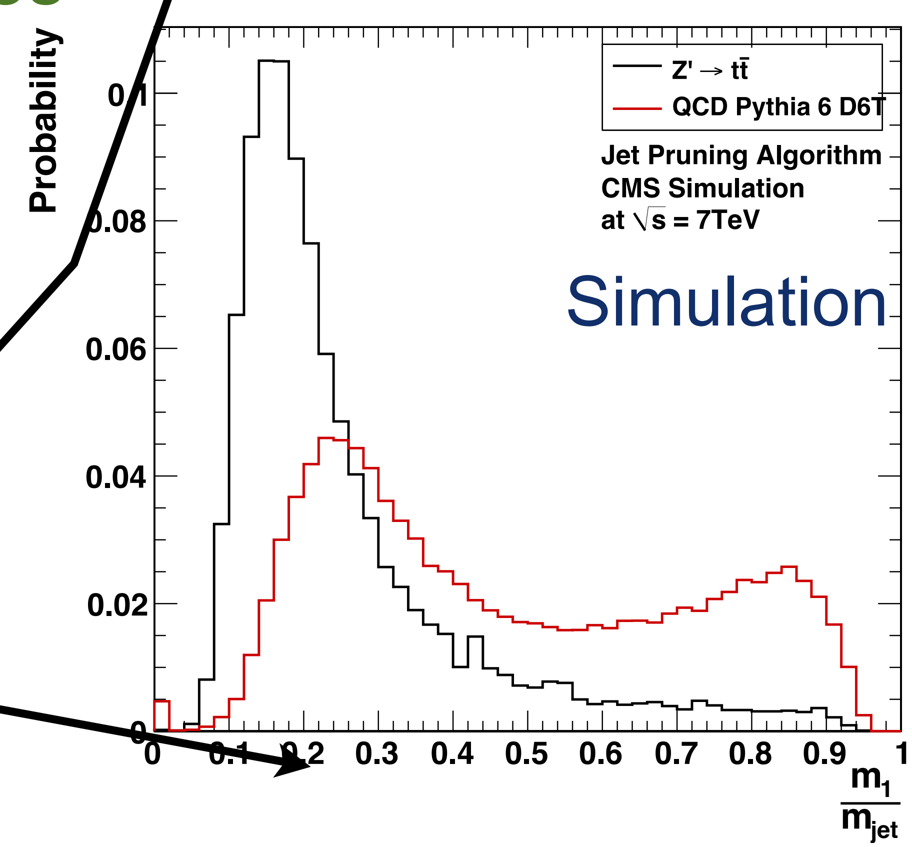
Removes soft particles

$$\Delta R_{12} < 0.5 \times \frac{m_{\text{jet}}}{p_T}$$

Removes wide angle particles

- For W tagging, require:
  - Jet mass in 60-100 GeV/c2
  - Mass drop (mu) < 0.4 $\mu = \frac{m_{j1}}{m_j}$

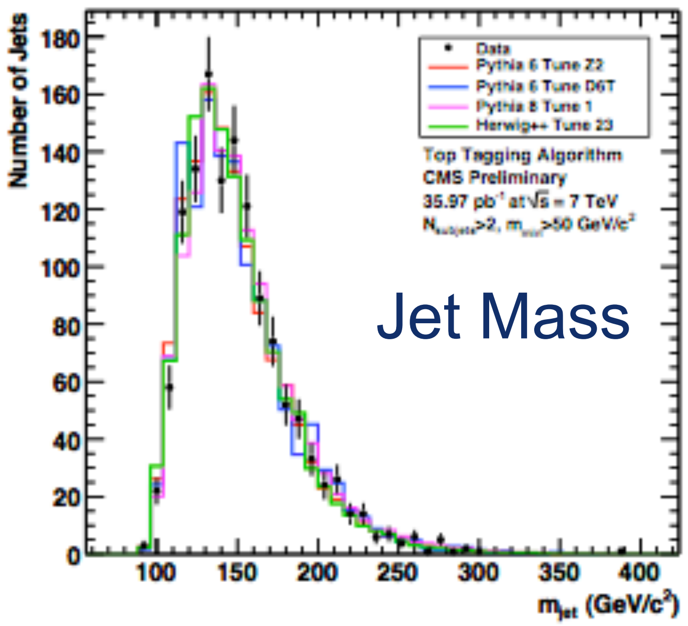


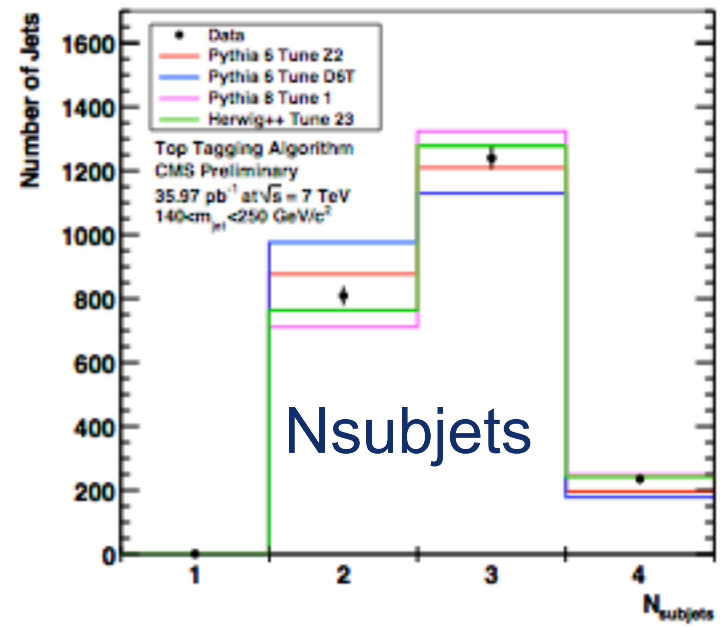


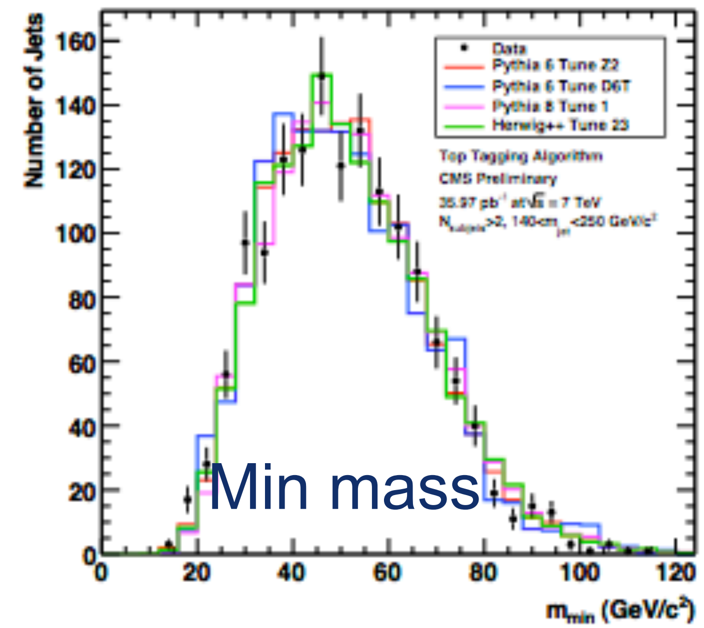
Data/MC Comparisons

- Extensive validation with 36 pb-1 of data outlined in technical report JME-10-013

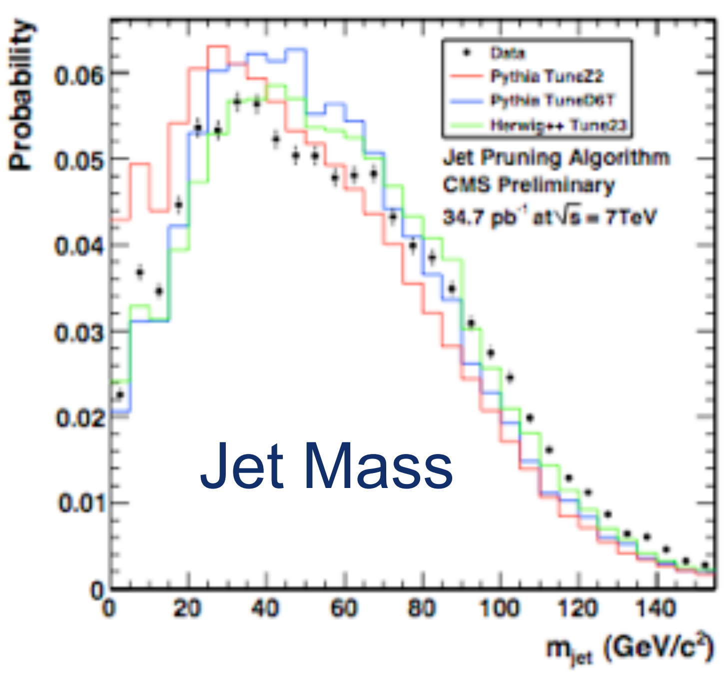
Top Tagging

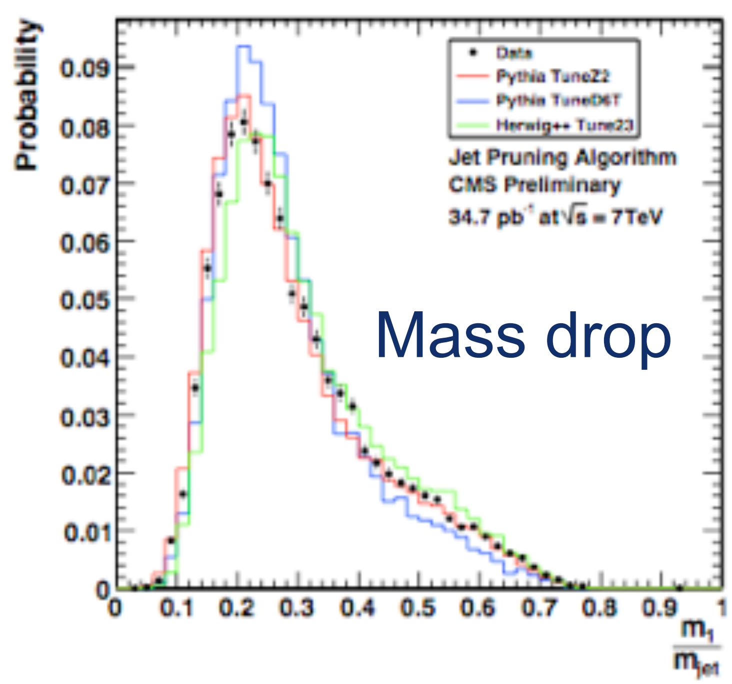






W Tagging

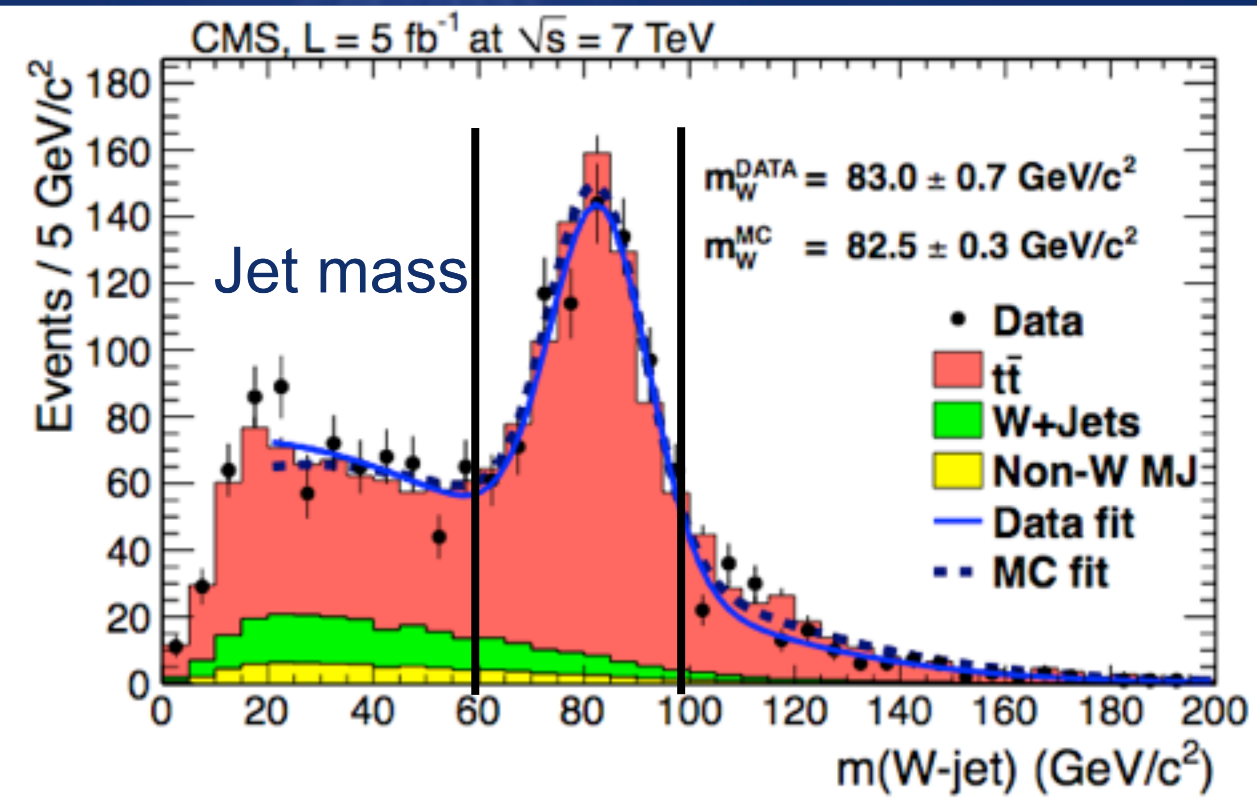


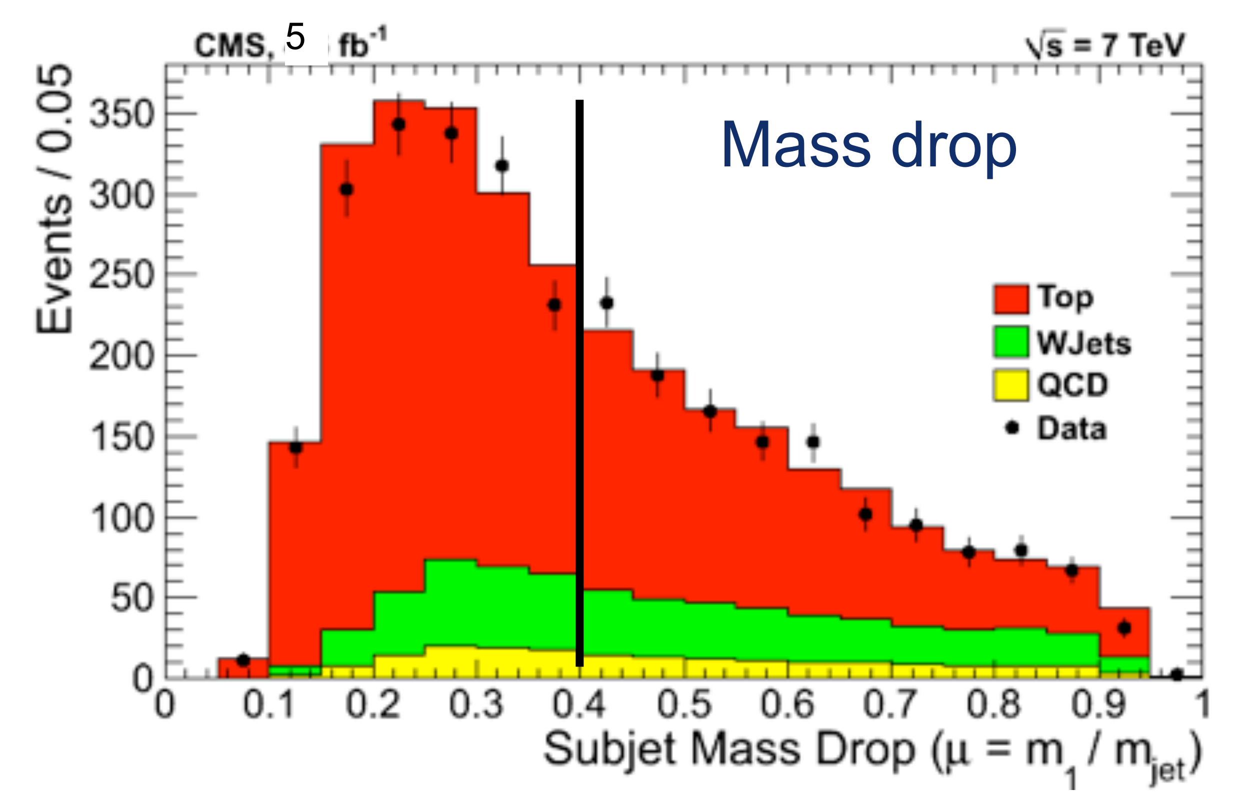


Excellent data/MC agreement, best with Herwig++

Data/MC Comparisons

- Also investigated real W's in semileptonic ttbar events
- Excellent agreement between data and simulation





Outline

- Motivation
- Experimental and algorithmic overview
- Applications

Analyses

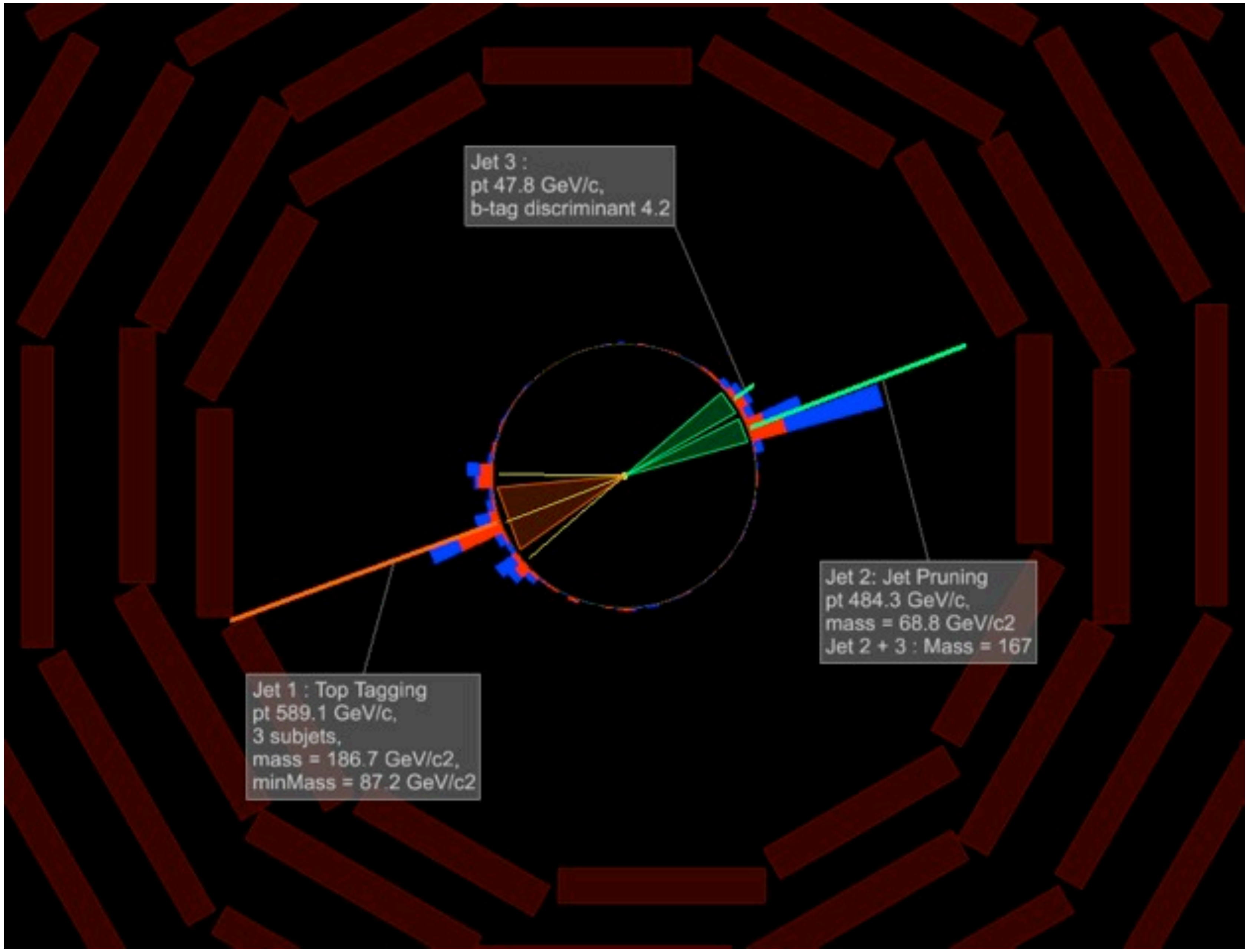
- Boosted ttbar search
- Boosted V + MET search
- Boosted V + ll search
- H->bb
- Dijet analysis

FOCUS

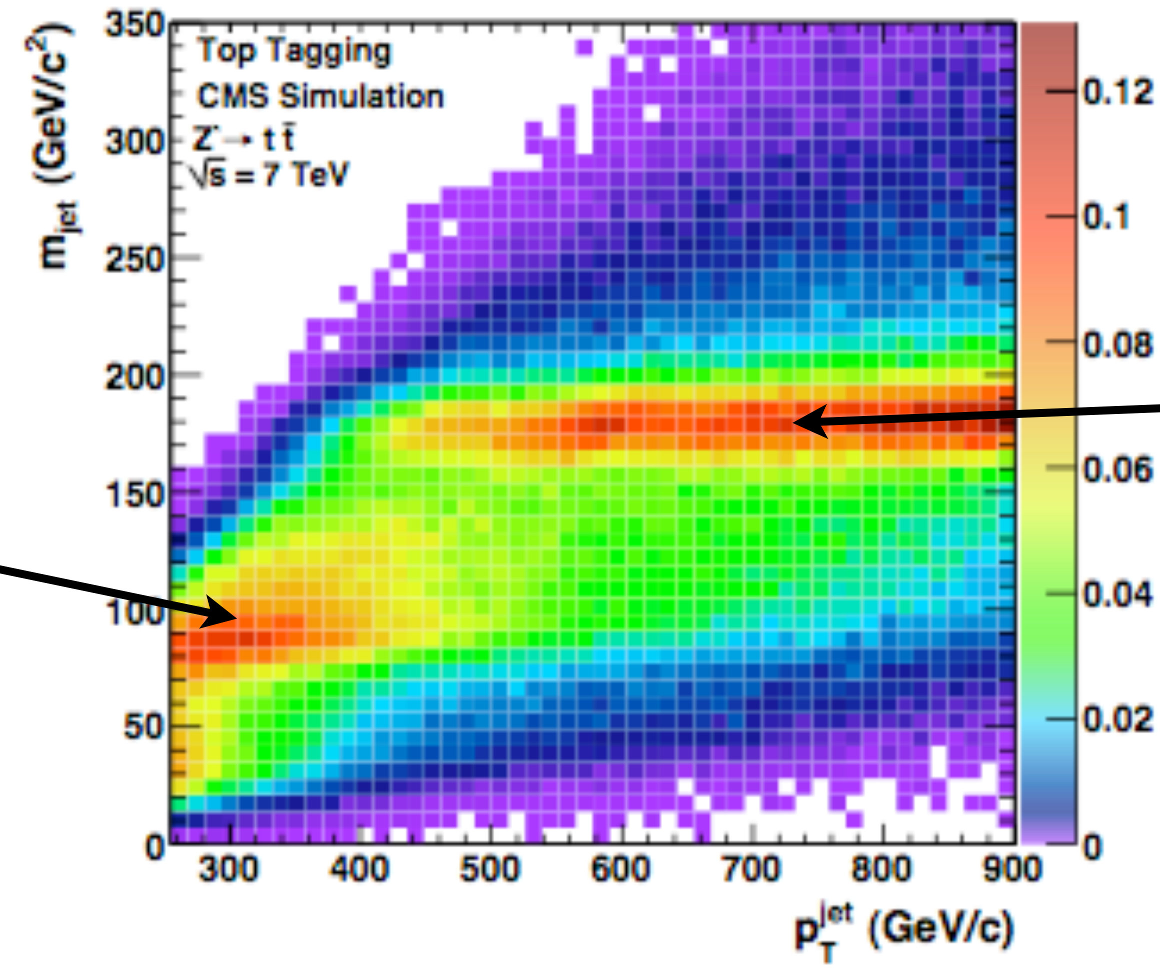
Covered elsewhere

Boosted ttbar

- Look for resonances decaying to boosted ttbar pairs
- Prototype models are Z’ and RS KK gluon
- Data-driven background estimate
- Also perform analysis of general enhancements to ttbar spectrum



Boosted ttbar

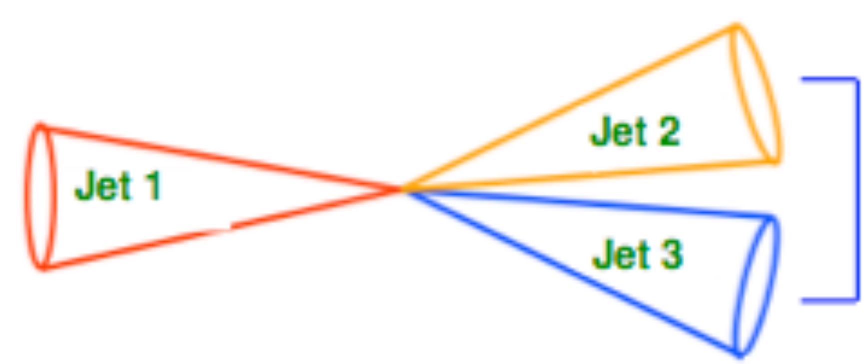


Type 2 : partially merged

Type 1 : fully merged

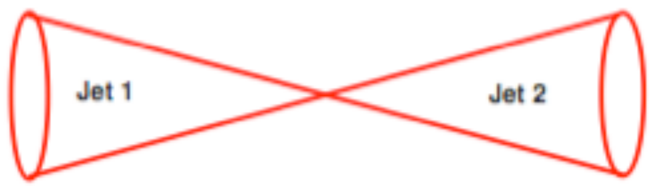
Type 1 + Type 1

Type 1 + Type 2

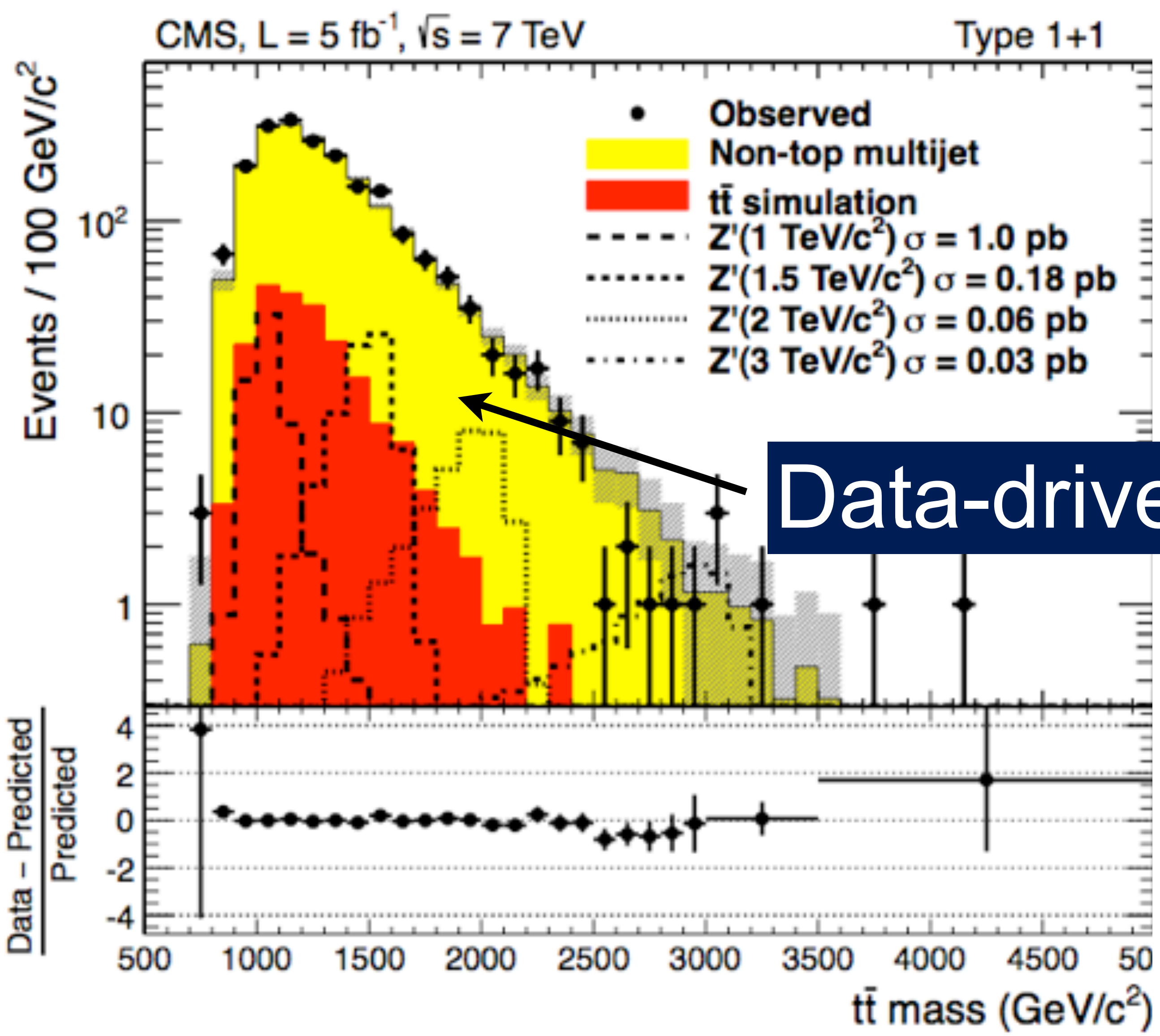


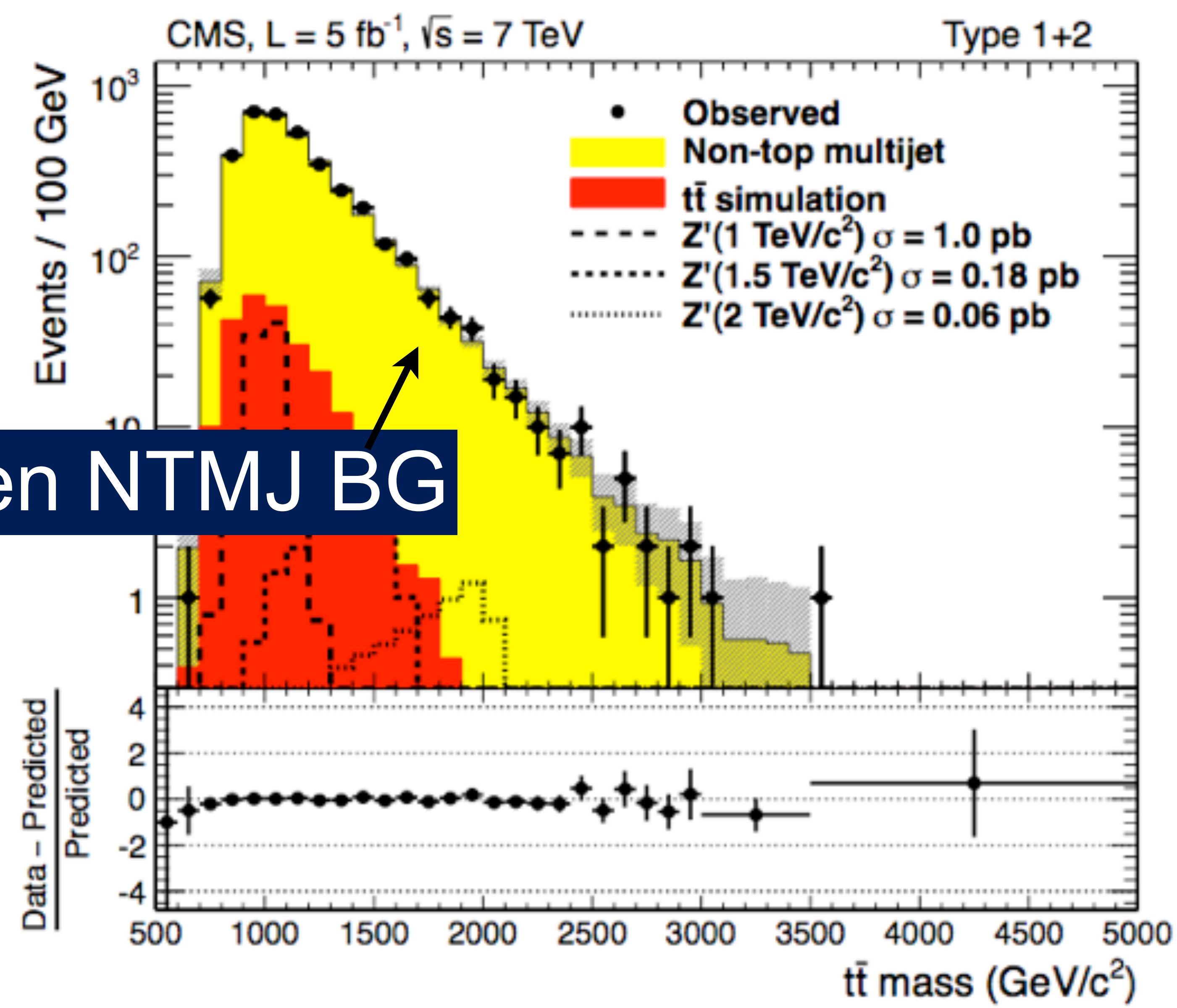
Boosted ttbar

Type 1 + Type 1



Type 1 + Type 2

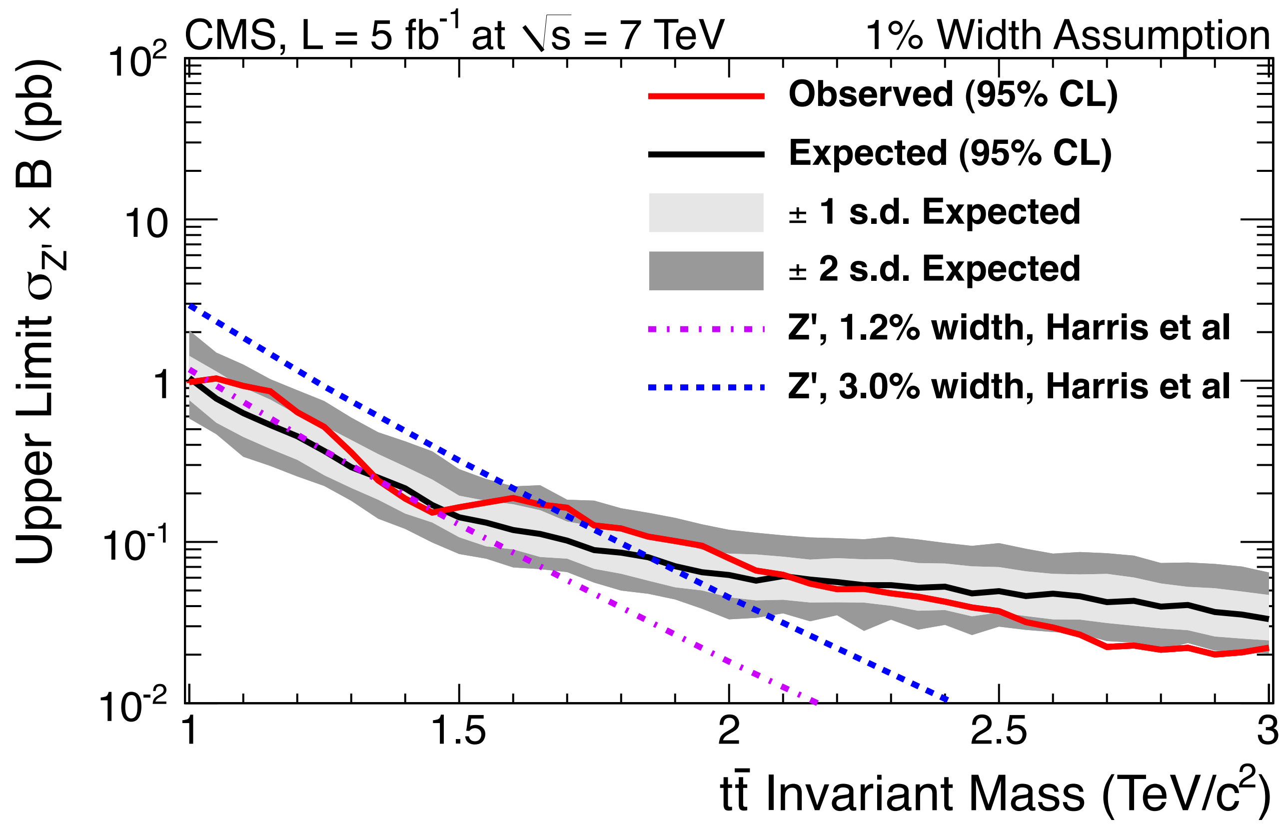


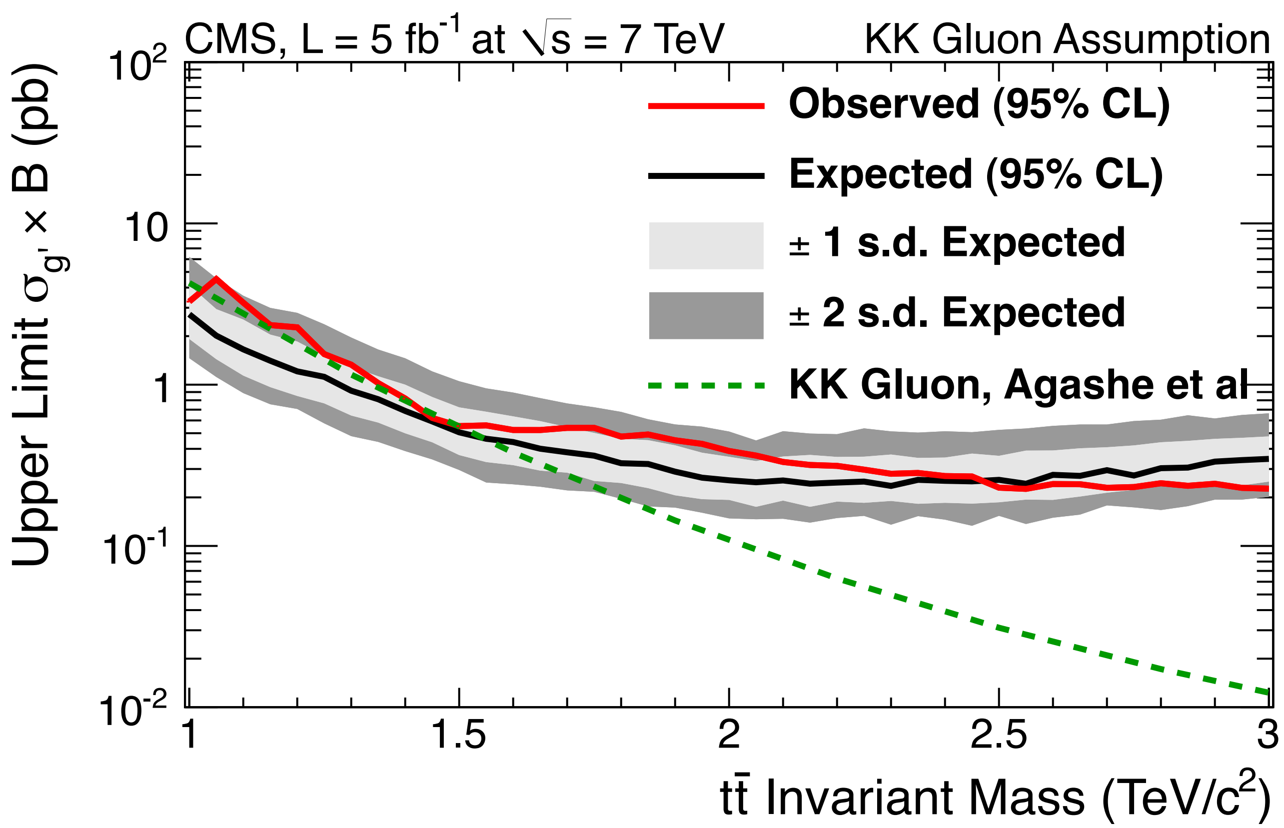


Data-driven NTMJ BG

Boosted ttbar

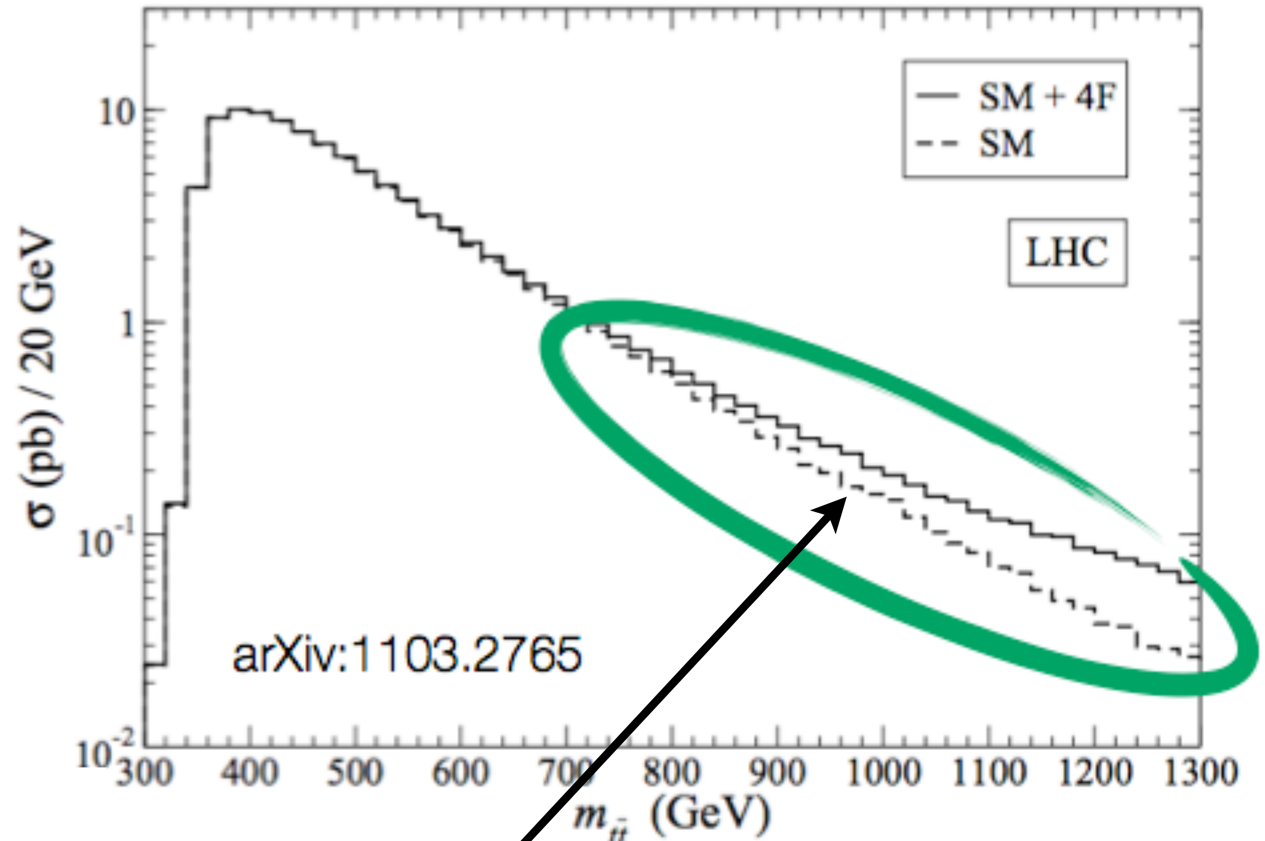
- Examined Z' with 1% width, 10% width (not shown), and KK gluon
- Extensive exclusions everywhere
- Combination with other ttbar analyses is ongoing!

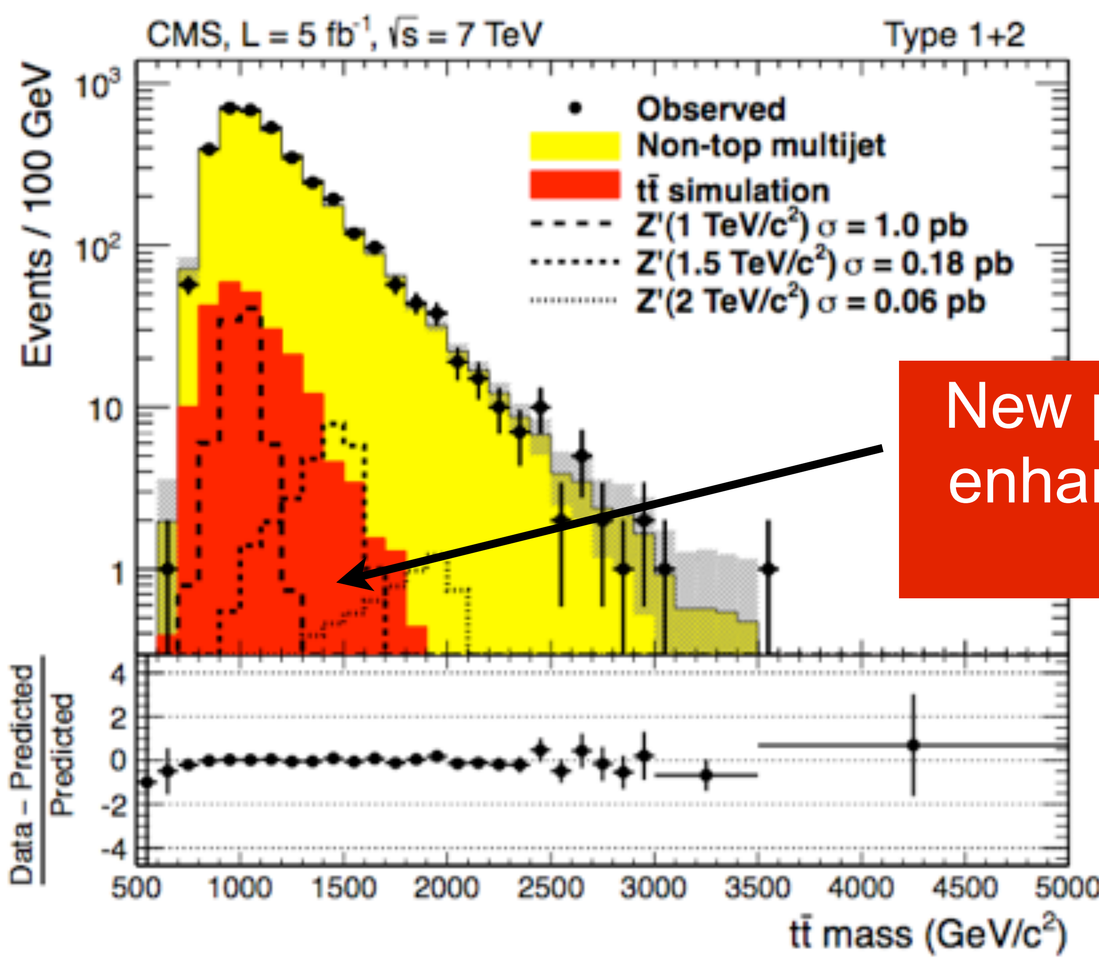




Boosted ttbar

- Can also check enhancement to ttbar XS
  - Use counting exp, absolute background normalization for mtt > 1 TeV
  - Aguilar-Saavedra, Perez-Victoria (arXiv:1103.2765v2 [hep-ph])
  - Delaunay et al (arXiv:1103.2297v3 [hep-ph])

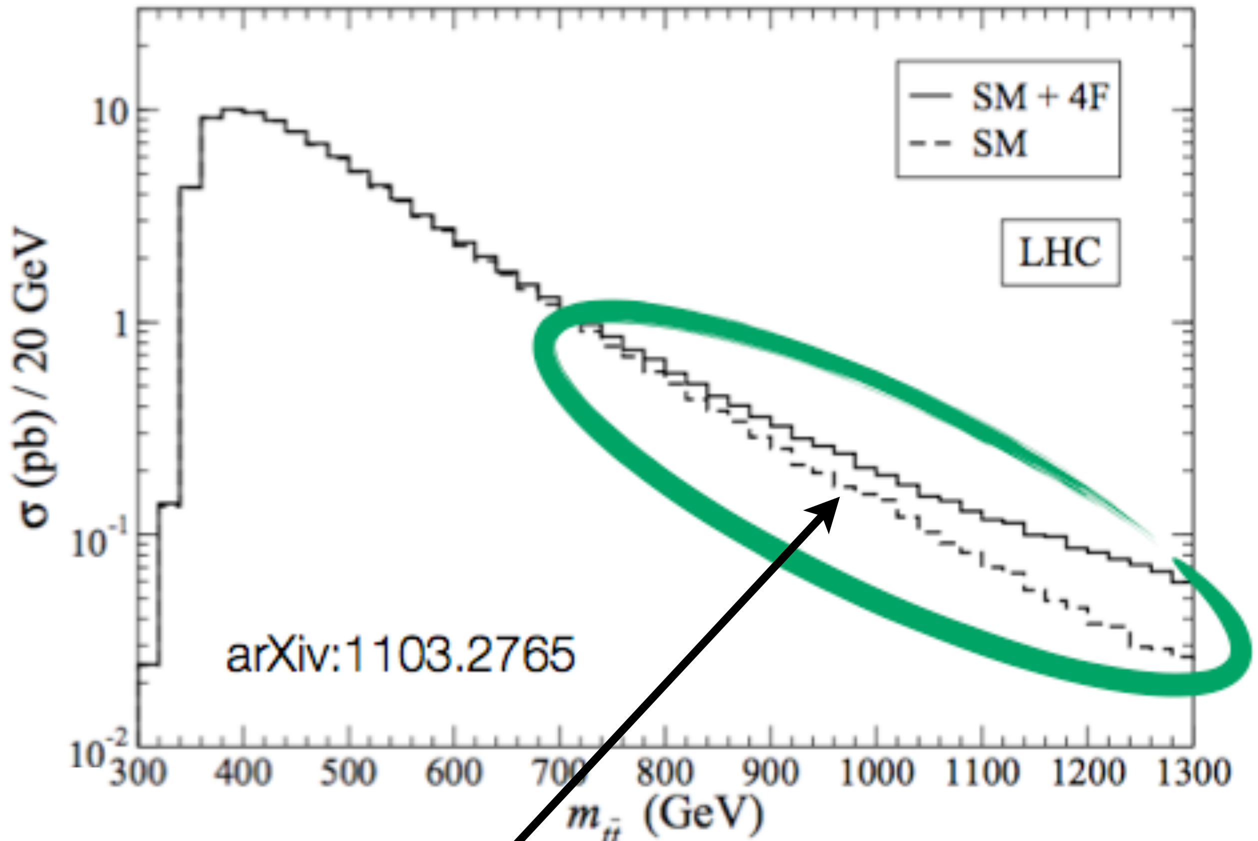


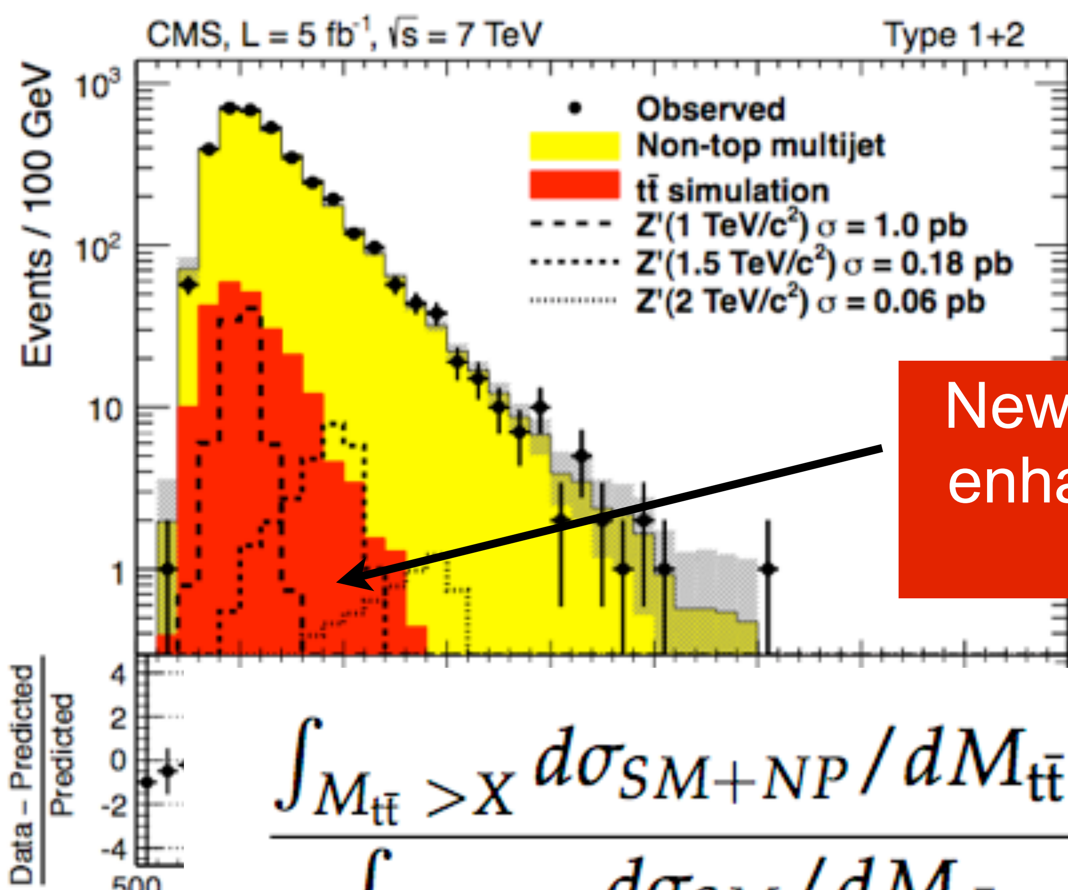


New physics would enhance ttbar from SM

Boosted ttbar

- Can also check enhancement to ttbar XS
  - Use counting exp, absolute background normalization for mtt > 1 TeV
  - Aguilar-Saavedra, Perez-Victoria (arXiv:1103.2765v2 [hep-ph])
  - Delaunay et al (arXiv:1103.2297v3 [hep-ph])



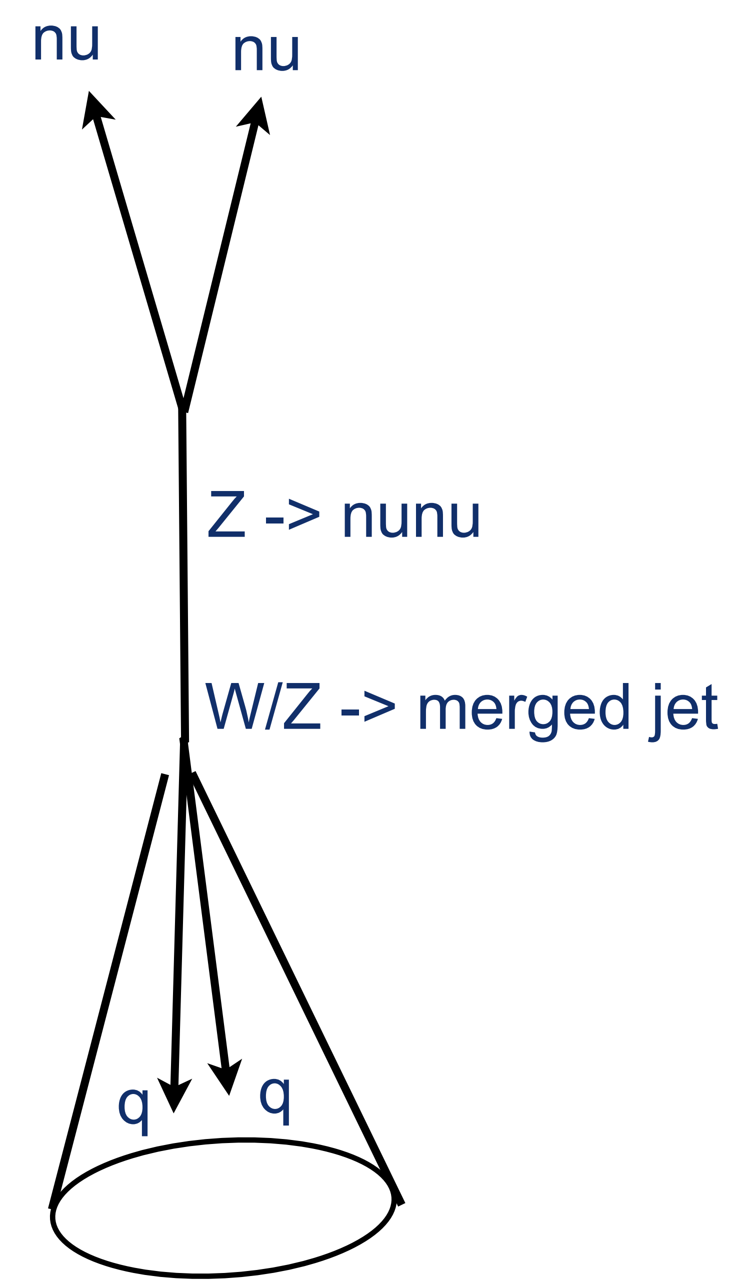


New physics would enhance ttbar from SM

$$\frac{\int_{M_{t\bar{t}}>X} d\sigma_{SM+NP}/dM_{t\bar{t}}}{\int_{M_{t\bar{t}}>X} d\sigma_{SM}/dM_{t\bar{t}}} < 2.6 \quad (X = 1\ \text{TeV})$$

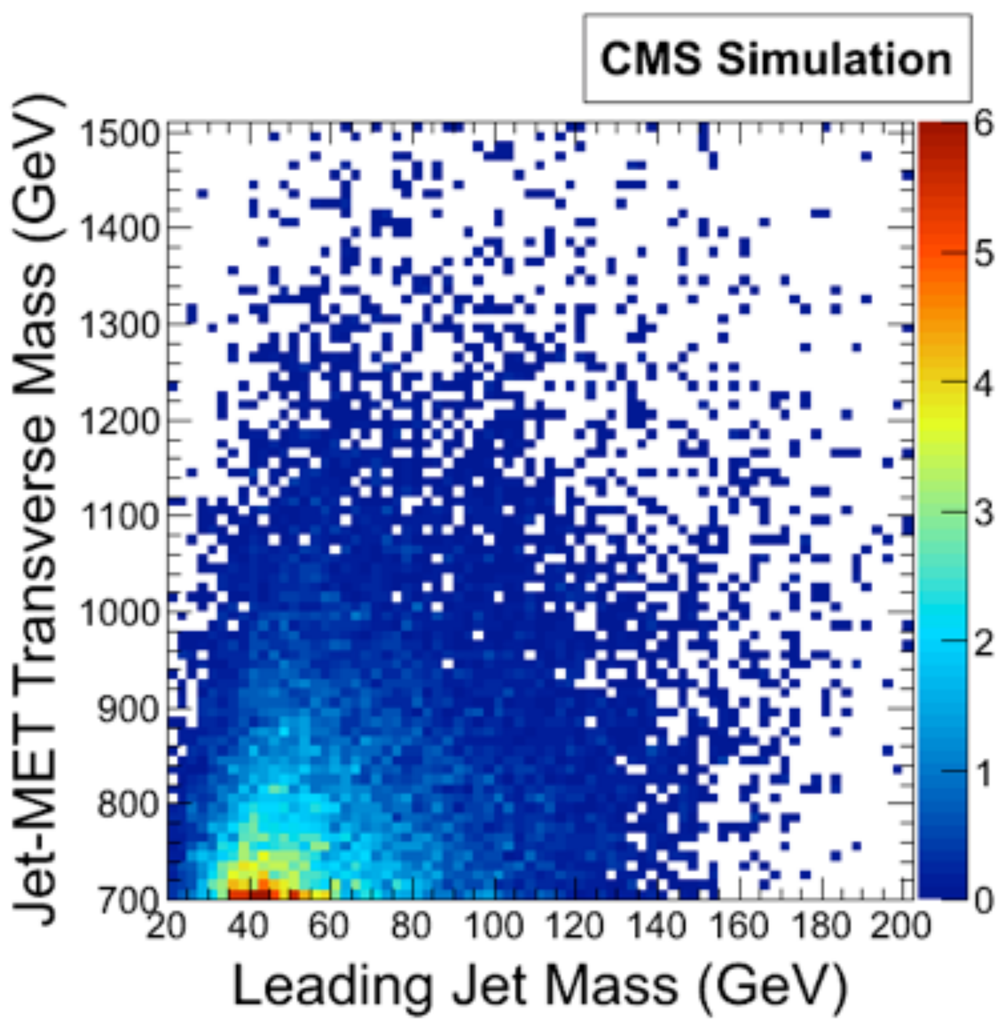
Boosted V+MET

- Look for resonances to V+Z -> merged jet + MET
- Prototype signals are W', RS gravitons, technirho
- Data-driven background estimate

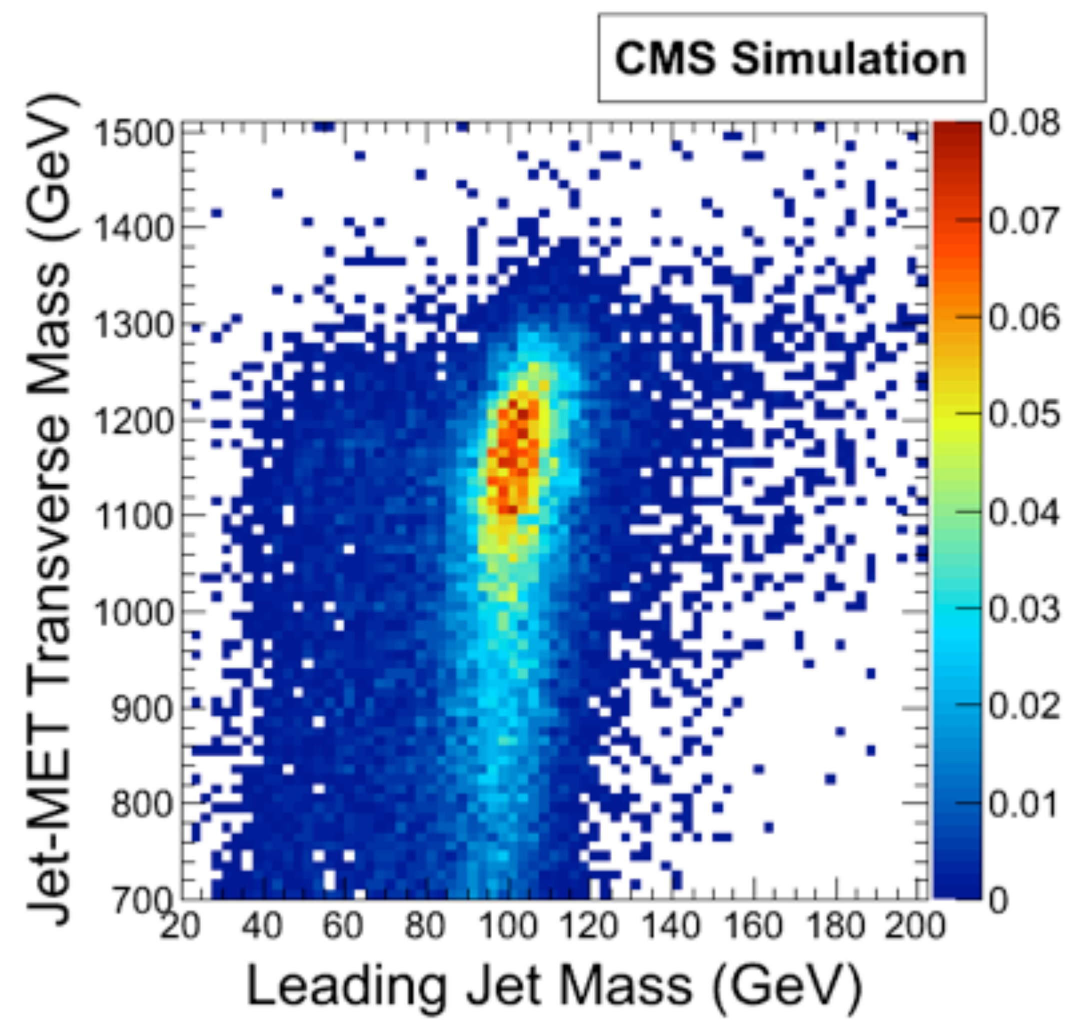


- Use simpler V-tagging :
  - anti-kT jets, D = 0.7, no jet grooming
  - $m_{jet} > 70$ GeV
-

SM backgrounds



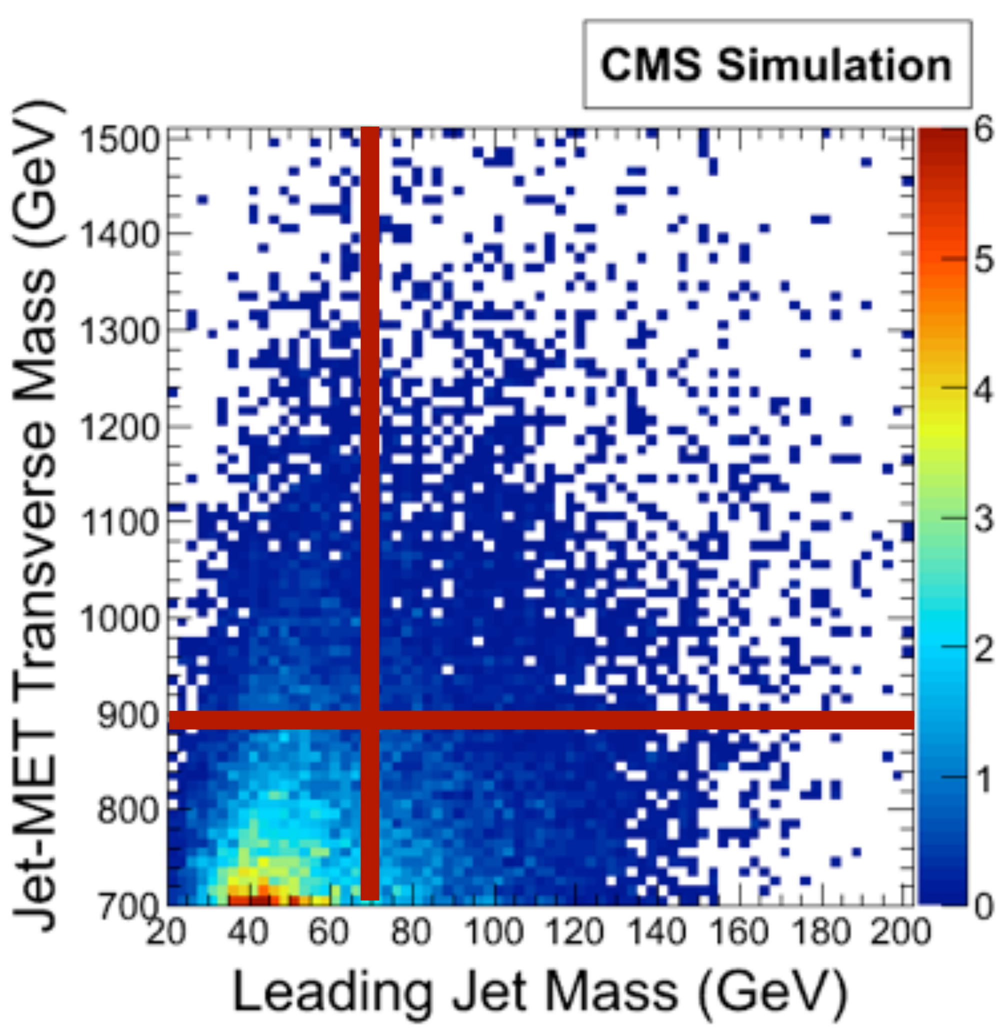
RS graviton, m=1250 GeV



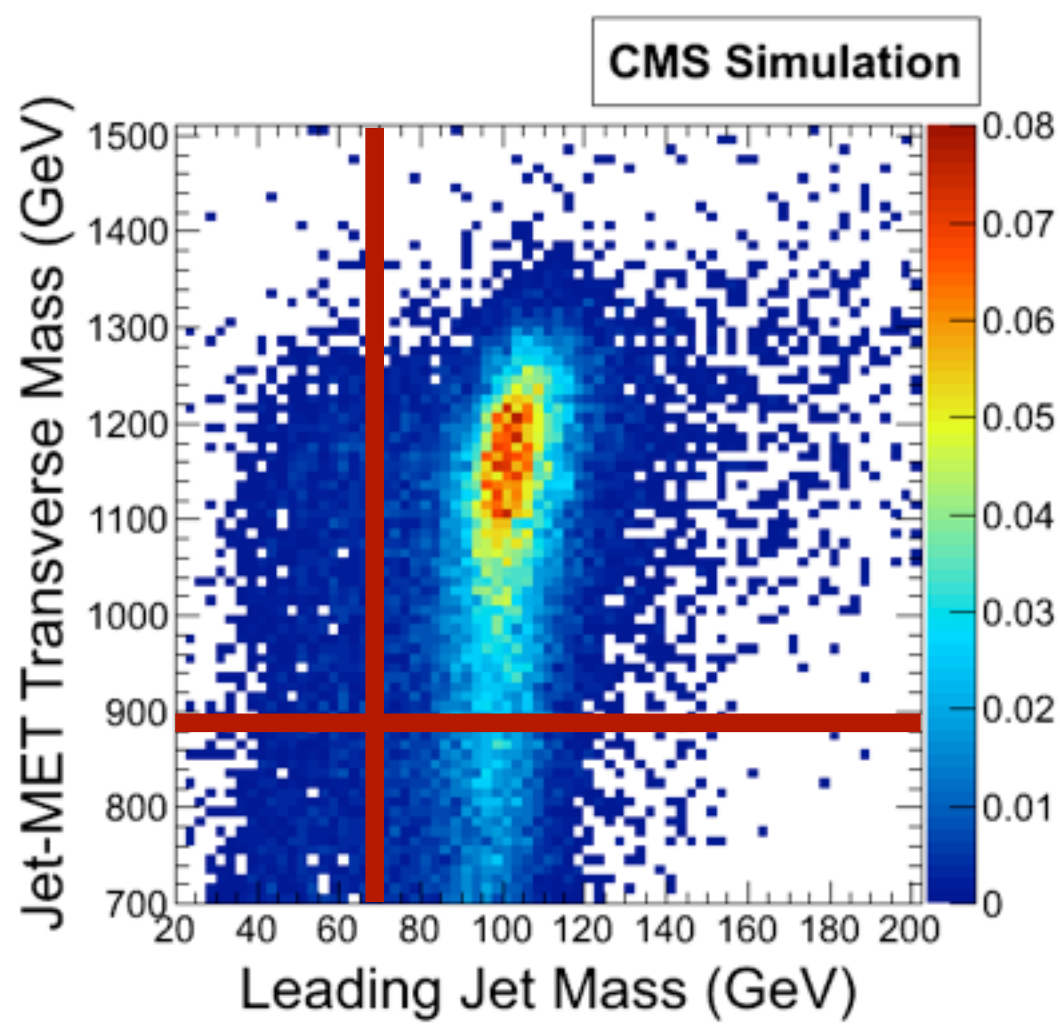
Boosted V+MET

- Use simpler V-tagging :
  - anti-kT jets, D = 0.7, no jet grooming
  - m_{jet} > 70 GeV
- Use data-driven background (ABCD method)

SM backgrounds

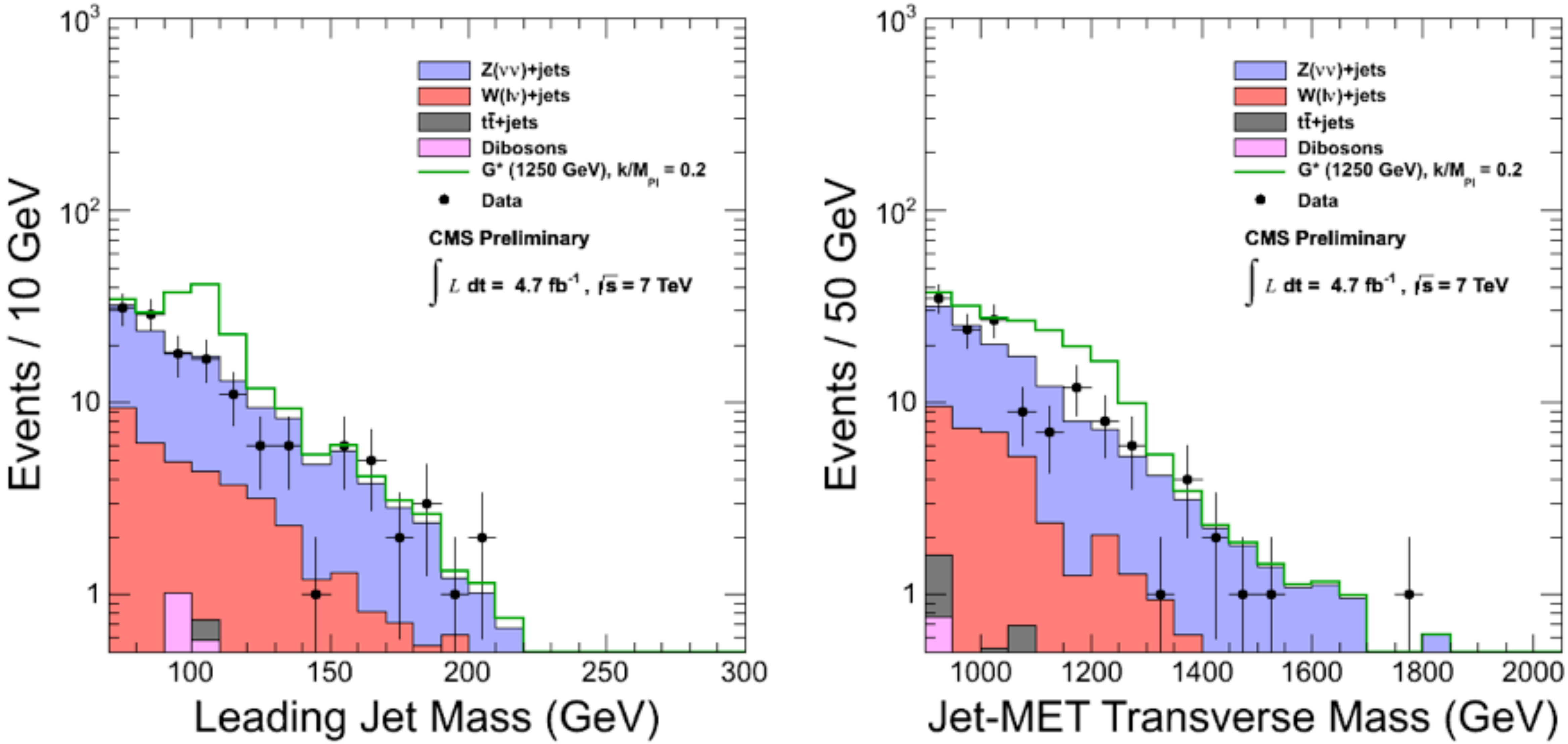


RS graviton, m=1250 GeV



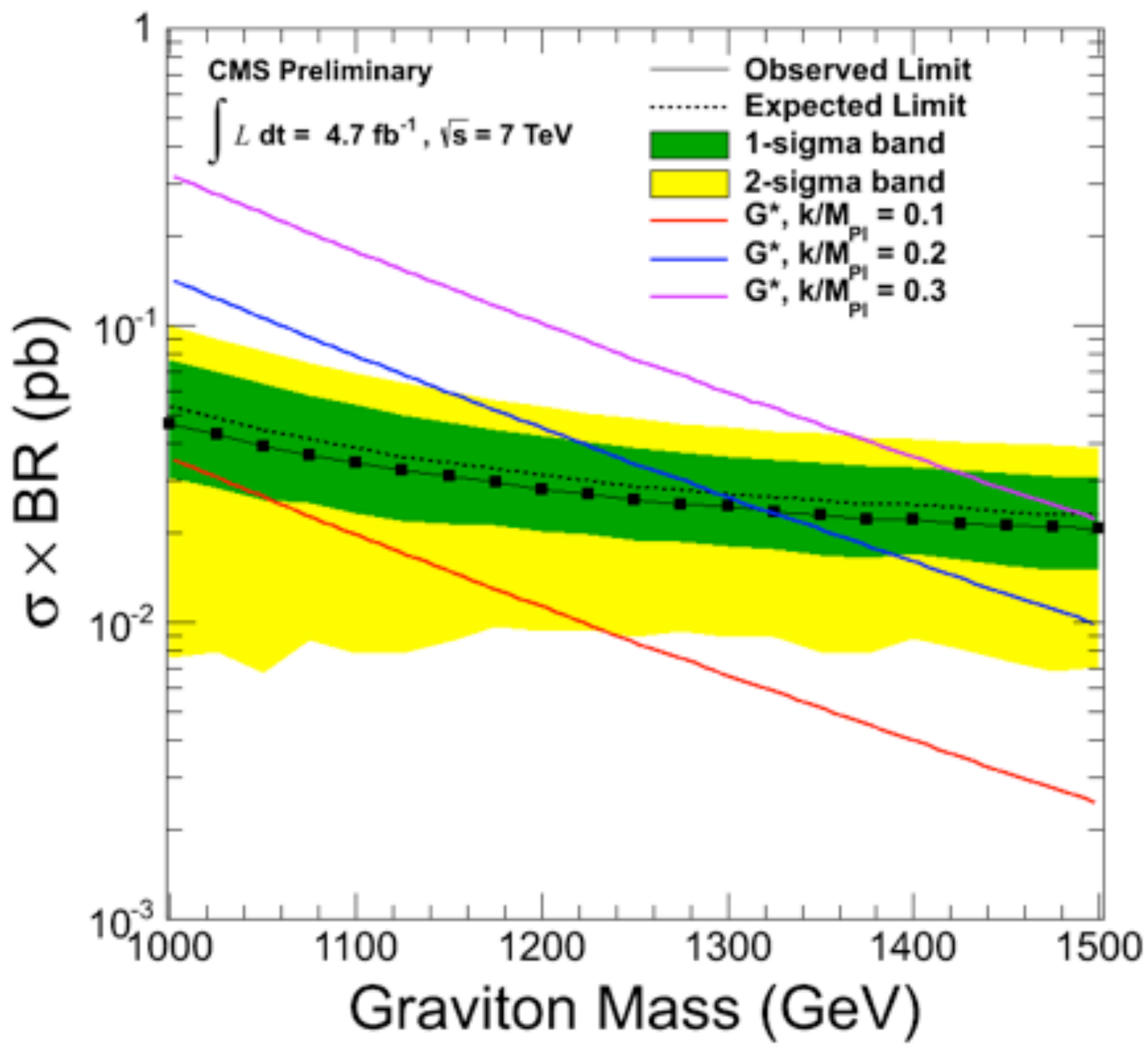
Boosted V+MET

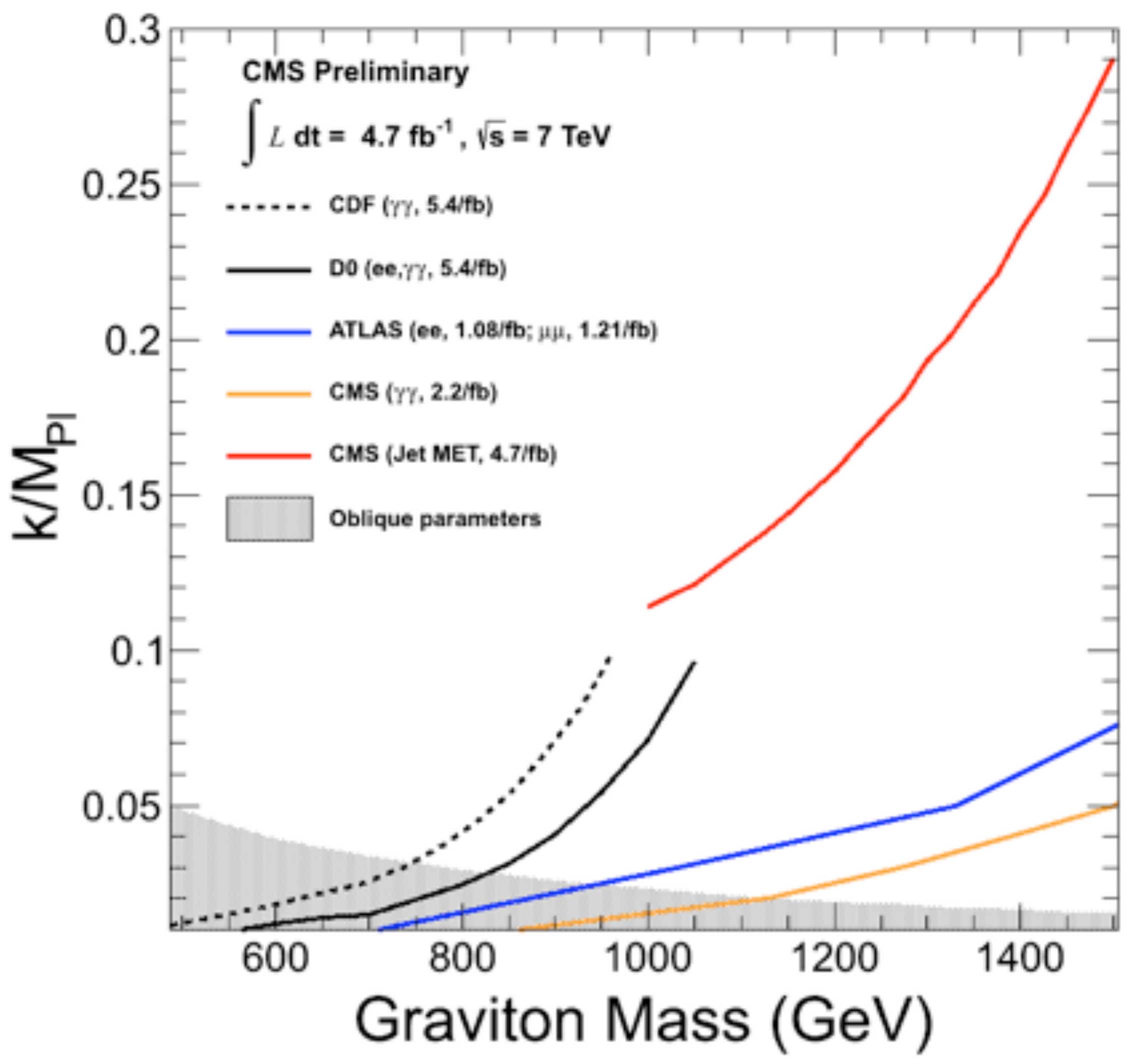
- Good agreement with MC expectations
- No signal observed



Boosted V+MET

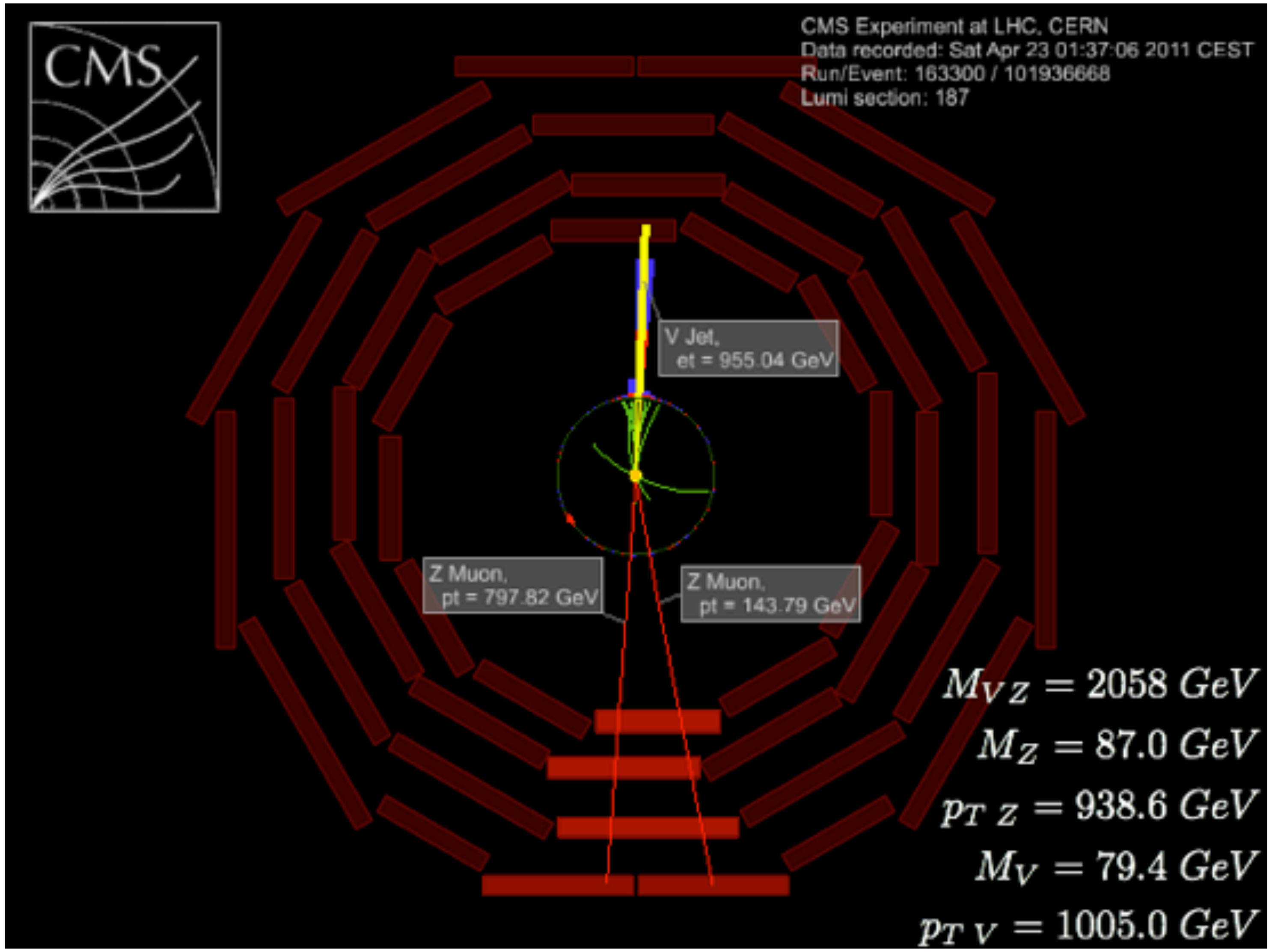
- Limits set on cross section, and also in k/M_{pl} vs m (graviton) interpretation





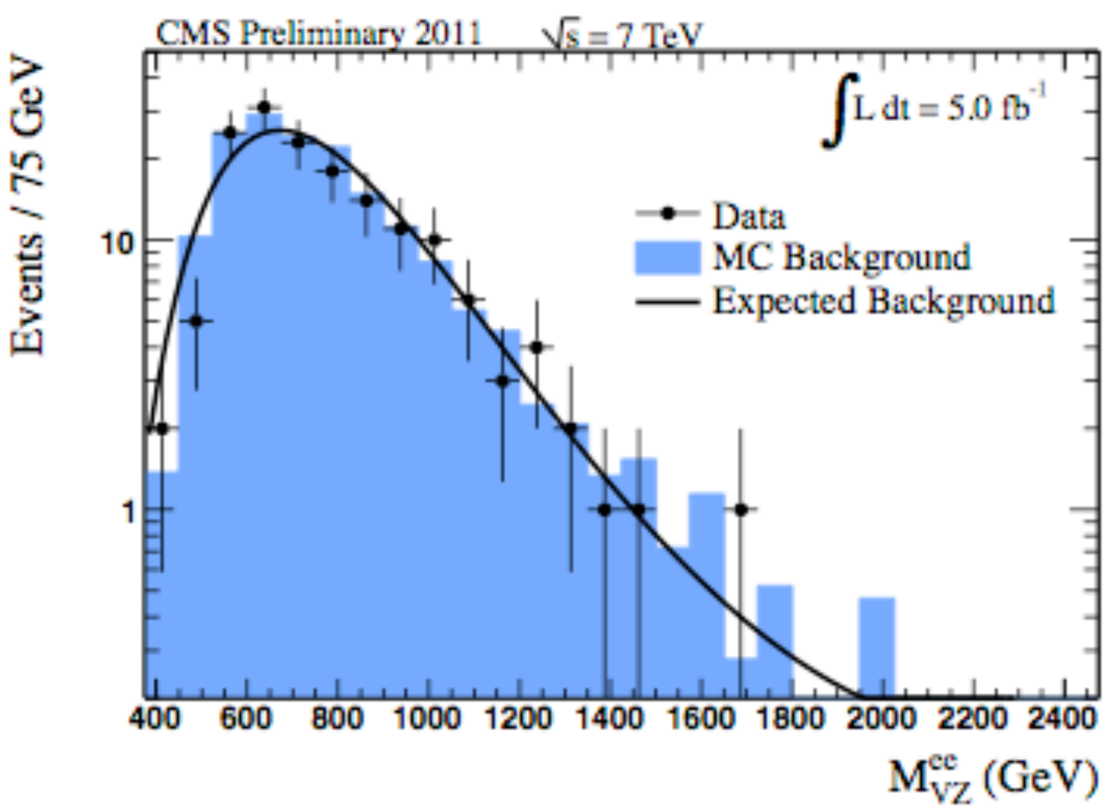
Boosted V+ll

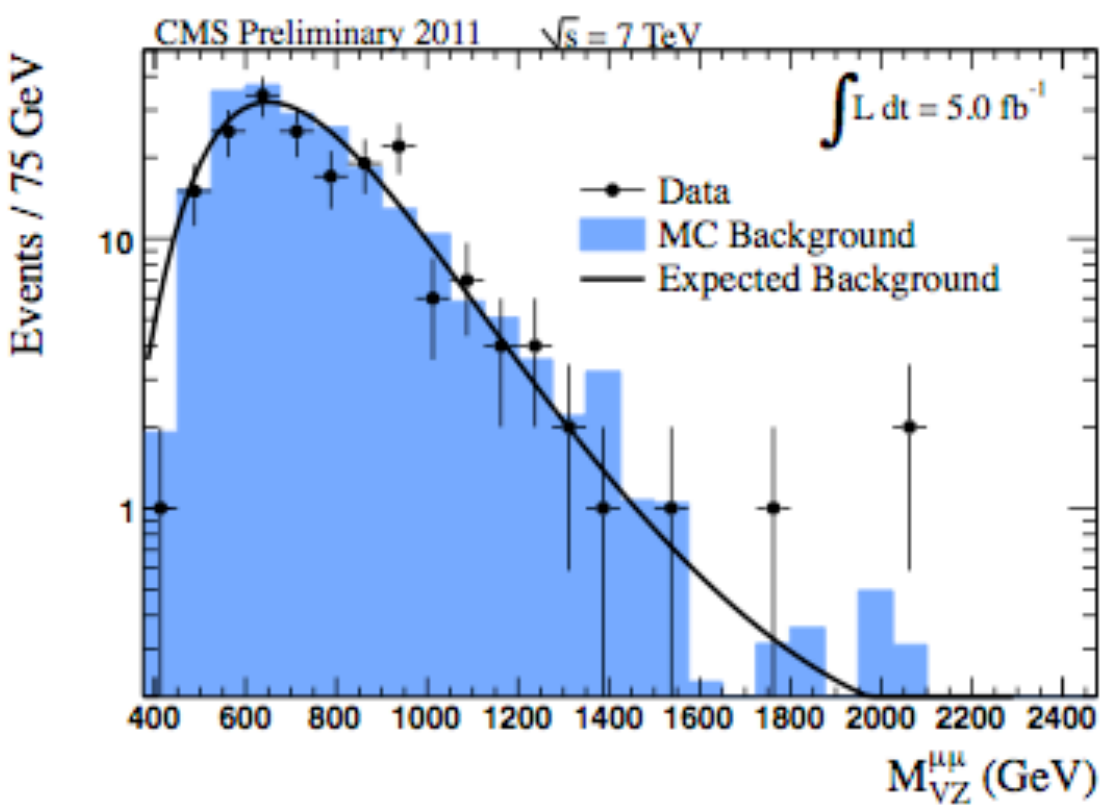
- Hot off the presses!
- Look for resonances to V+Z -> ll + merged jet
- Prototype signals are W', RS gravitons, technirho
- Data-driven background estimate
- Use same V-tagging as boosted V + MET search



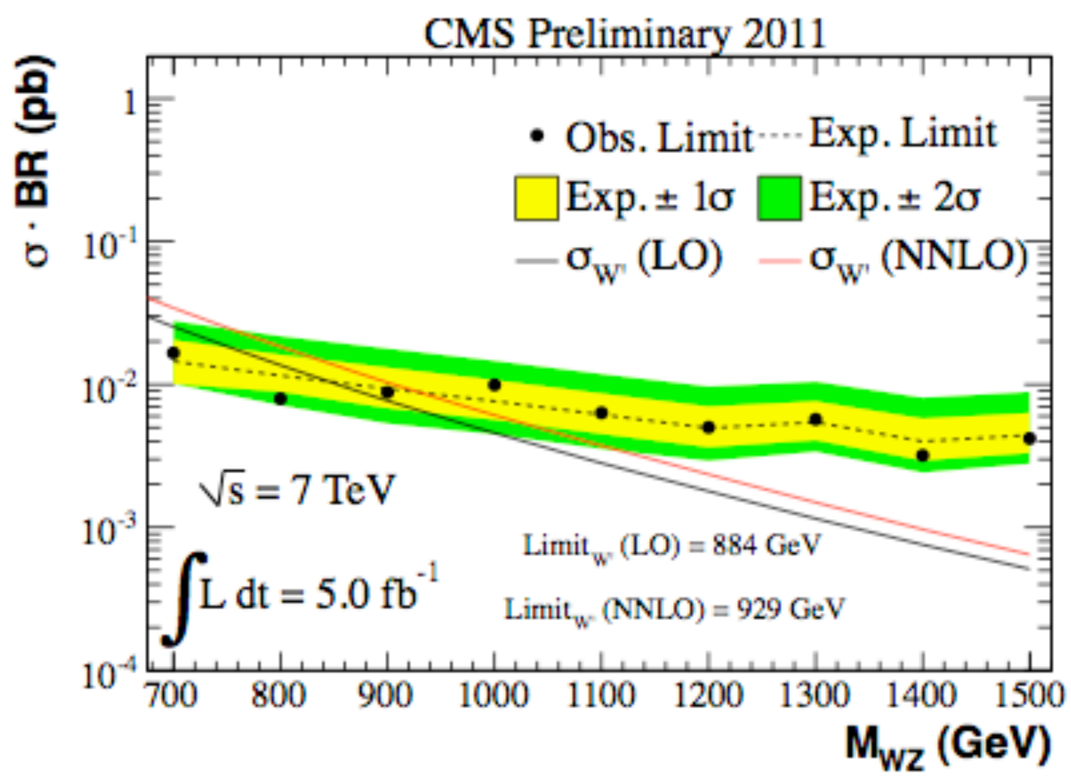
Boosted V+ll

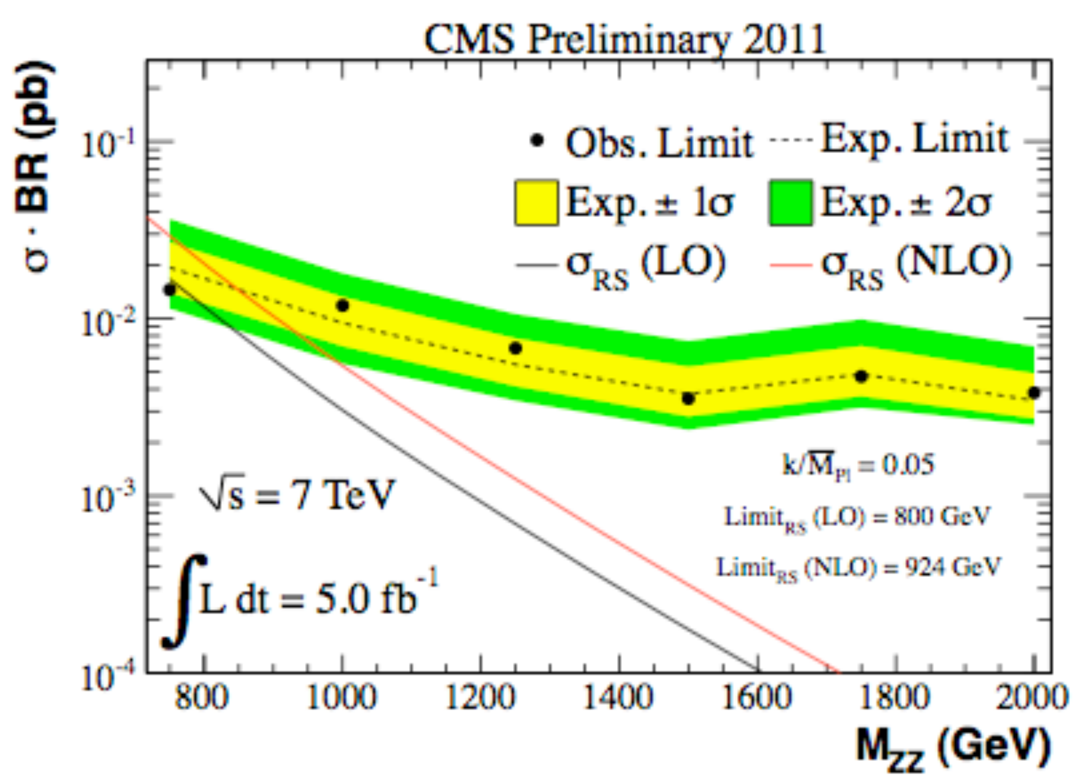
- Use low-jet-mass sideband as control region
- Extrapolate to signal region (accounting for kinematic effects taken from MC)
- Good agreement between data-driven background estimate, and MC expectation



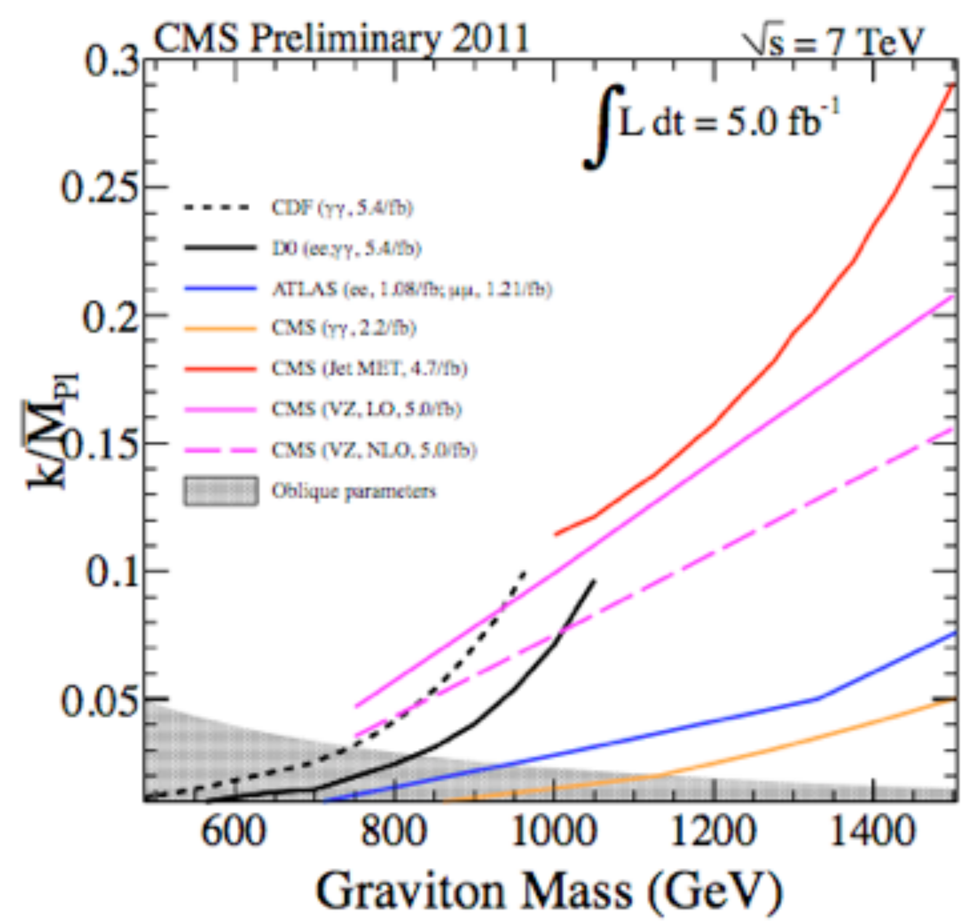


Boosted V+ll





- Limits set on cross section for W' and RS graviton models
- Also interpretation of RS graviton parameters (as in previous case)



Coming soon to a conference near you...

top tagging with trimming

qcd measurements

stops with boosted tops

t' with W-tagging

boosted V+V in dijets

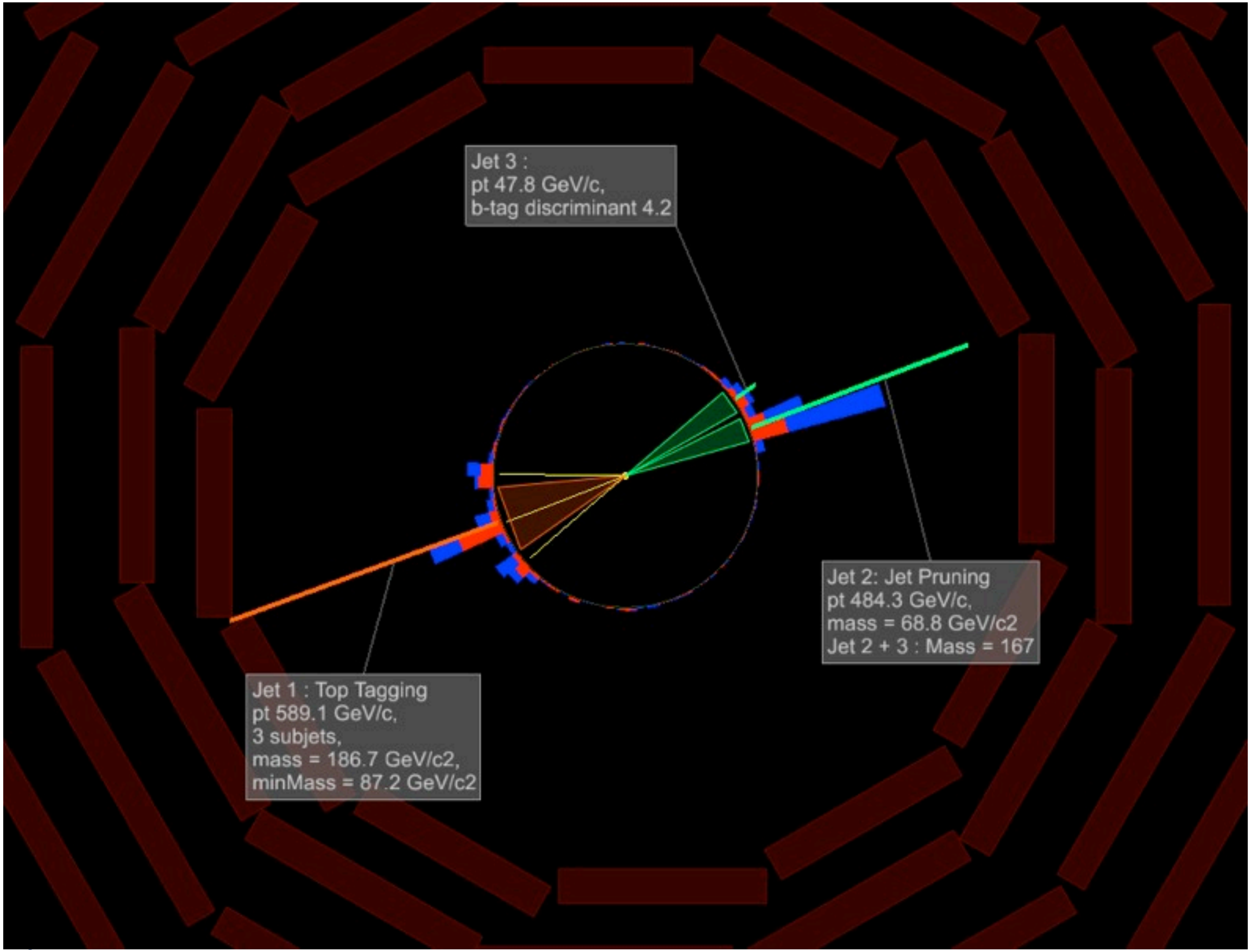
top tagging with n-subjettiness

Summary and conclusions

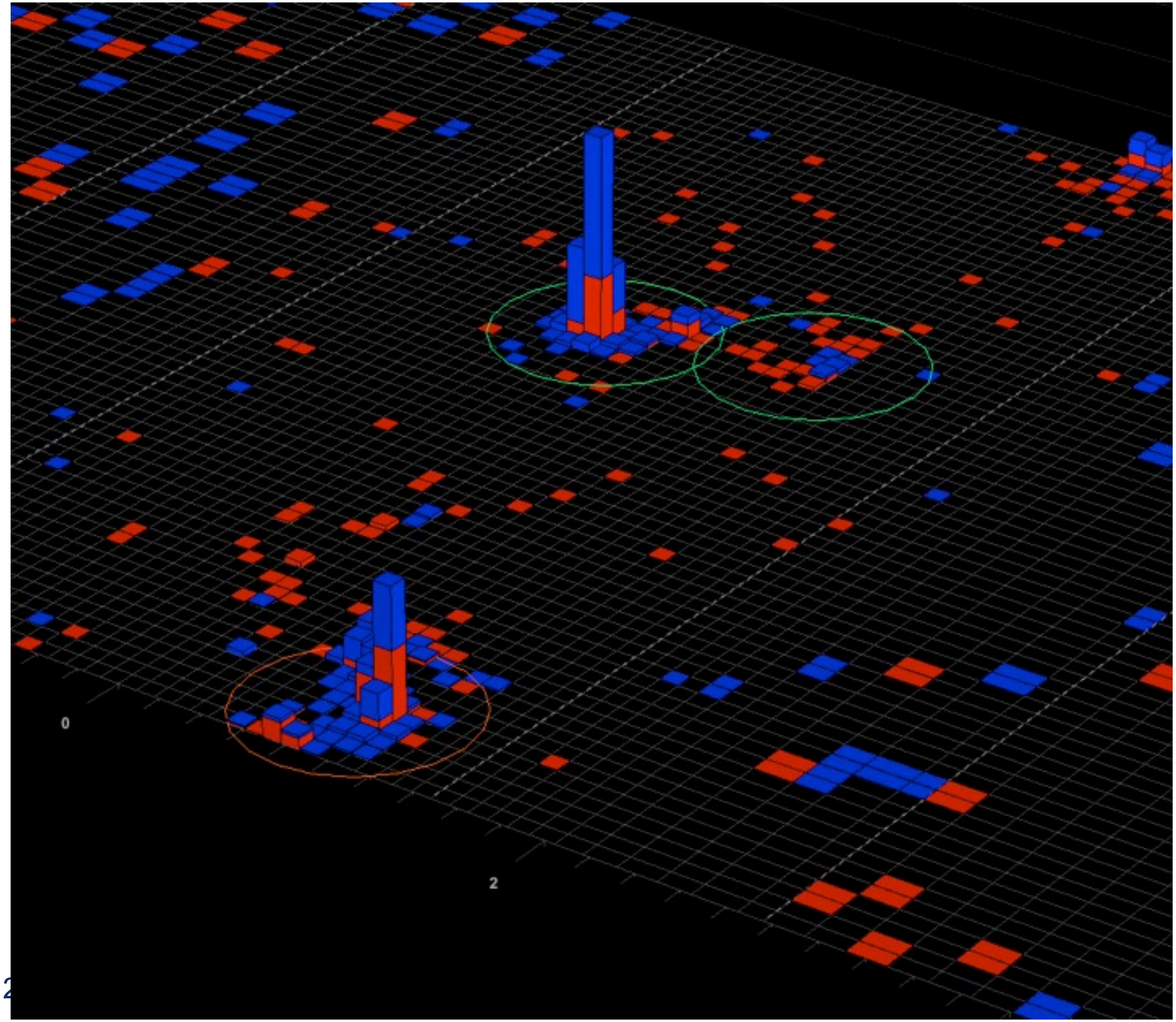
- Jet substructure tools are becoming widely accepted at CMS
- "Out of the gate" the boosted channels are giving good performance relative to leptonic channels
- Already some publications and approved analyses, and many more on the way

Backups

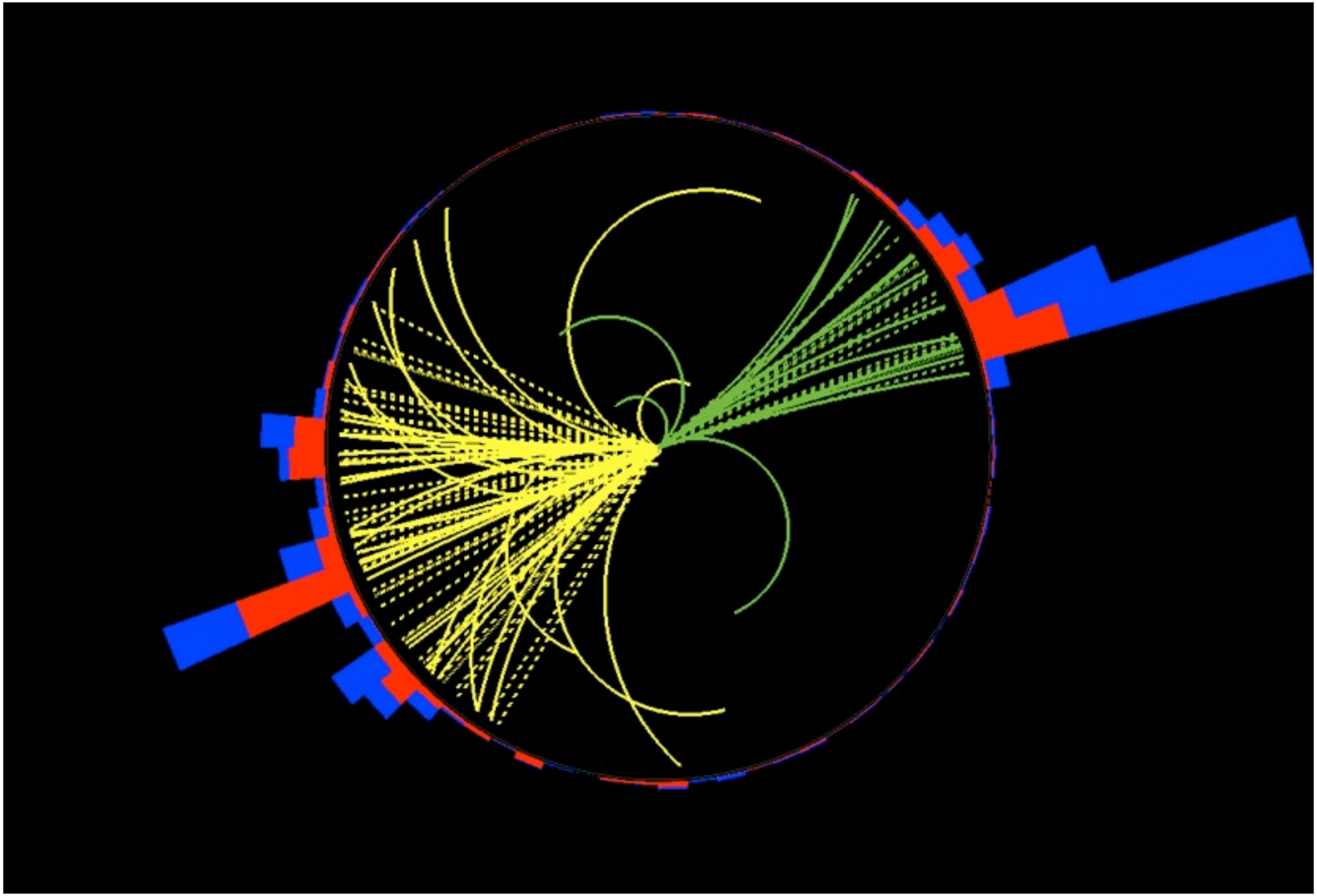
Rho/Phi : Subjet view



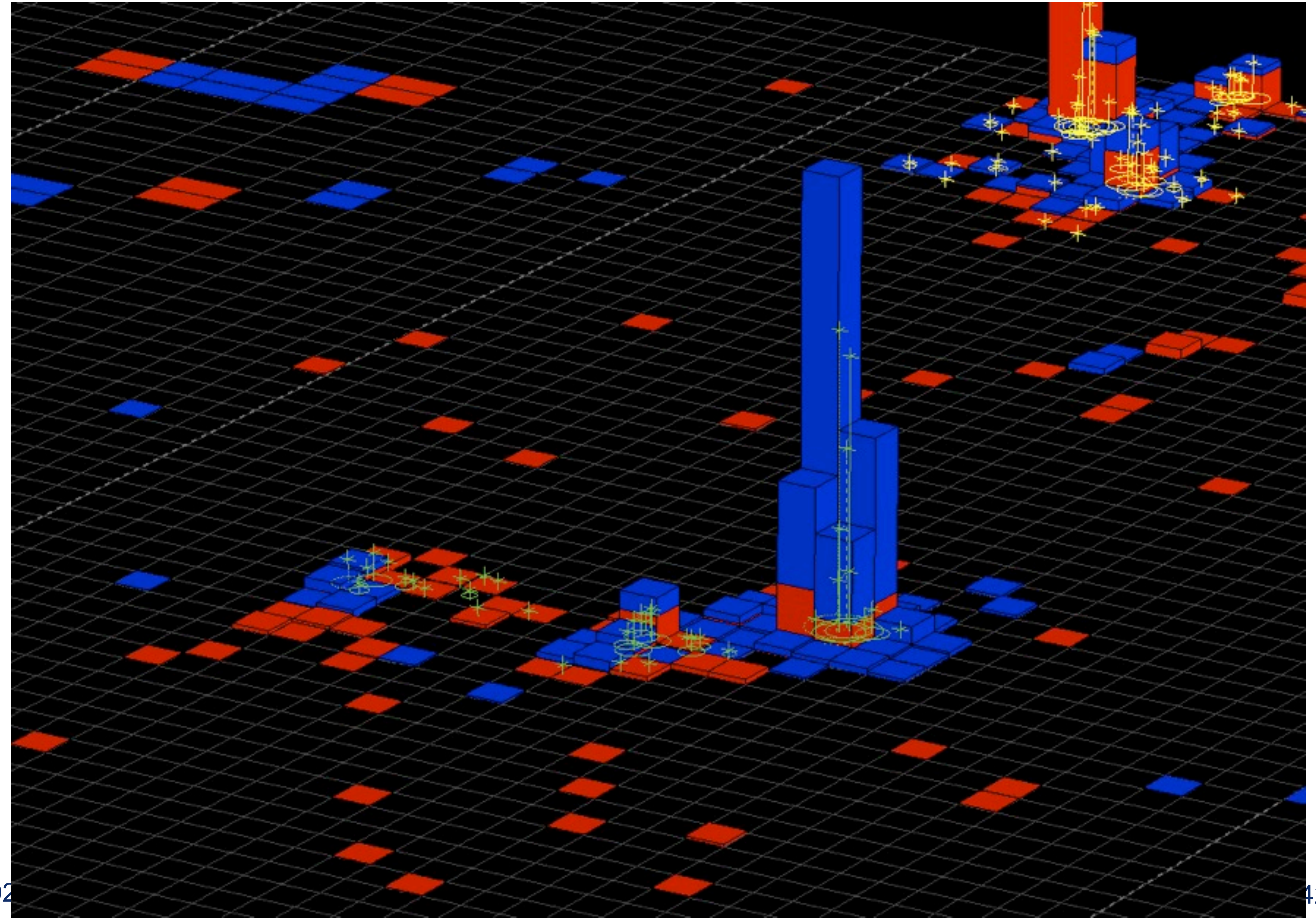
Lego : Subjet view



Rho/Phi: PF View



Lego : PF View



Sequential Clustering Algorithms

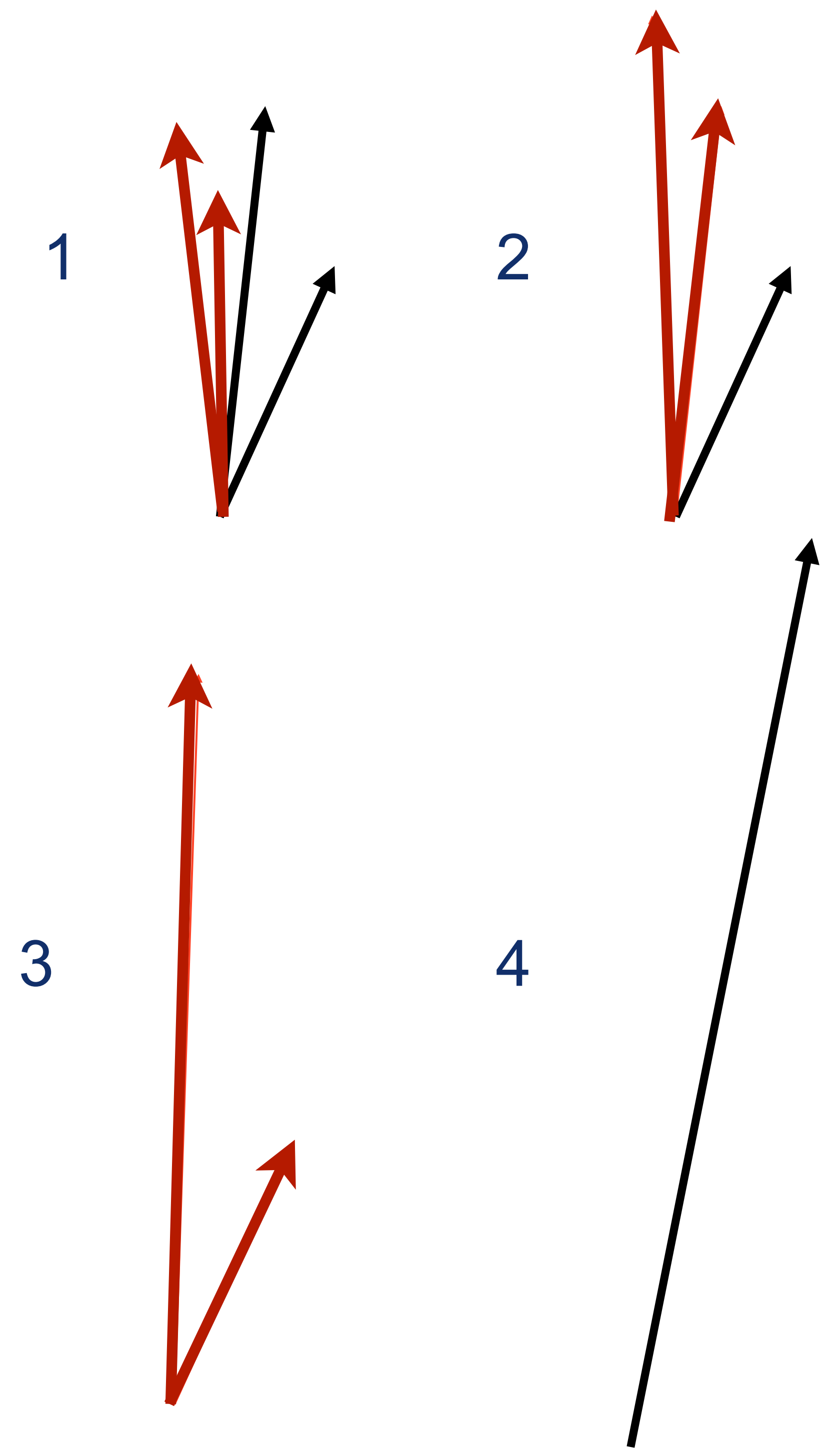
- Like "QCD played backwards!"
- Pairwise examination of input 4-vectors
- Calculate d_{ij}

$$d_{ij} = min(k_{ti}^{n}, k_{tj}^{n})\Delta R_{ij}^{2}/R^{2}$$

- Also find the "beam distance"

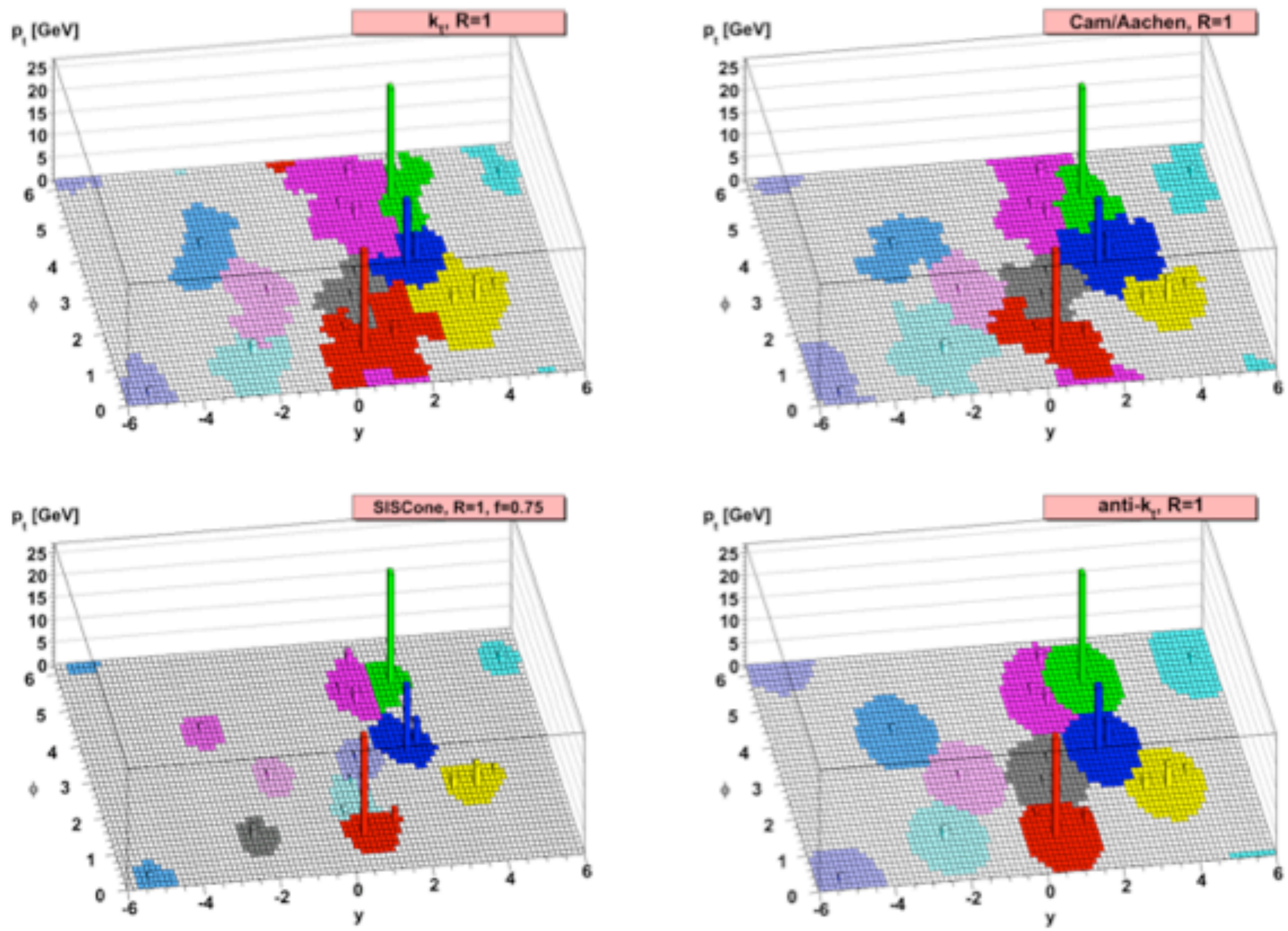
$$d_{iB} = k_{T,i}^{n}$$

- Find min of all d_{ij} and d_{iB}
  - If min is a dij, merge and iterate
  - If min is a d_{iB}, classify as a final jet
- Continue until list is exhausted



Sequential Clustering Algorithms

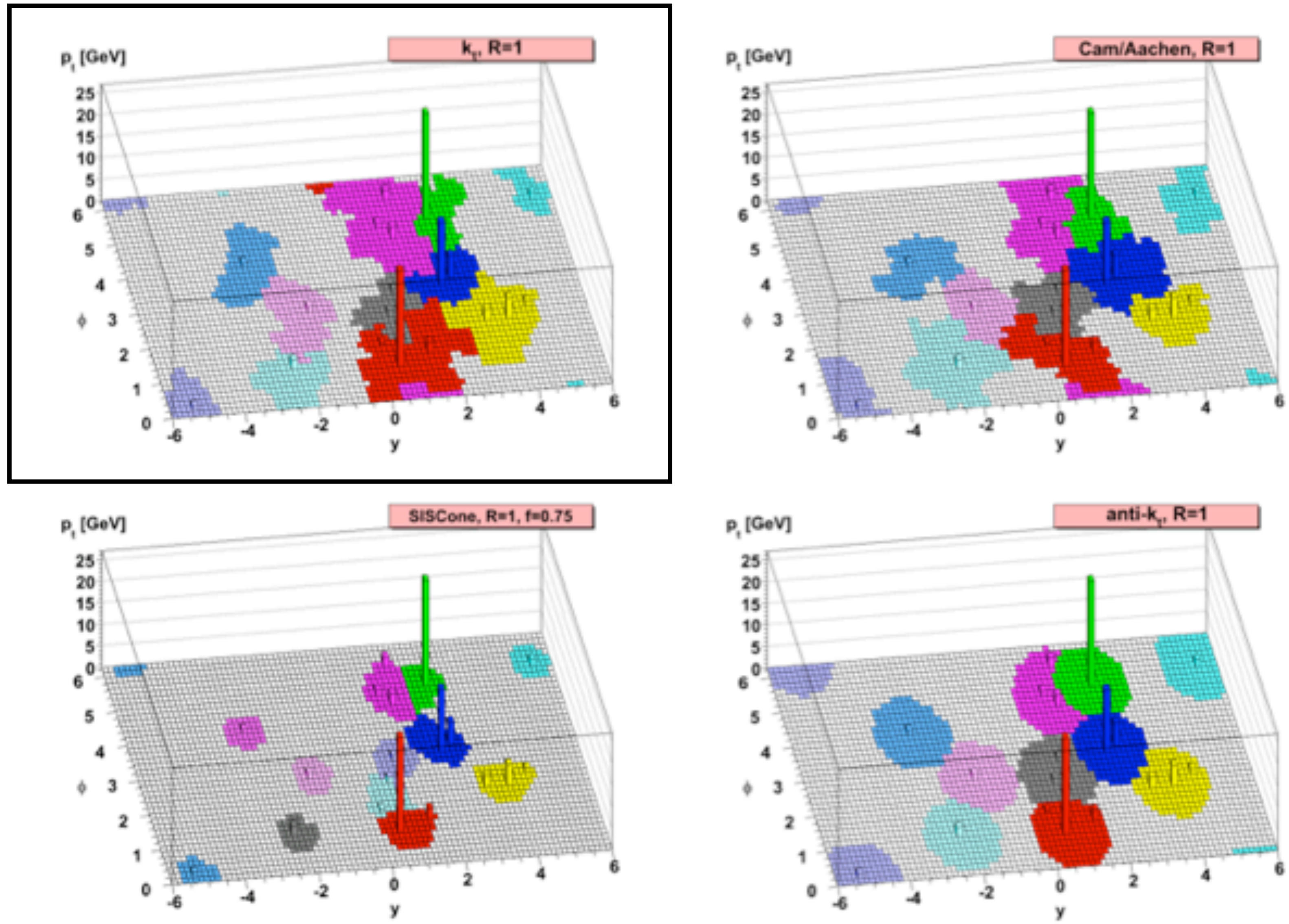
- Properties depend on sequence
  - N = 2:
    - "kT"
  - N = 0 :
    - "Cambridge-Aachen" (CA)
  - N = -2:
    - "anti-kT"



arXiv:0802.1189v2 [hep-ph]
Cacciari, Salam, Soyez

Sequential Clustering Algorithms

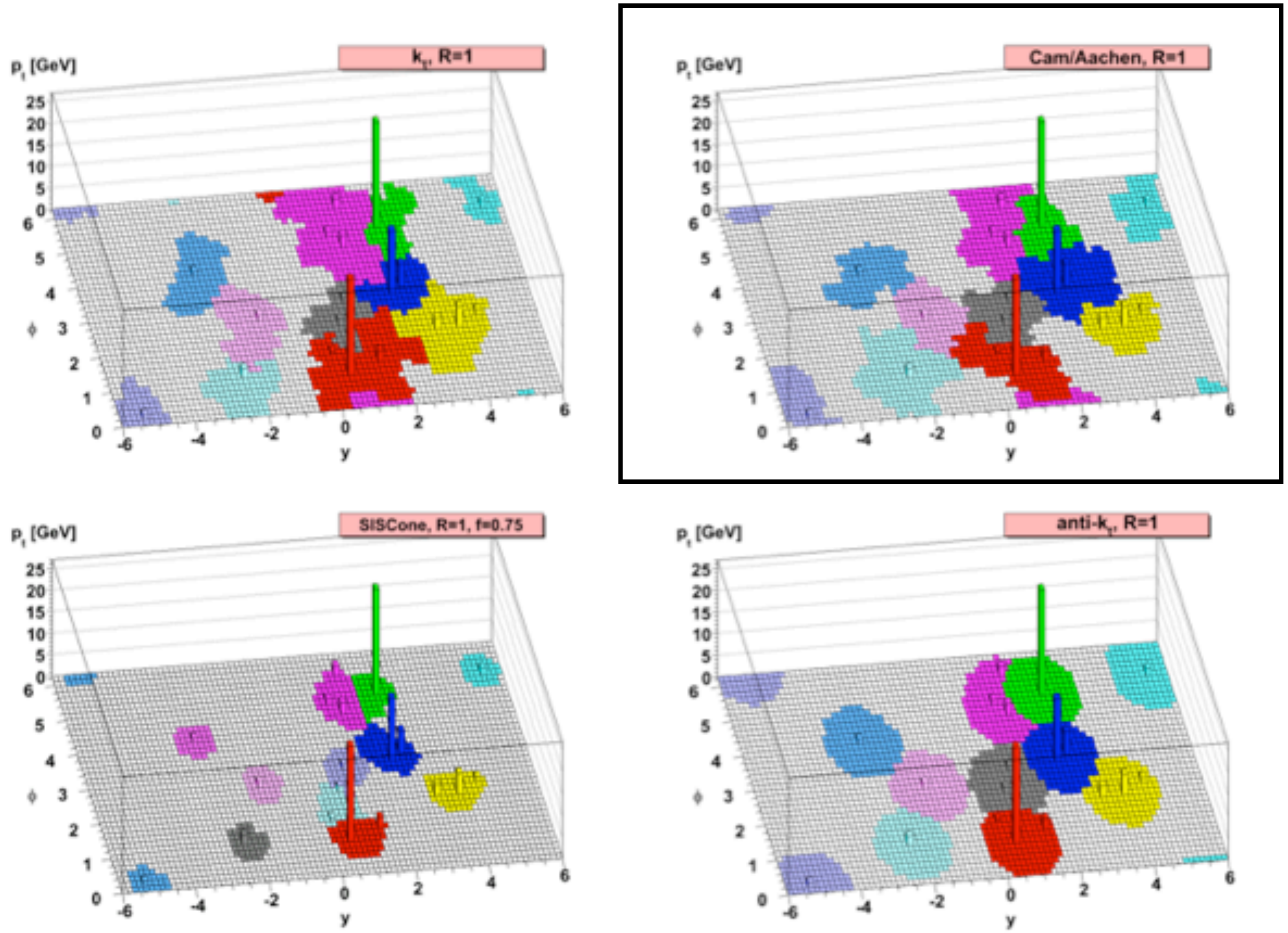
- Properties depend or sequence
  - N = 2:
    - "kT"
  - N = 0 :
    - "Cambridge-Aachen" (CA)
  - N = -2:
    - "anti-kT"
      - "QCD in reverse"
      - Clusters soft particles first
      - Good for low-pt jets
        - Good for jet area computation



arXiv:0802.1189v2 [hep-ph]
Cacciari, Salam, Soyez

Sequential Clustering Algorithms

- Properties depend on sequence
  - N = 2:
    - "kT"
  - N = 0 :
    - "Cambridge-Aachen" (CA)
  - N = -2:
    - "anti-kT"

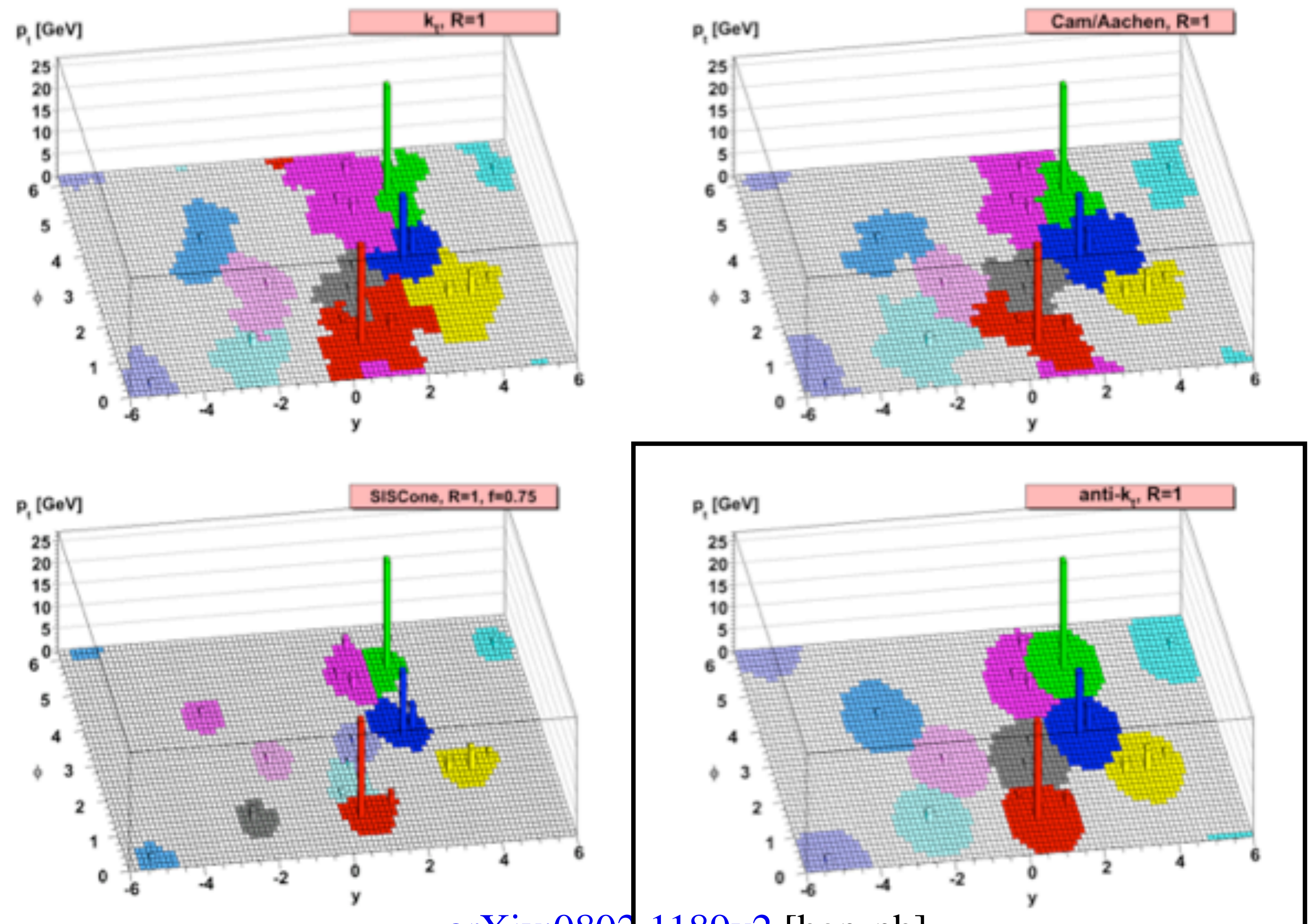


arXiv:0802.1189v2 [hep-ph]
Cacciari, Salam, Soyez

- Angular information only
- Good for finding substructure!

Sequential Clustering Algorithms

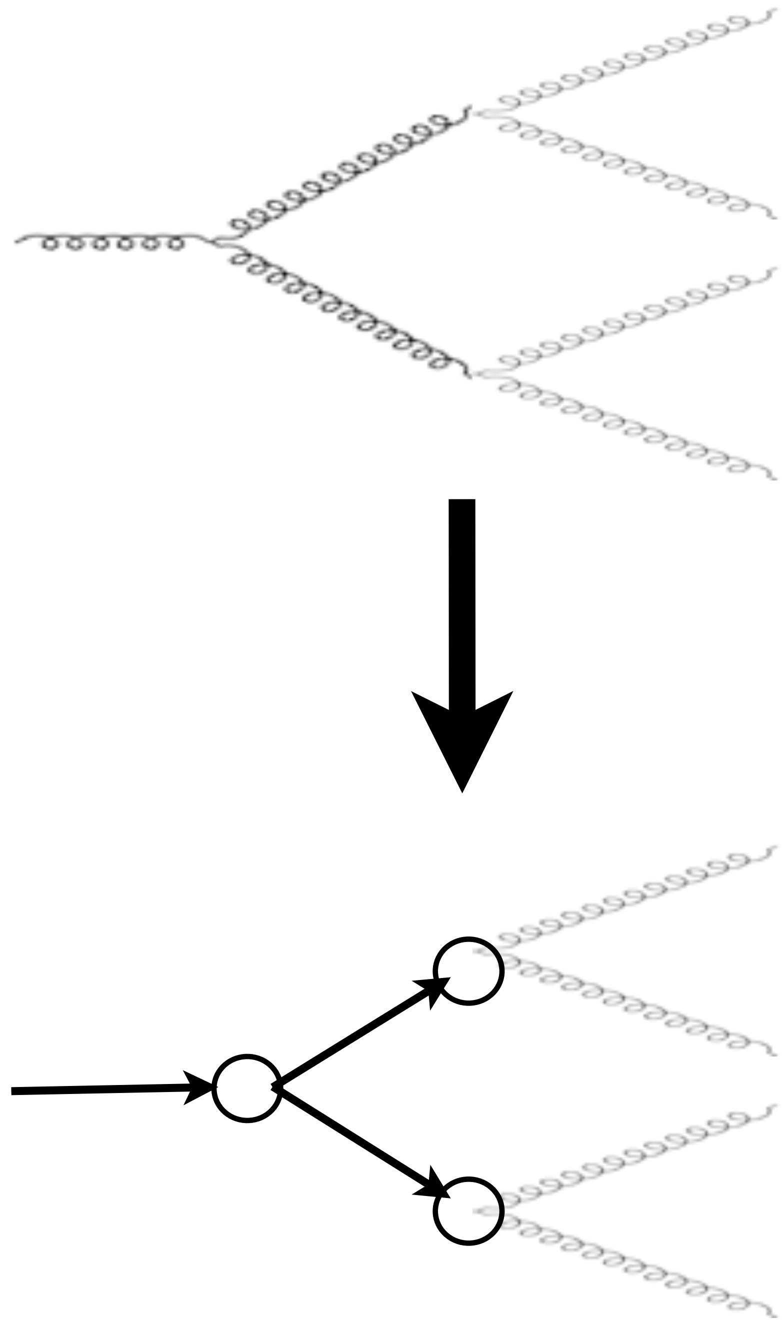
- Properties depend or sequence
  - N = 2:
    - "kT"
  - N = 0 :
    - "Cambridge-Aachen" (CA)
  - N = -2:
    - "anti-kT"



arXiv:0802.1189v2 [hep-ph]
Cacciari, Salam, Soyez

- Clusters hard particles first
- Idealized cone algorithm
- Best for jet counting

- Now that we understand how to construct a jet, we're ready to take a look at massive ones



Jet Substructure

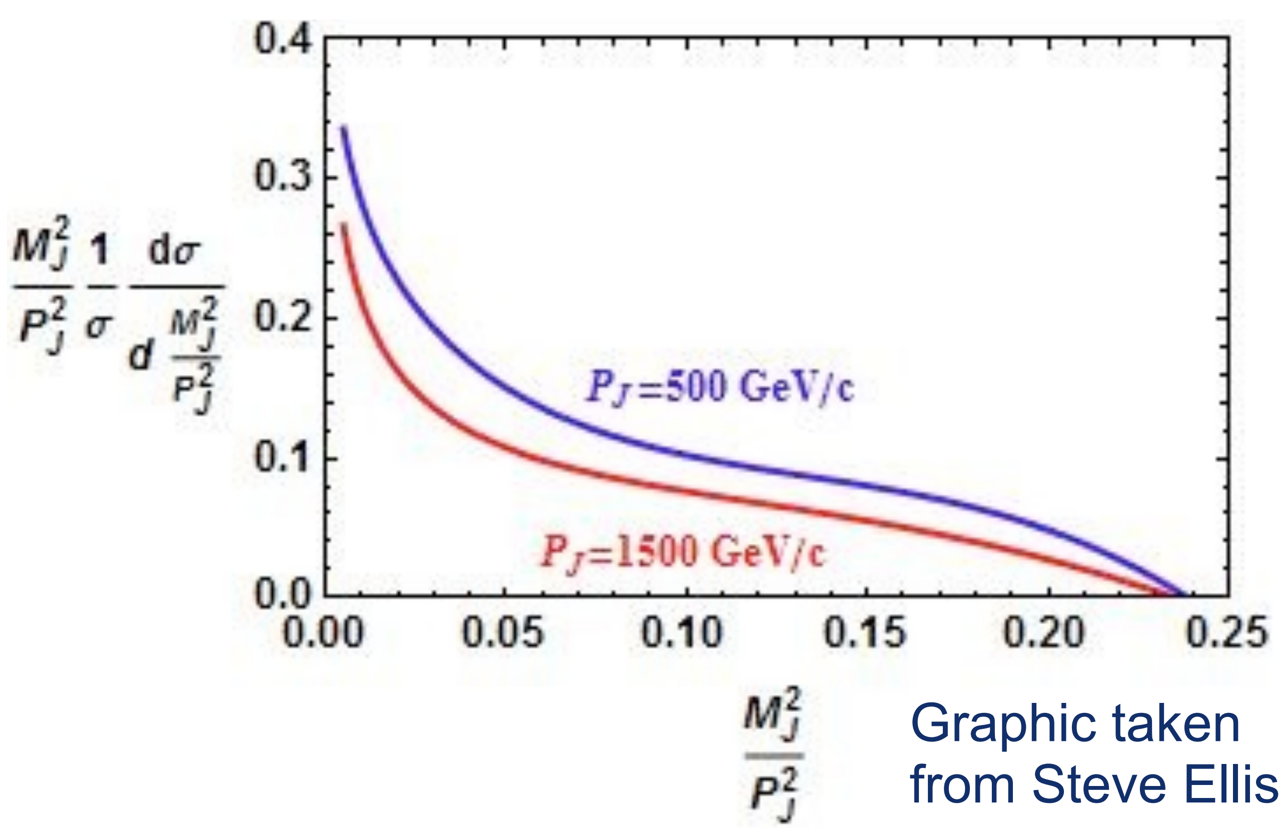
- What's the typical mass scale of QCD?
  - See e.g. Ellis et al (arXiv:0712.2447v1)

Finite-size effects from cutoff

Scales ~linearly with momentum

Log. divergence at low mass

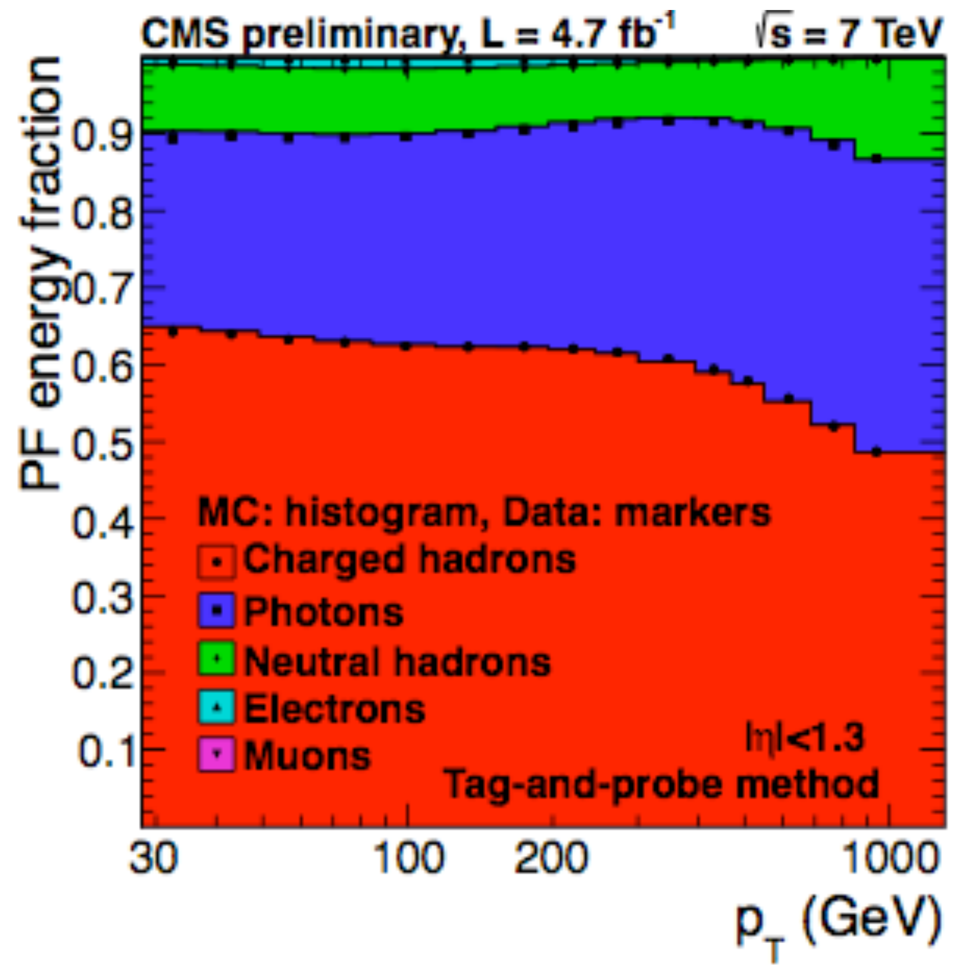
$$\langle M_J^2 \rangle_{NLO} \simeq \overline{C} \left(\frac{p_J}{\sqrt{s}} \right) \alpha_s \left(\frac{p_J}{2} \right) p_J^2 R^2,$$



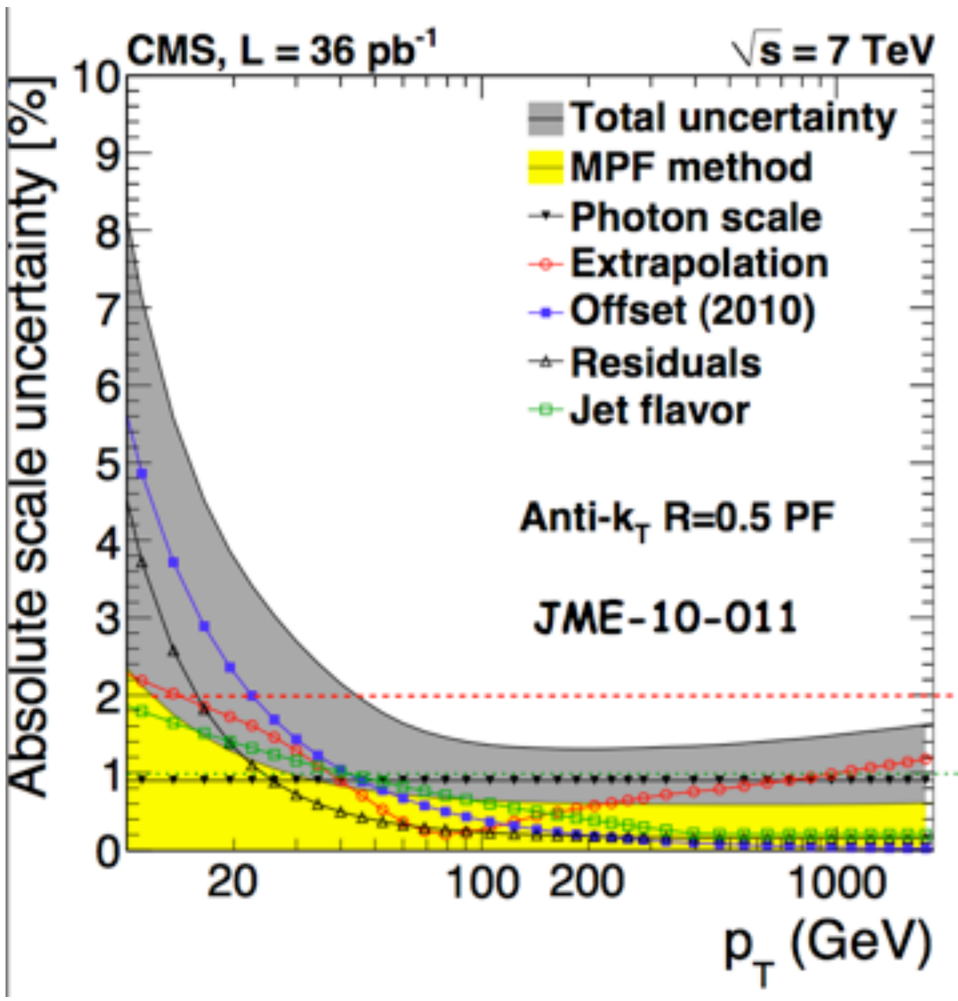
Graphic taken from Steve Ellis

Experimental Overview

- Simulation of PF composition of jets is astonishingly good!

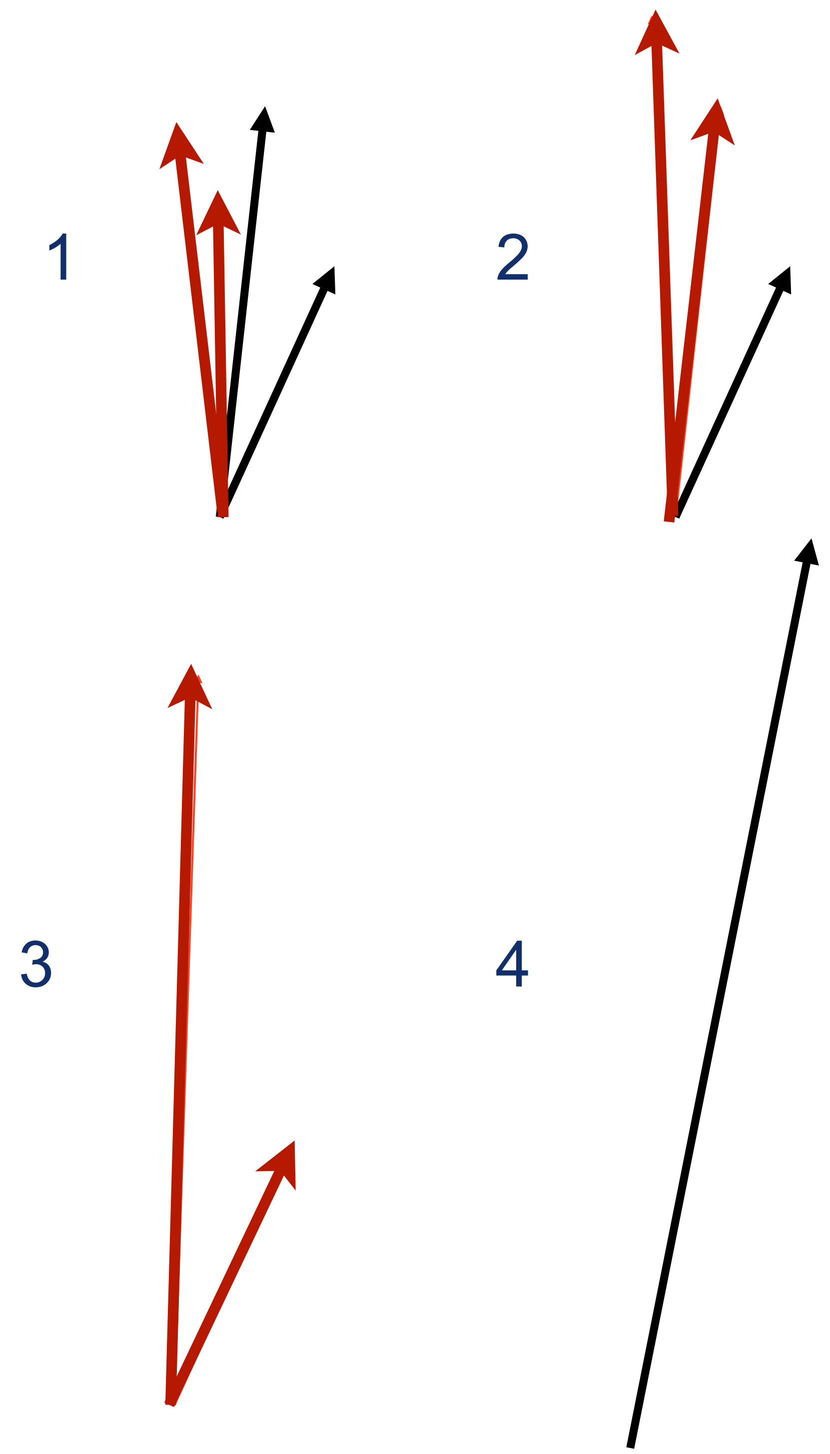


- Uncertainty on JES is ~1-4% depending on pt



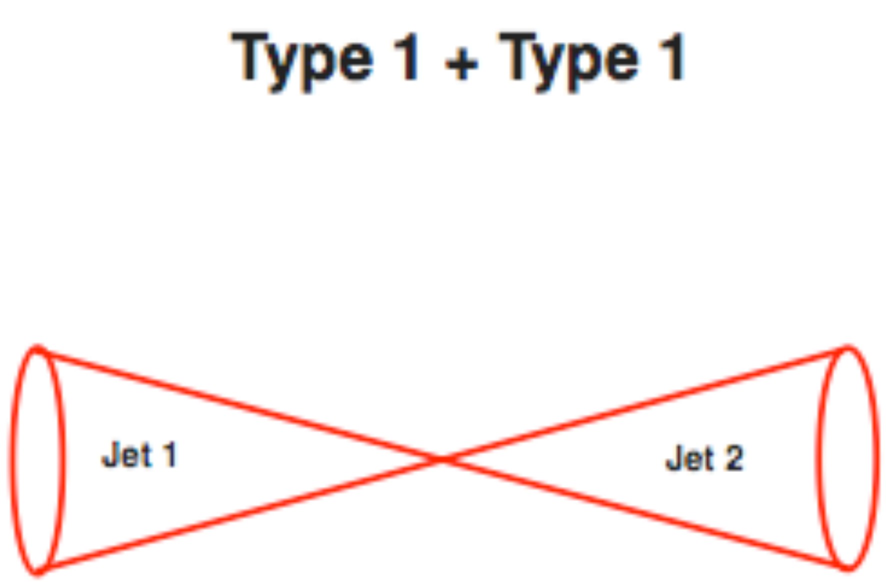
Jet Substructure

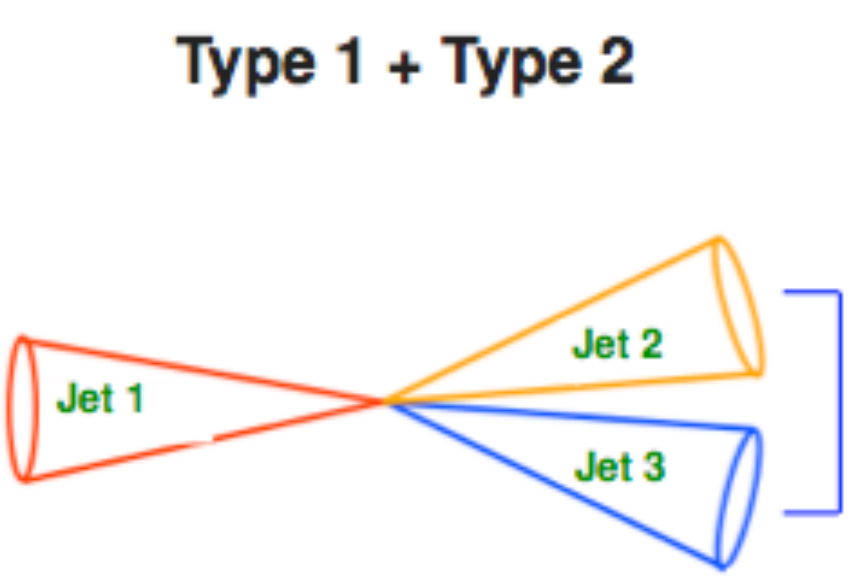
- Examine the cluster sequence "in reverse!"
  - 4 -> 3 -> 2 -> 1
- Around massive particles, clustering will behave differently than general QCD
- Groom away unwanted bits
- You're left with the interesting parts



Analysis Strategy

- Assume m_{ttbar} > 4x m_{top}
  - Creates “boosted” final state
  - Hemispheric topology
- Classify hemispheres by number of “total” jets
  - “Type 1” : 1 fully merged jet
  - “Type 2” : 1 partially merged jet, 1 non-merged jet
- Look at “Type 1 + 1” and “Type 1 + 2” events
- Apply “cascading selection”
  - No event gets left behind!

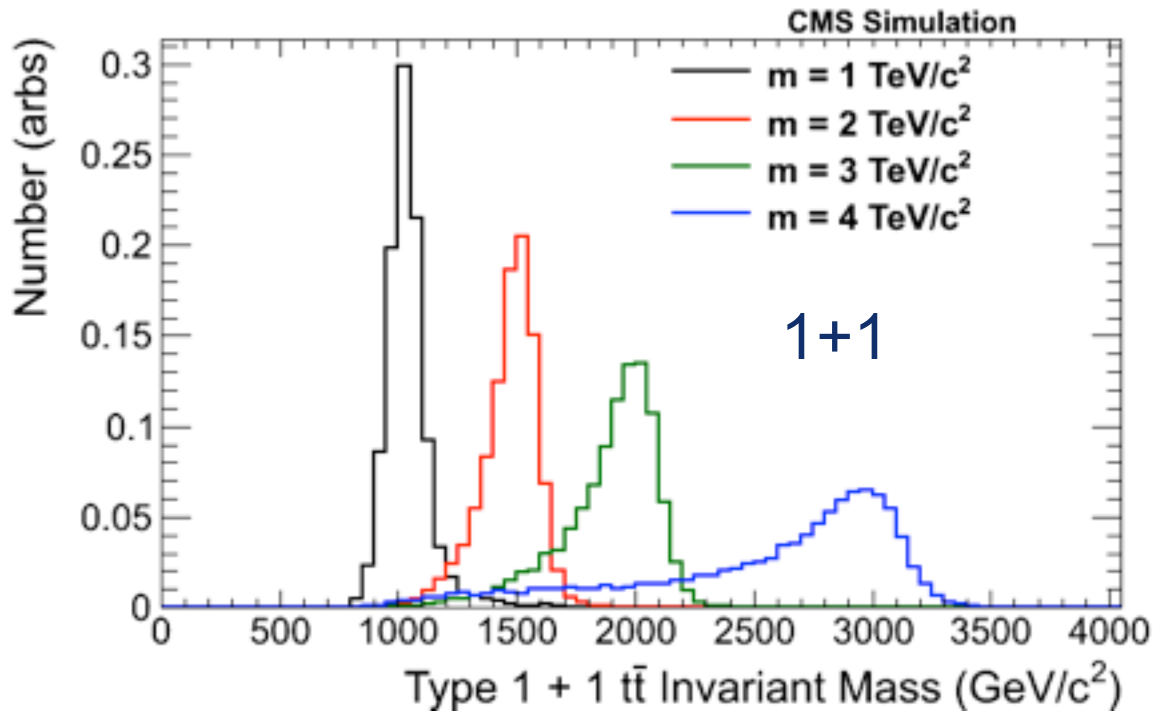


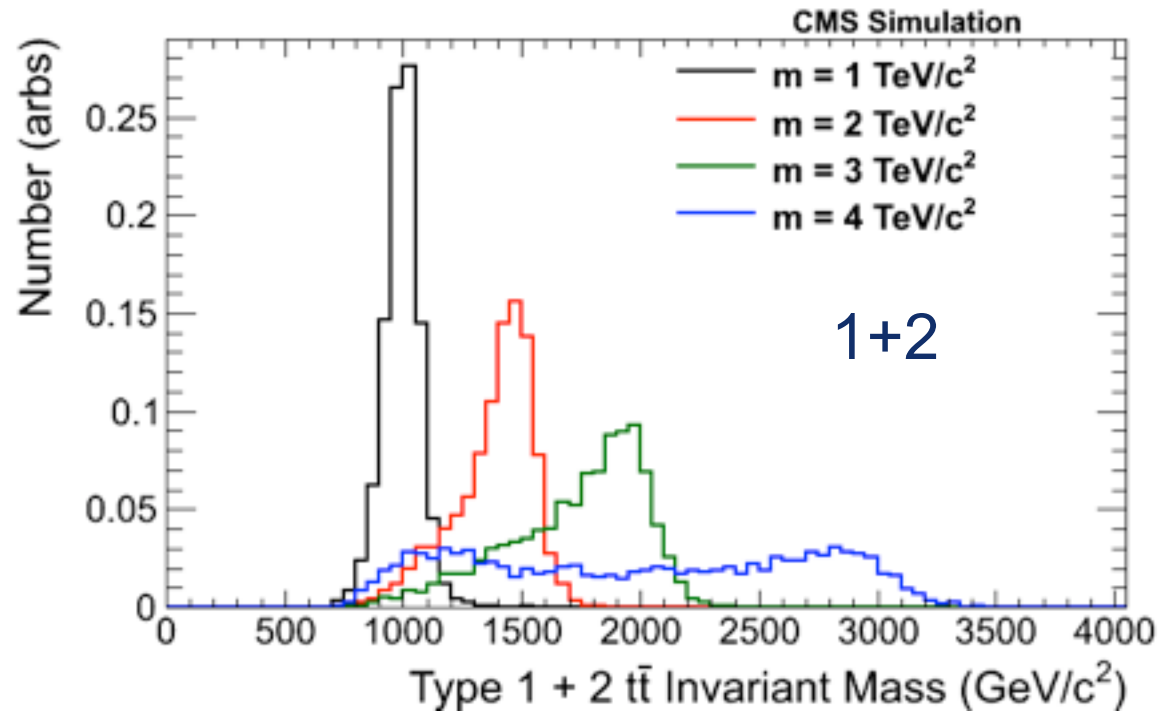


Statistical Treatment

- Shape uncertainties:
  - JES
    - 4% (standard JES), 3% (substructure jets), 1% (subjet JES relative to full JES)
  - Trigger
    - Half-difference between trigger-weighted MC and unweighted MC
  - Background
    - Half-difference between mass-modified-mistags and non-mass-modified
- Rate uncertainties:
  - Data-to-MC efficiency scale factor (6% in double-tagged sample)
  - Luminosity (2.2%)
  - TTbar background theoretical uncertainties (50% renormalization/factorization scale)

Signal





- Use samples of Z’ from Madgraph with width = 1% of mass
- Also check width=10%, and a Pythia8 KK gluon sample (width ~20%)
- Weight by trigger efficiency (plateaus around 1.5 TeV)
- Correct for efficiency with data-to-MC scale factor (97 +- 3%)

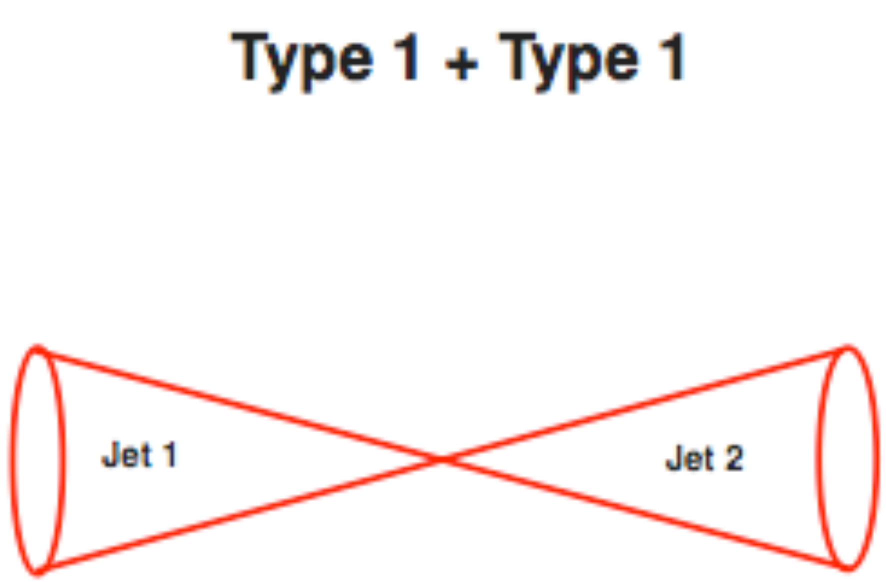
Signal

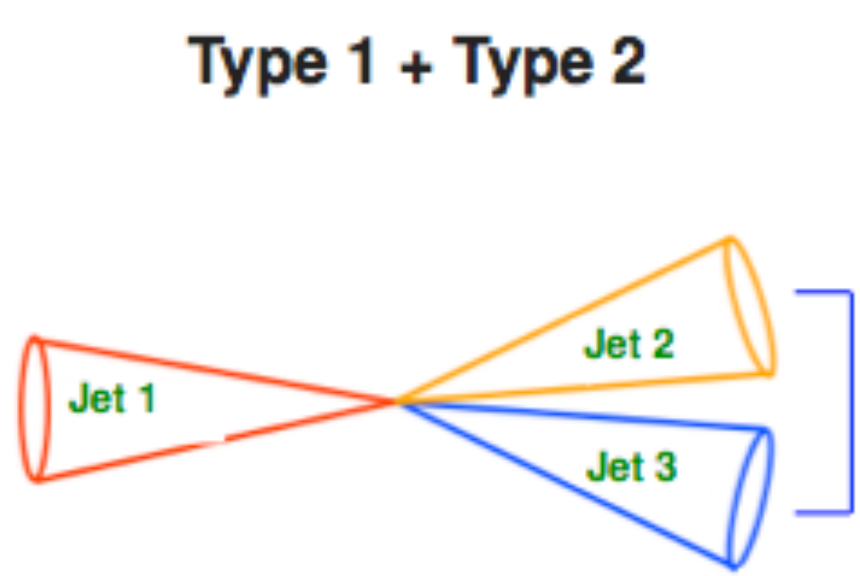
Z' mass	Signal window	ϵ_{MC}	ϵ_{trig}	ϵ_{total}	$\sigma_{\epsilon_{total}}$	B	σ_B	N_{obs}
Type 1 + 1								
1000	900-1100	1.2%	79.2%	0.8%	0.2%	501.2	26.2	506
1500	1200-1600	7.1%	98.6%	5.7%	0.6%	766.1	36.8	774
2000	1300-2400	10.8%	99.6%	8.7%	0.8%	808.7	39.8	809
2500	1700-2800	8.5%	99.7%	6.9%	0.6%	238.6	9.7	222
3000	2000-3300	6.9%	99.8%	5.6%	0.6%	97.0	4.2	79
Type 1 + 2								
1000	900-1100	2.1%	71.9%	1.2%	0.4%	1334.5	40.6	1383
1500	1200-1600	2.6%	91.3%	1.9%	0.2%	911.8	51.2	900
2000	1300-2400	2.2%	94.4%	1.7%	0.2%	841.9	39.4	841

- Shown in mass windows for indication of relative importance of systematic effects
- Overall efficiencies in the 3-12% range for signal

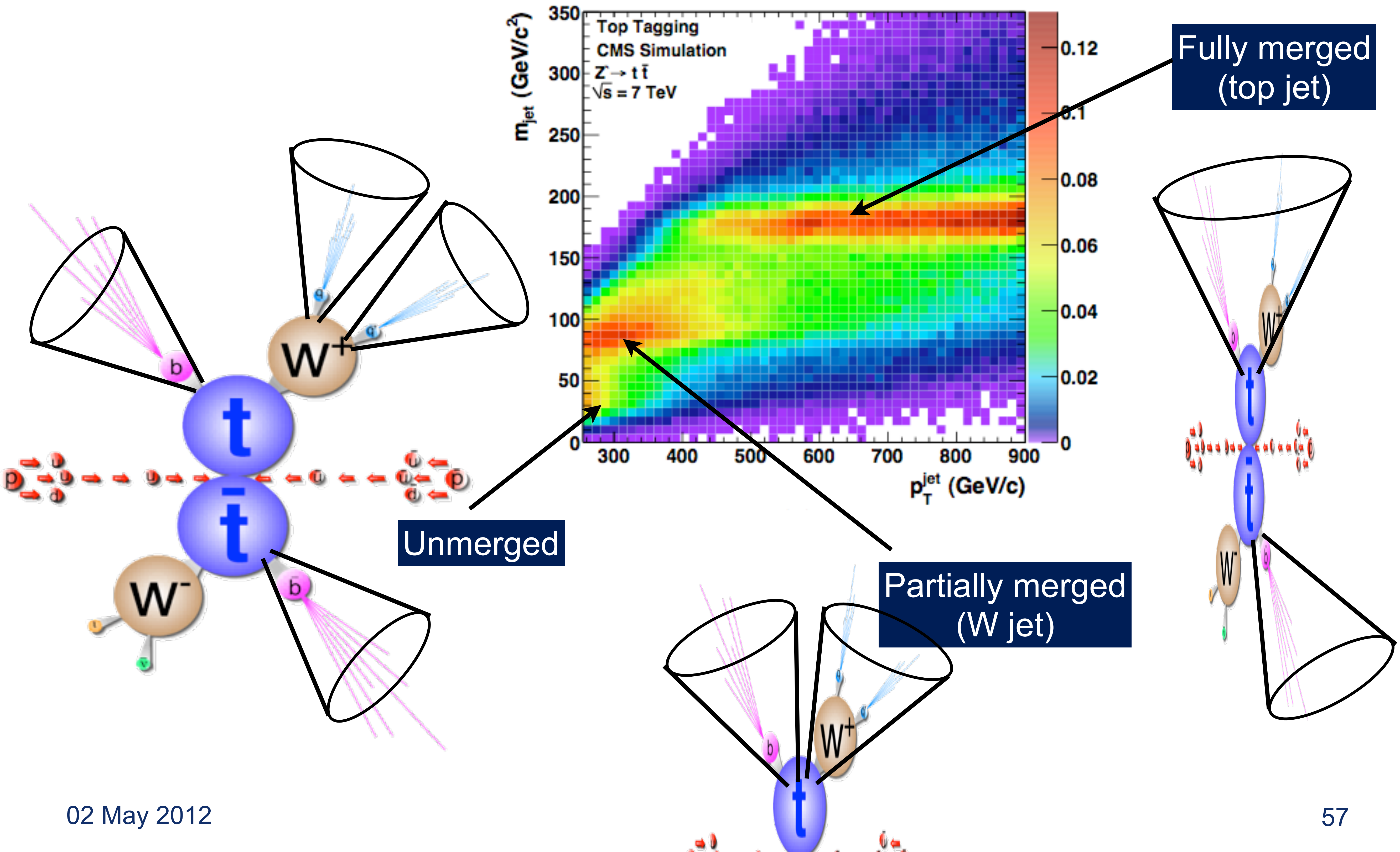
Event Selection

- Type 1 + 1
  - Jet pt > 350 GeV/c
  - Both jets satisfy “top tagger” requirements
- Type 1 + 2
  - Veto 1 + 1 (<1% overlap)
  - Jet pt > 350, 200, 30 GeV/c
  - Jet 1 (type 1 jet) satisfies “top tagger” requirements
  - Jet 2 (type 2 hemisphere) satisfies “W tagger” requirements
  - Jet 3 (type 2 hemisphere) has no requirements





Motivation


Top Tagging
CMS Simulation
Z' → t t̄
√s = 7 TeV
m_jet (GeV/c²)
p_T^jet (GeV/c)
0
50
100
150
200
250
300
350
300
400
500
600
700
800
900
0
0.02
0.04
0.06
0.08
0.1
0.12
Fully merged
(top jet)
Partially merged
(W jet)
Unmerged
t
t̄
W+
W-
b
b̄