Jet Substructure in the Pre-ILC Era

Brock Tweedie
Boston University
18 March 2011

We're Making Progress

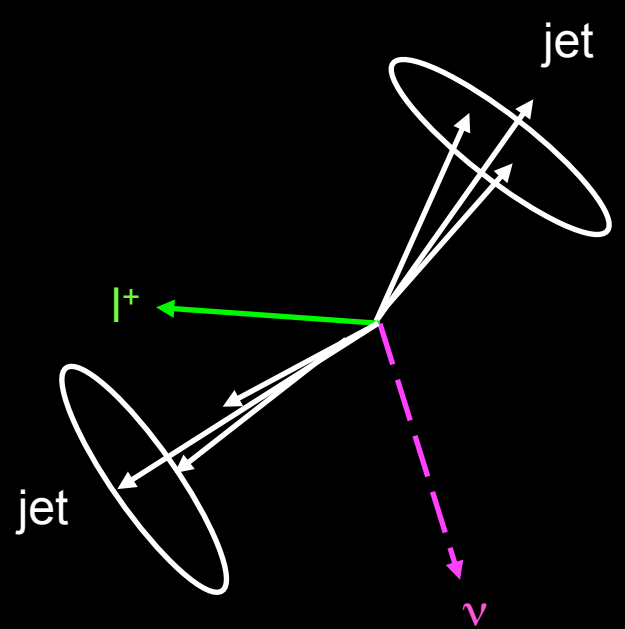
- Everything "heavy" becomes "light" eventually
  - 1970's: Taus, c-quarks, and b-quarks freshly discovered, exploring the multi-GeV mass scale!
  - 1980's: Tau-jets, c-jets, and b-jets are secondary signals of interesting new things
    - jet substructure with small # hadrons

We're Making Progress

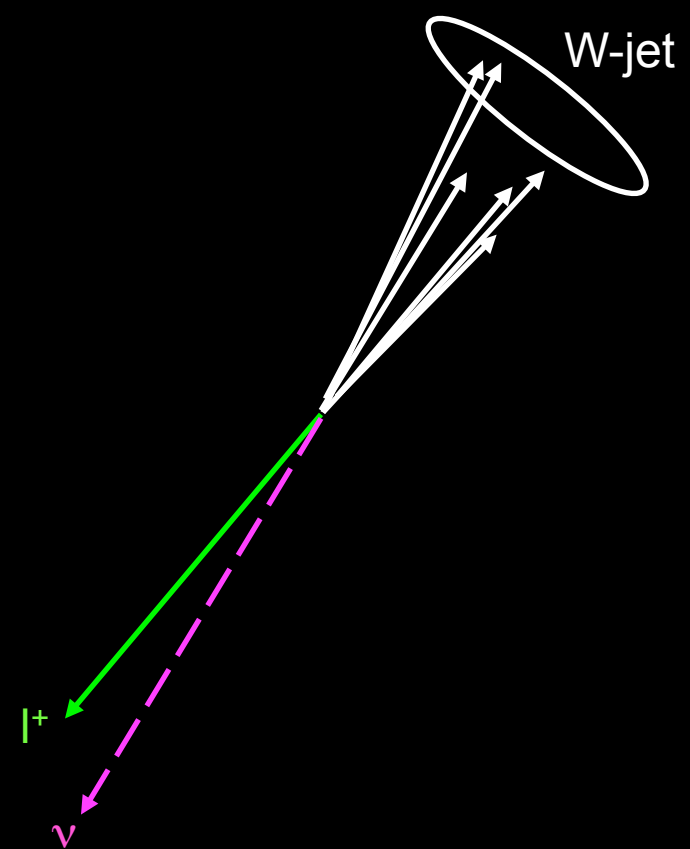
- Everything "heavy" becomes "light" eventually
  - 1980/90's: W-bosons, Z-bosons, and top-quarks freshly discovered, exploring the ~100-GeV mass scale!
  - 2010's: W-jets, Z-jets, and top-jets (and Higgs-jets if we're lucky) are secondary signals of interesting new things
    - jet substructure with large # hadrons

The New Jets

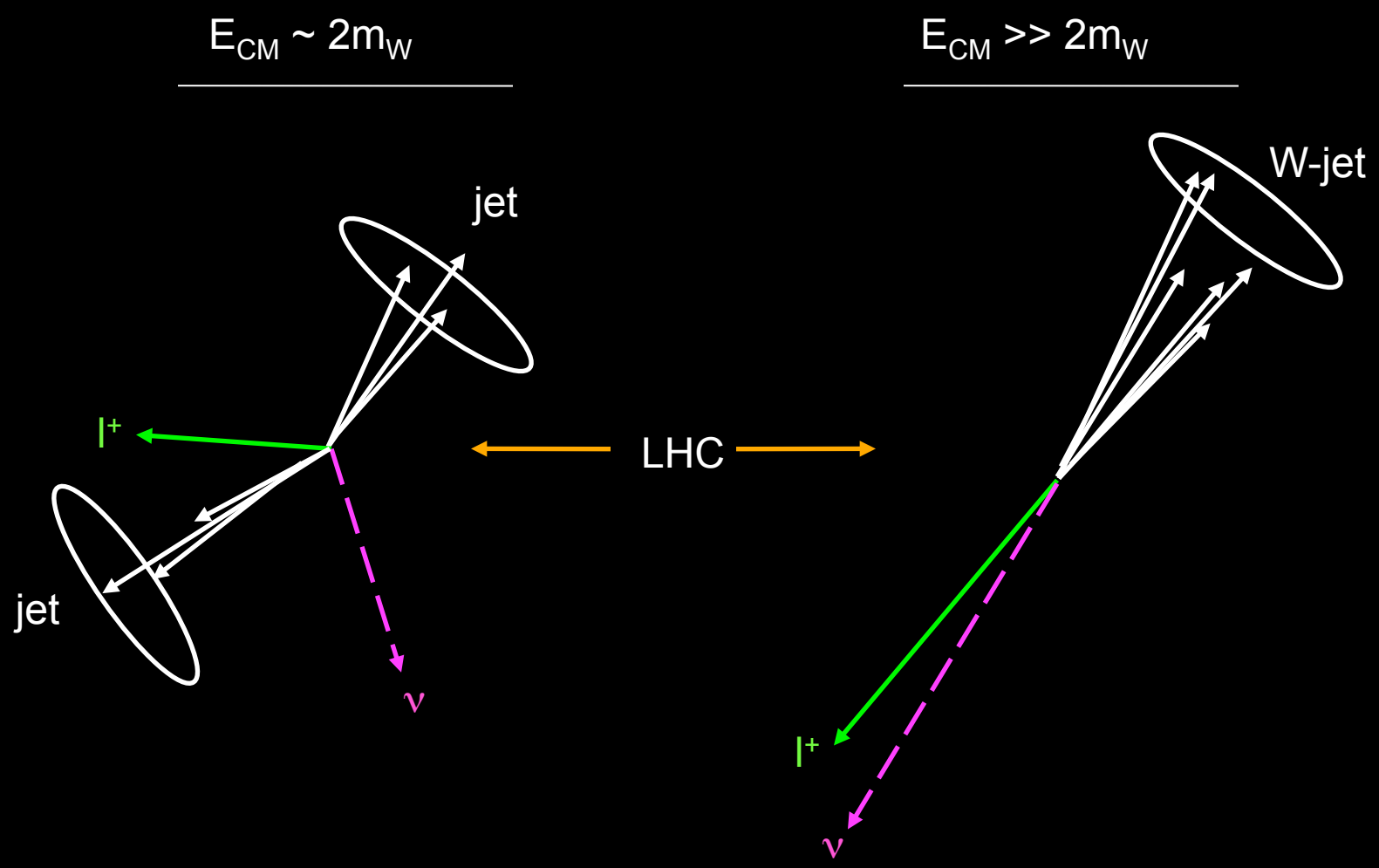
$E_{CM} \sim 2m_W$



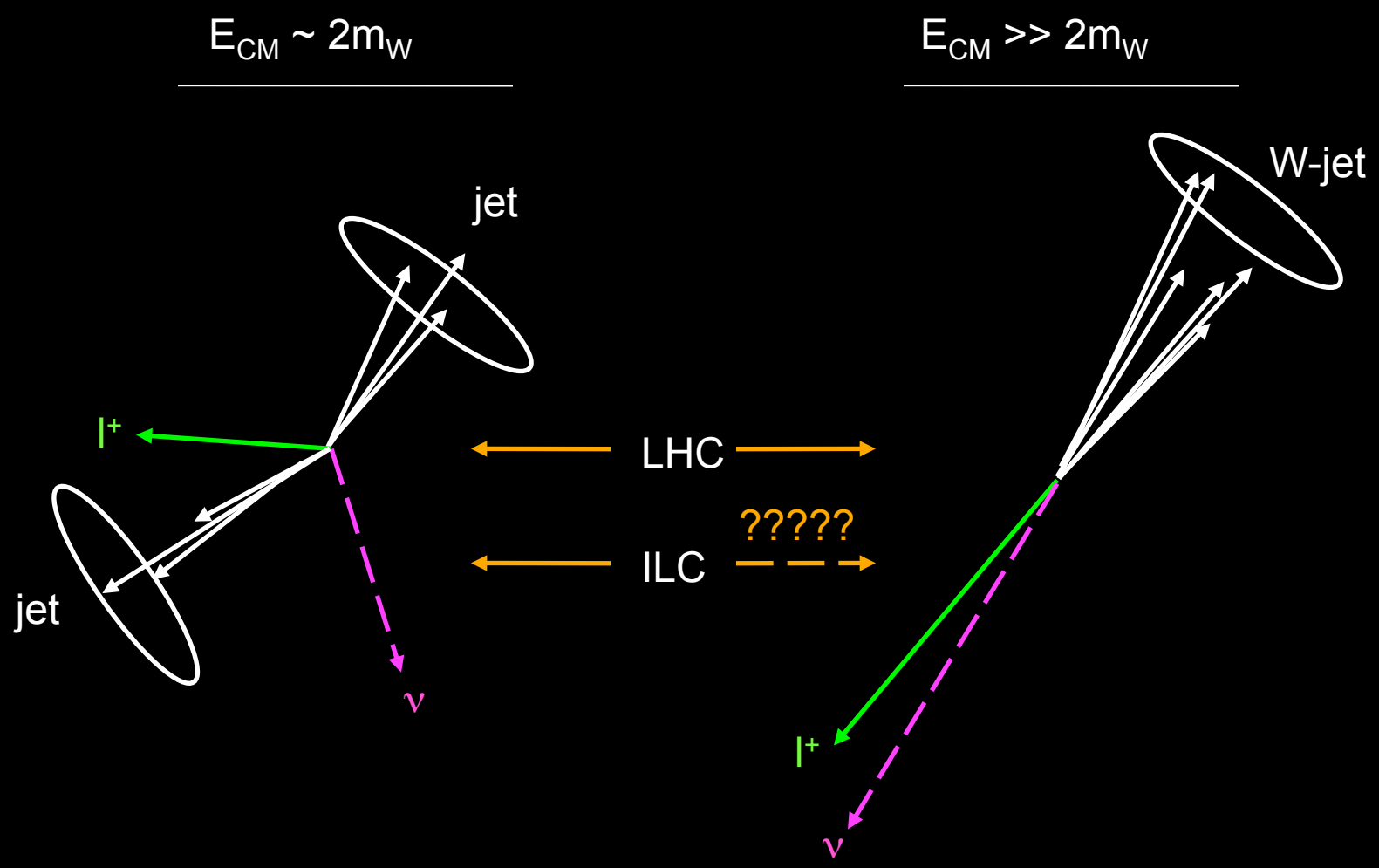
$E_{CM} >> 2m_W$



The New Jets



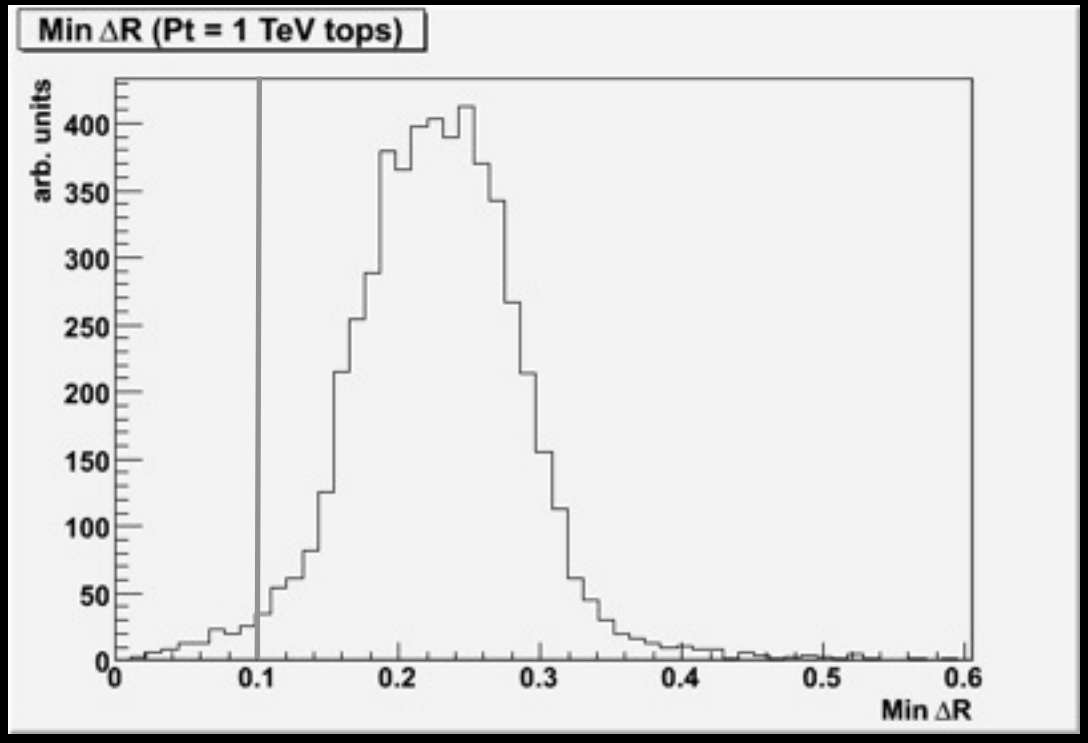
The New Jets

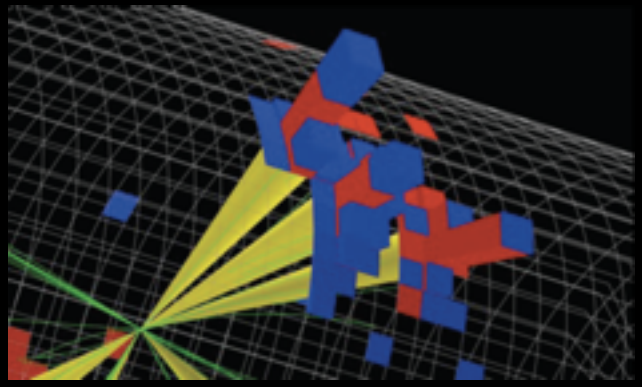


Some Relevant Angular Scales

- W->qq
  - $\Delta R \sim 2m / p_T$
  - $p_T = 400$ GeV => $\Delta R \sim 0.4$
- Typical LHC jet size
  - $\Delta R \sim 0.4$
- HCAL cells
  - $\Delta R \sim 0.1$
- ECAL cells
  - $\Delta R \sim 0.02$
- Tracker
  - $\Delta R \sim 0.001$

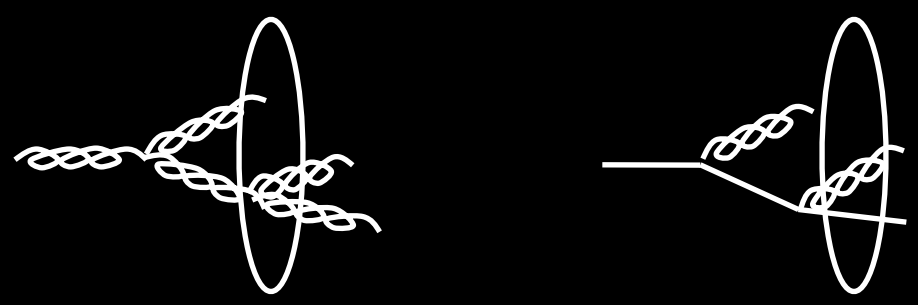
Angular Scales in Top Decay





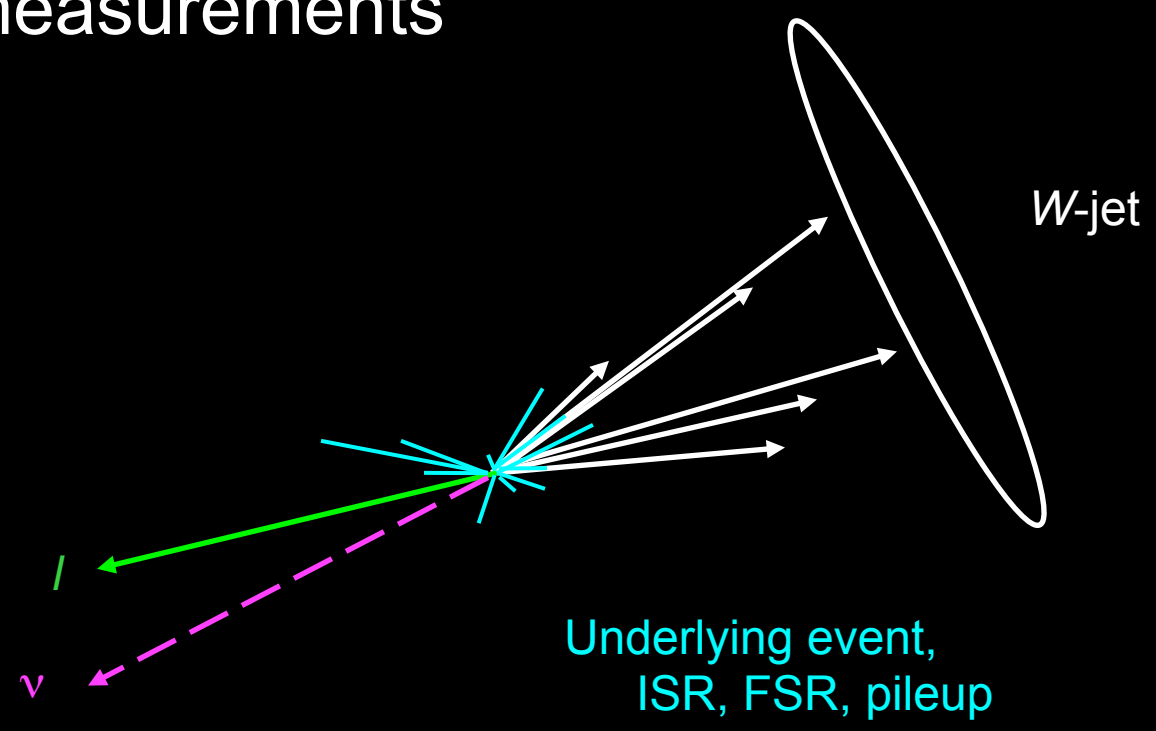
So What?

- QCD is our enemy
  - collinear splittings => easy to get several "hard" partons emitted into a small solid angle
  - jet mass ~ sqrt(α_s)*p_T*R ~ 100 GeV



So What?

- QCD is our enemy
  - All jets are full of junk!
  - Size too large compared to ΔR => messed-up jet-mass measurements



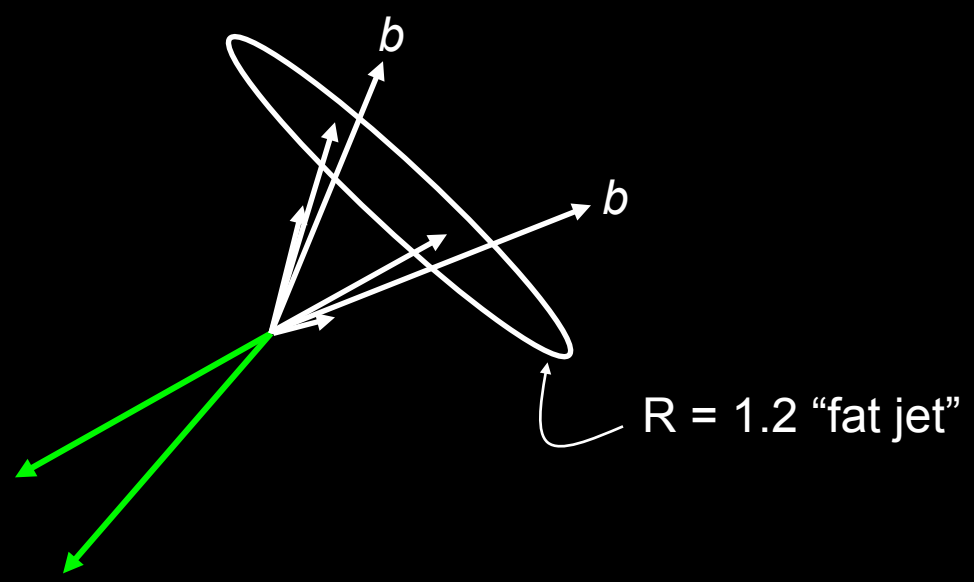
Goals of "All-Hadronic" Substructure

- Figure out the relevant ΔR scales on an event-by-event basis
  - jets -> subjets
- Discrimination against QCD parton splittings
  - multibody kinematics at small angles
- Keep the radiation we want, toss the junk

Brief Early History (2-Body)

- 1994: Seymour
  - 600 GeV Higgs->WW
- 2002: Butterworth Cox Forshaw
  - TeV-scale WW scattering
- 2007: Butterworth Ellis Raklev
  - W-jets from SUSY decays
- 2008: Butterworth Davison Rubin Salam
  - (semi)boosted Higgs->bb in LHC Higgsstrahlung

Butterworth, Davison, Rubin, Salam (BDRS)



W/Z -> leptons/neutrinos

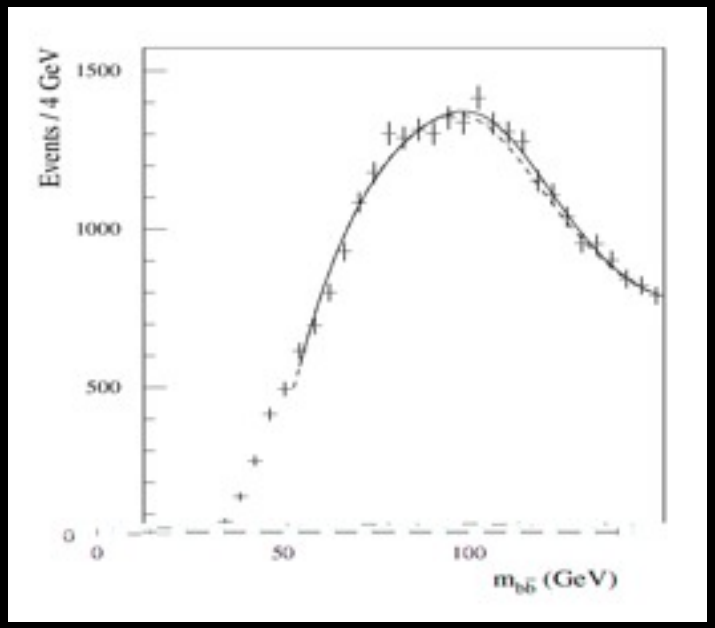
$p_T(V) \sim p_T(h) \sim 200$ GeV ($\Delta R \sim 1.0$)

high-p_T kills backgrounds faster than signal

also: Agrawal, Bowser-Chao, Cheung, Dicus, DPF Conf.1994:488-492

Butterworth, Davison, Rubin, Salam (BDRS)

Wh -> ($l\nu$)(bb)

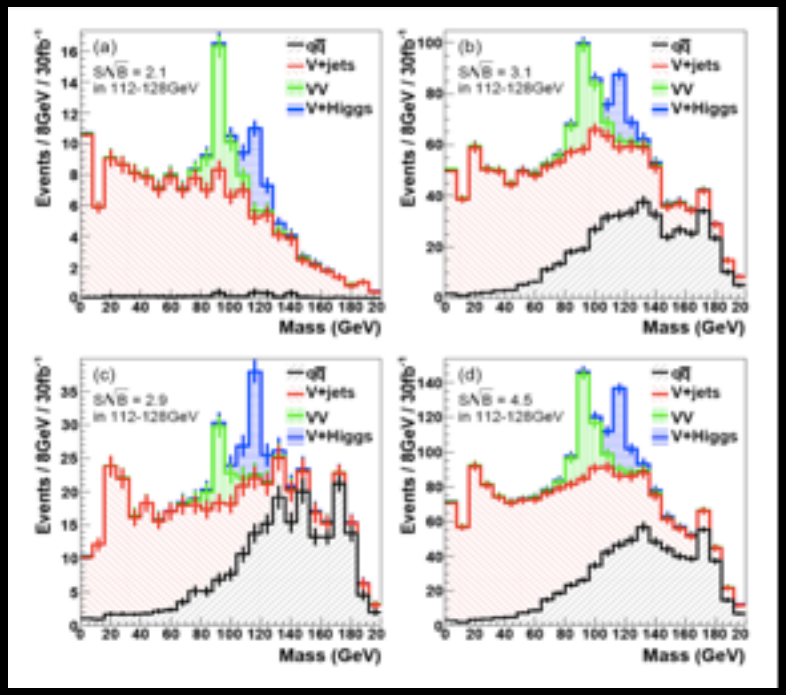


ATLAS TDR

30/fb, m_H = 100 GeV

High p_T

Zh -> (l^+l^-)(bb) *Zh* -> ($\nu\nu$)(bb)

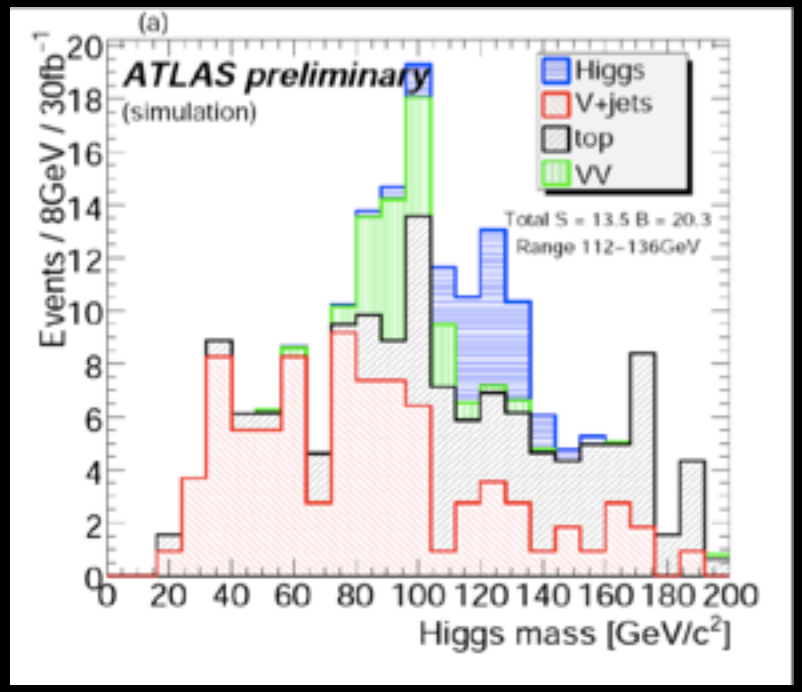


Wh -> ($l\nu$)(bb) combination

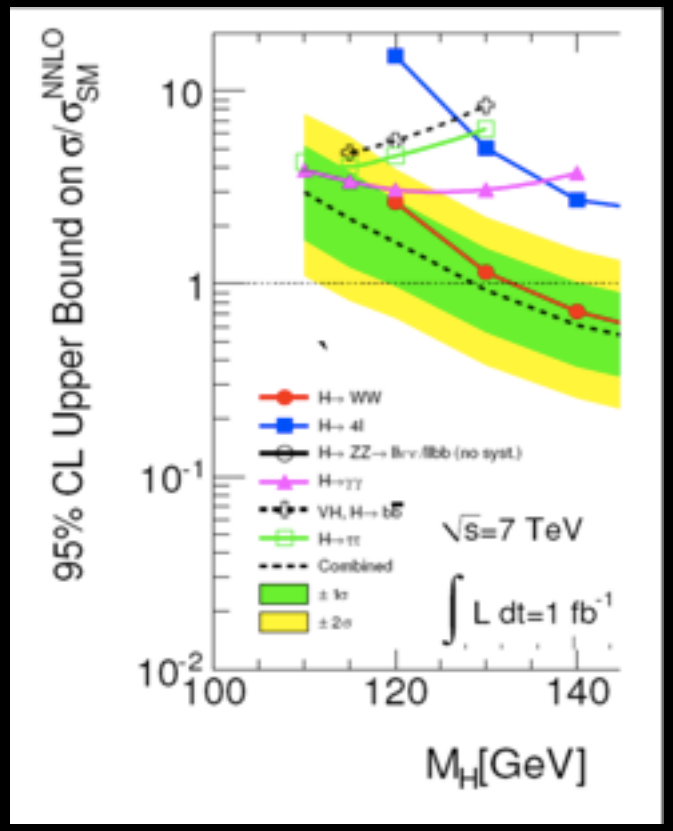
S/sqrt(B) ~ 4.5 at 30/fb LHC14

Butterworth, Davison, Rubin, Salam (BDRS)

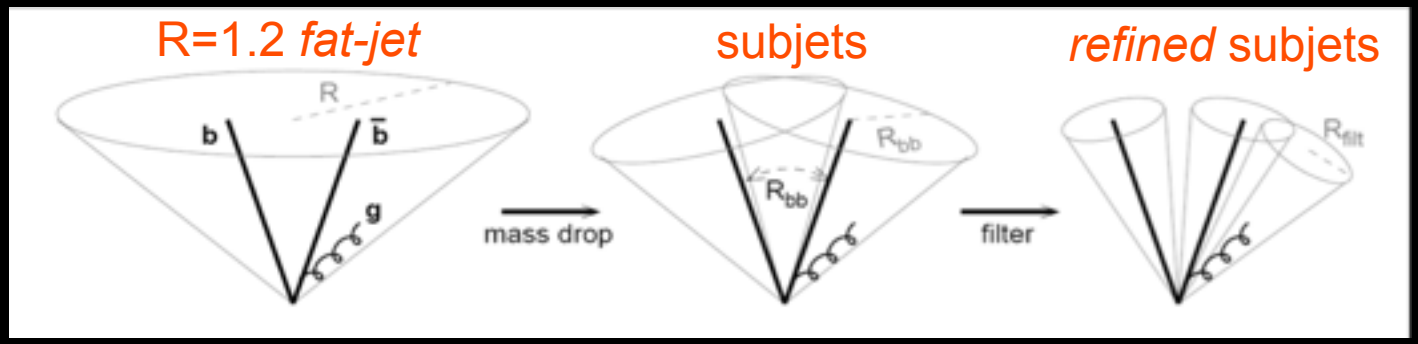
Wh -> (*l*ν)(*bb*)

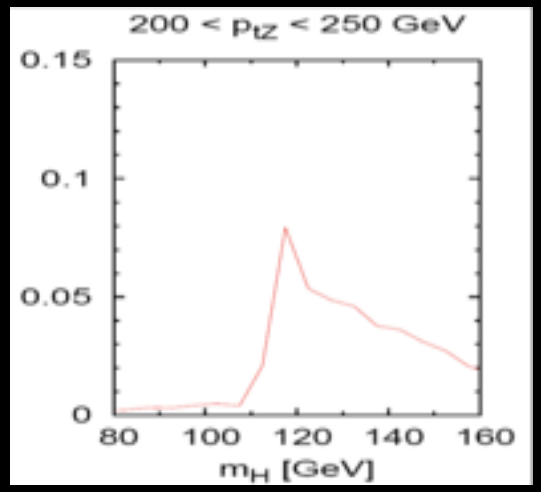


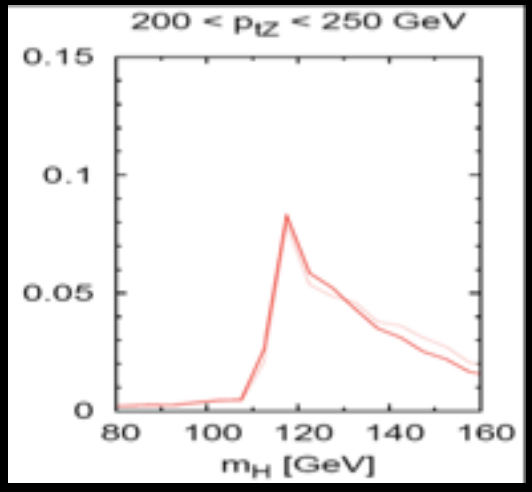
New ATLAS: ~2.6σ

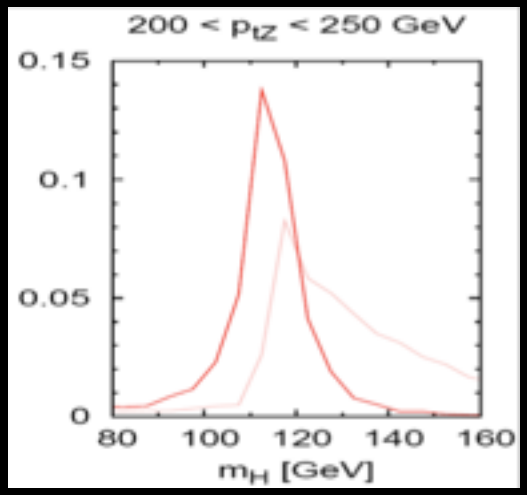


1/fb LHC7 Higgs exclusion



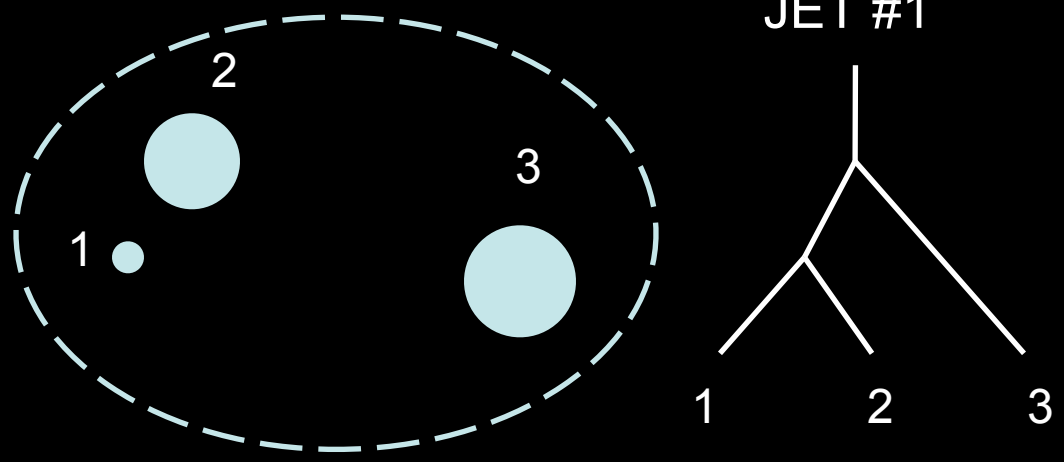


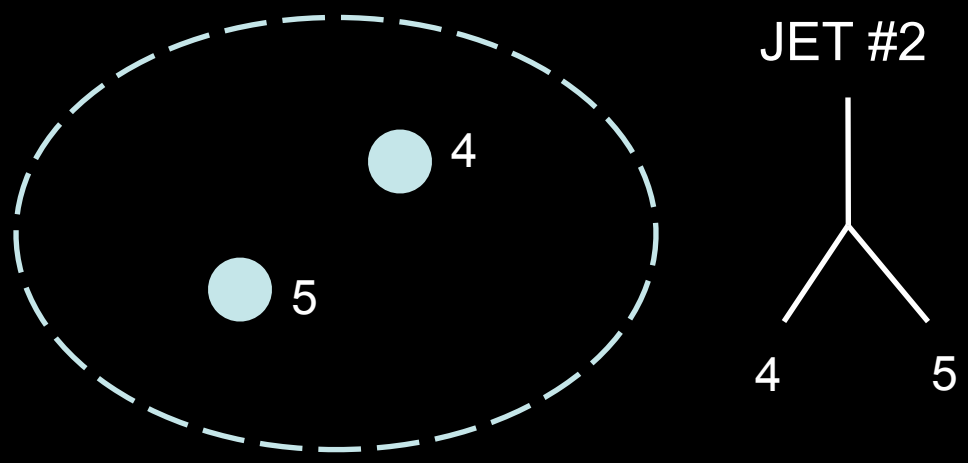




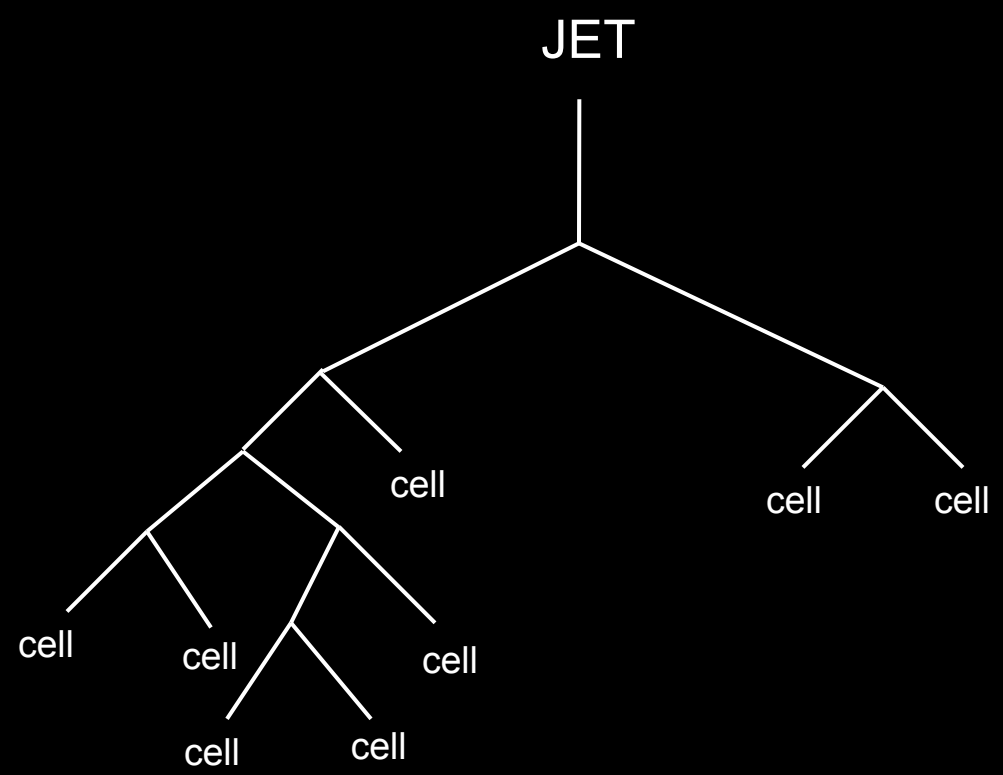
Butterworth, Davison, Rubin, Salam (BDRS)

Cambridge/Aachen Clustering History

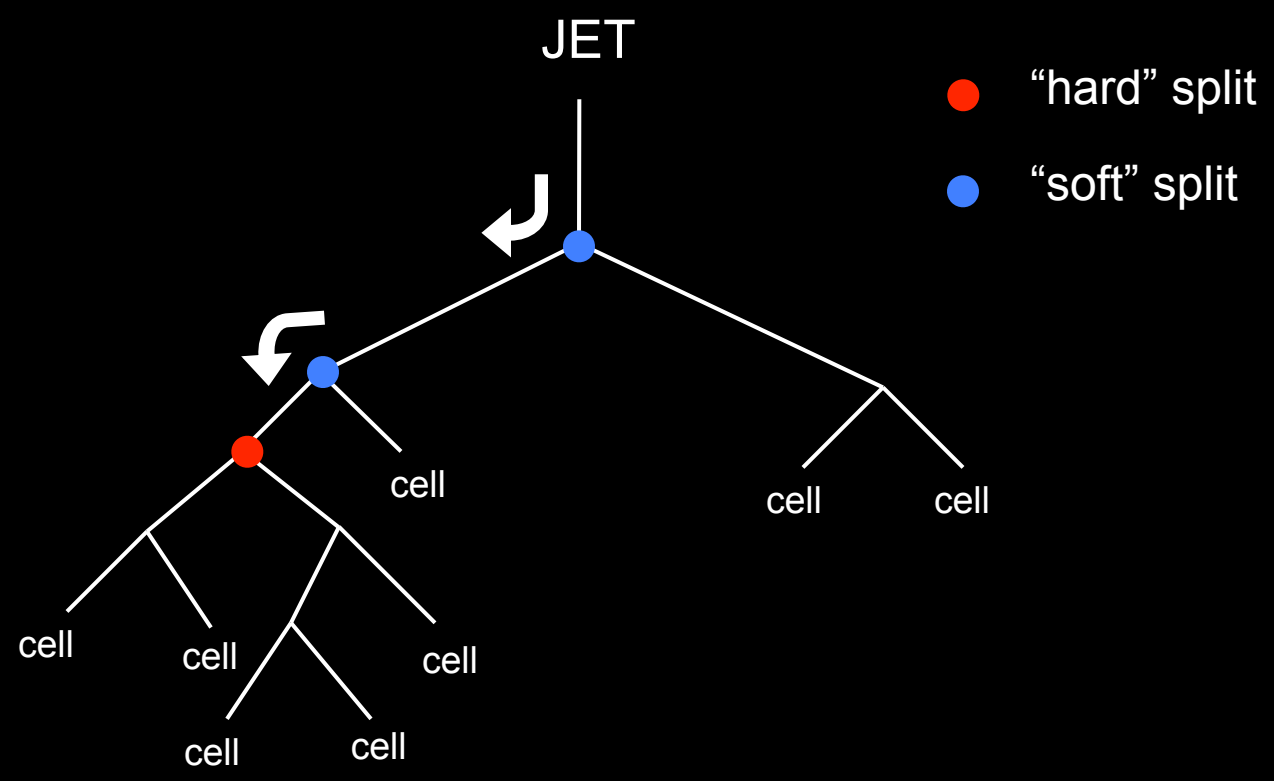




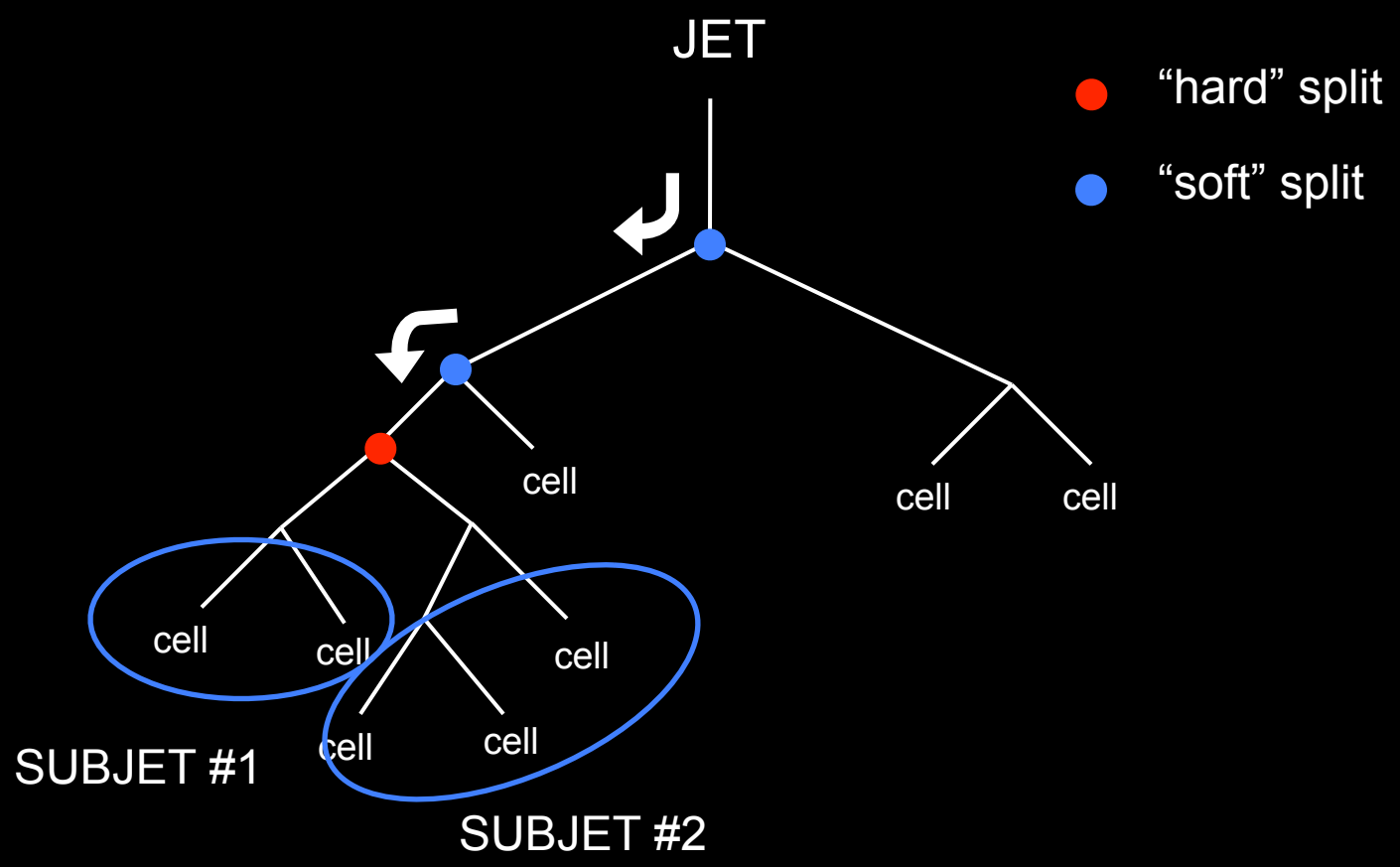
Jet Declustering



Jet Declustering



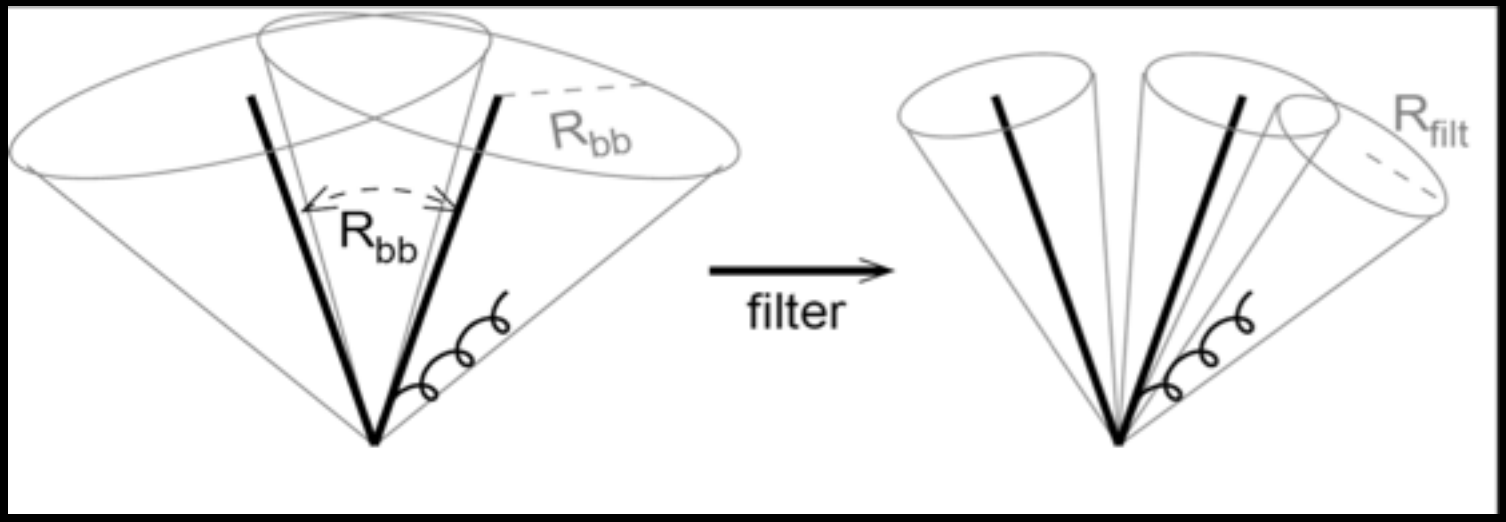
Jet Declustering



BDRS Declustering Criteria

- Mass-drop: $\mathrm{Max}(m_1,m_2) / m_{1+2} < 0.67$
- Symmetry: $\mathrm{Min}(p_{T1},p_{T2})*\Delta R_{12} / m_{1+2} > 0.3$
  - or, $\mathrm{Min}(p_{T1},p_{T2}) / p_{T1+2} > 0.1$

BDRS Filtering



Recluster subjet constituents using $R \sim \Delta R_{12}/2$ to get rid of remaining junk (these are still big subjets!)

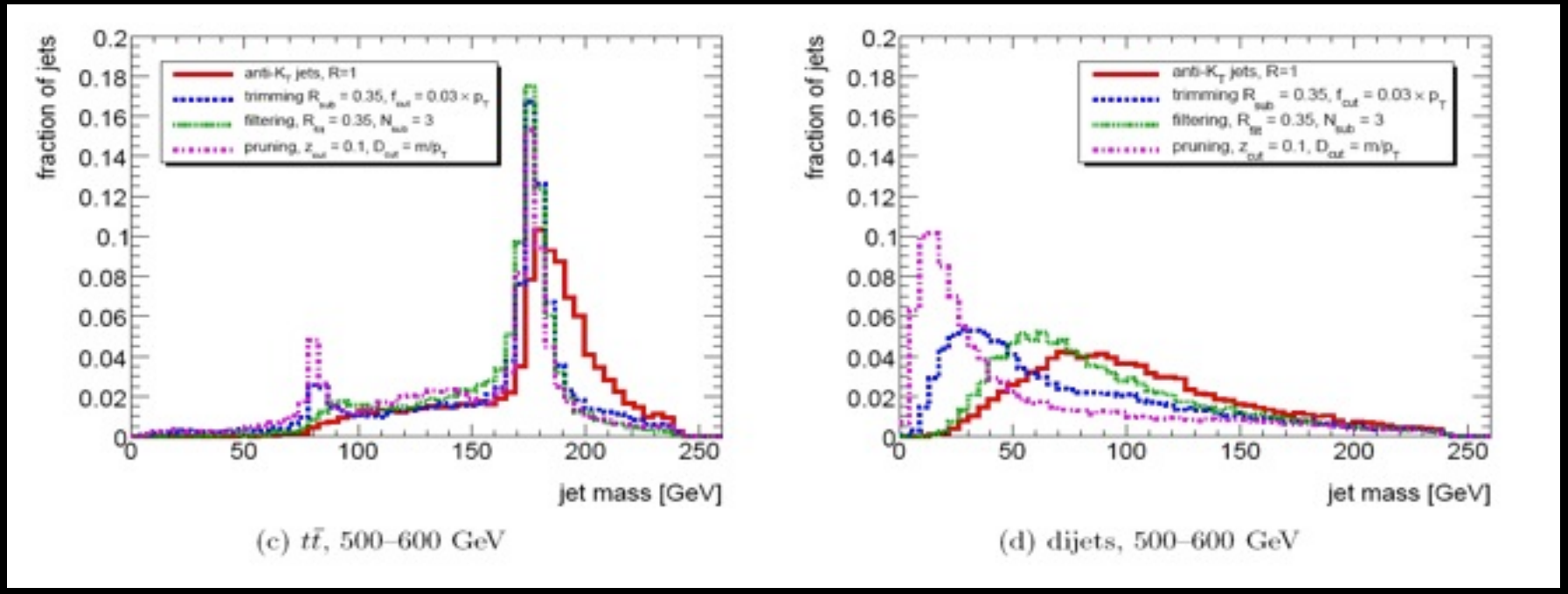
More "Jet Grooming" Strategies

- Pruning (Ellis, Vermilion, Walsh)
  - jet is selectively reclustered, ignoring the junk ("bottom-up" approach)
- Trimming (Krohn, Thaler, Wang)
  - remove regions of the jet with low energy density
    - recluster jet with miniature R (~0.2), throw away too-soft minijets

More "Jet Grooming" Strategies

- Pruning (Ellis, Vermilion, Walsh)
  - jet is selectively reclustered, ignoring the junk ("bottom-up" approach)
- Landscaping (Krohn, Thaler, Wang)
  - remove regions of the jet with low energy density
    - recluster jet with miniature R (~0.2), throw away too-soft minijets

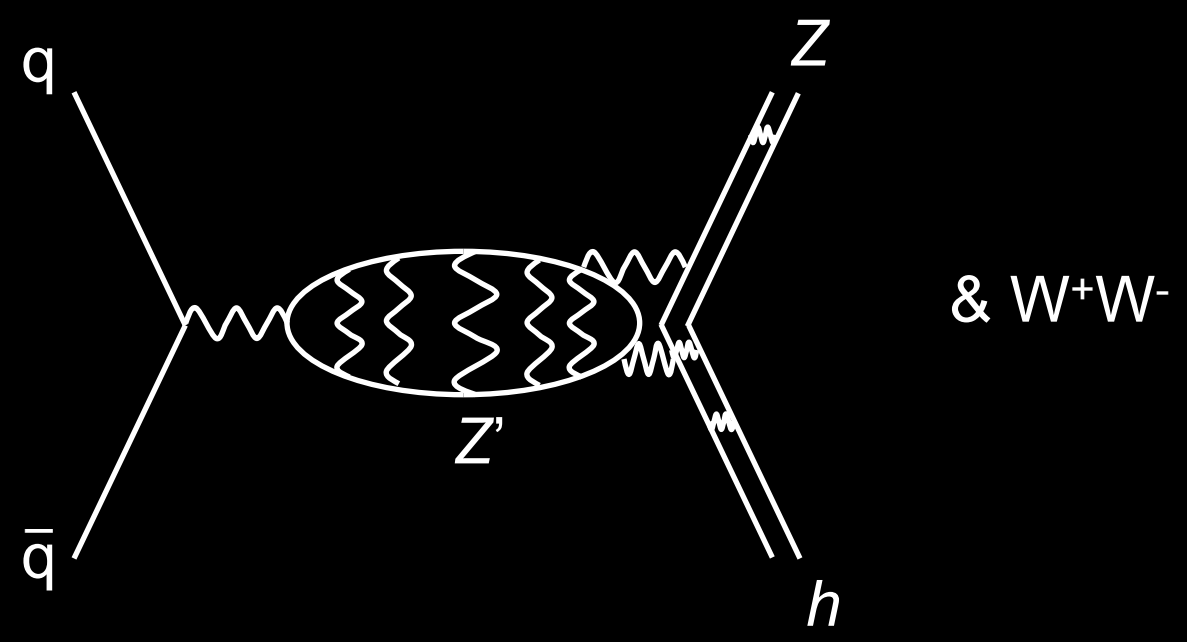
Grooming Comparison on Top-Jets



BOOST 2010, arXiv:1012.5412

complimentarity study: **Soper & Spannowsky, arXiv:1005.0417**

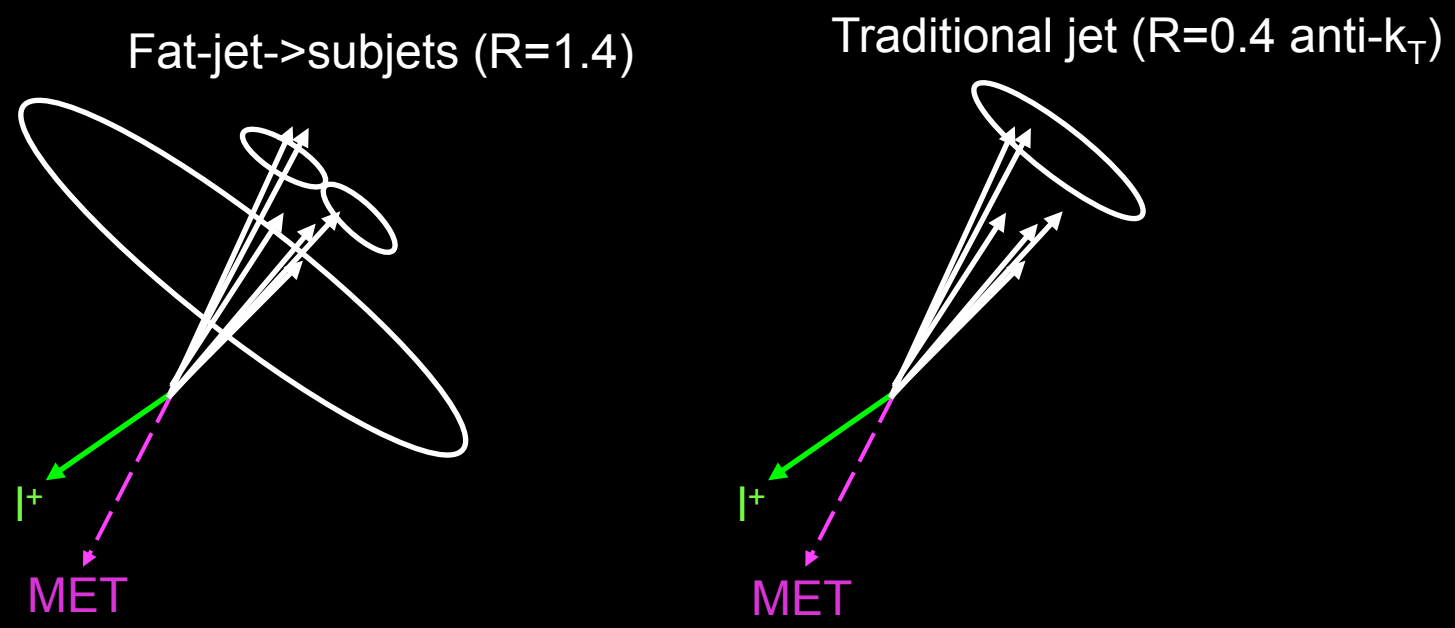
BDRS on Steroids: Z' -> Electroweak Bosons



Katz, Son, Tweedie, arXiv:1010.5253

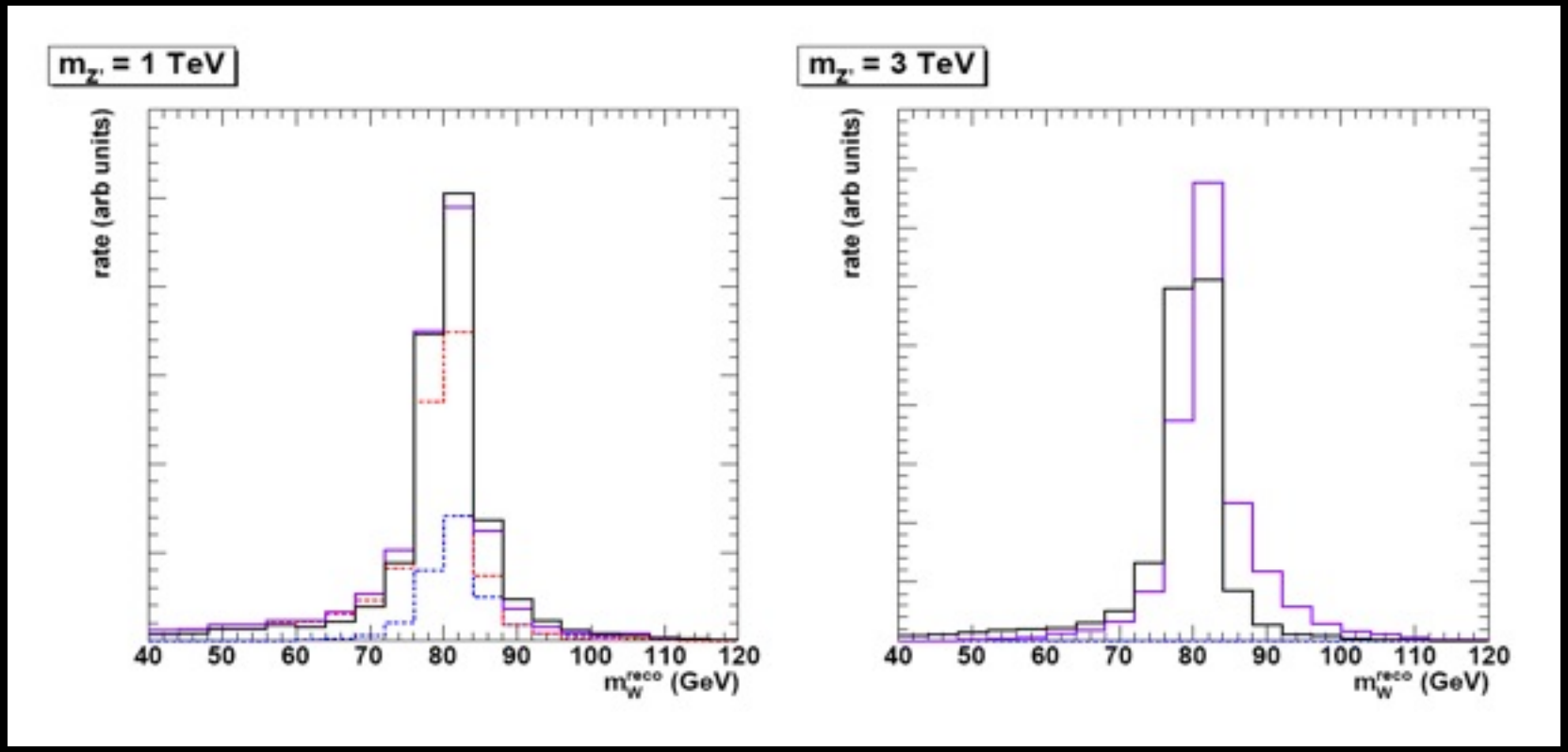
Related study (heavy h->ZZ): **Hackstein & Spannowsky, arXiv:1008.2202**

What is this Buying Us?: Substructure vs Traditional Jets



- Case study using particle-level *Z'*->*WW*->(*l*ν)(*qq'*) and *W*+jets->(*l*ν)+jets background events
  - Hunt for the hadronic *W* using either decomposed fat-jet or hybrid of traditional monojet and dijet searches

Substructure vs Jets: Signal



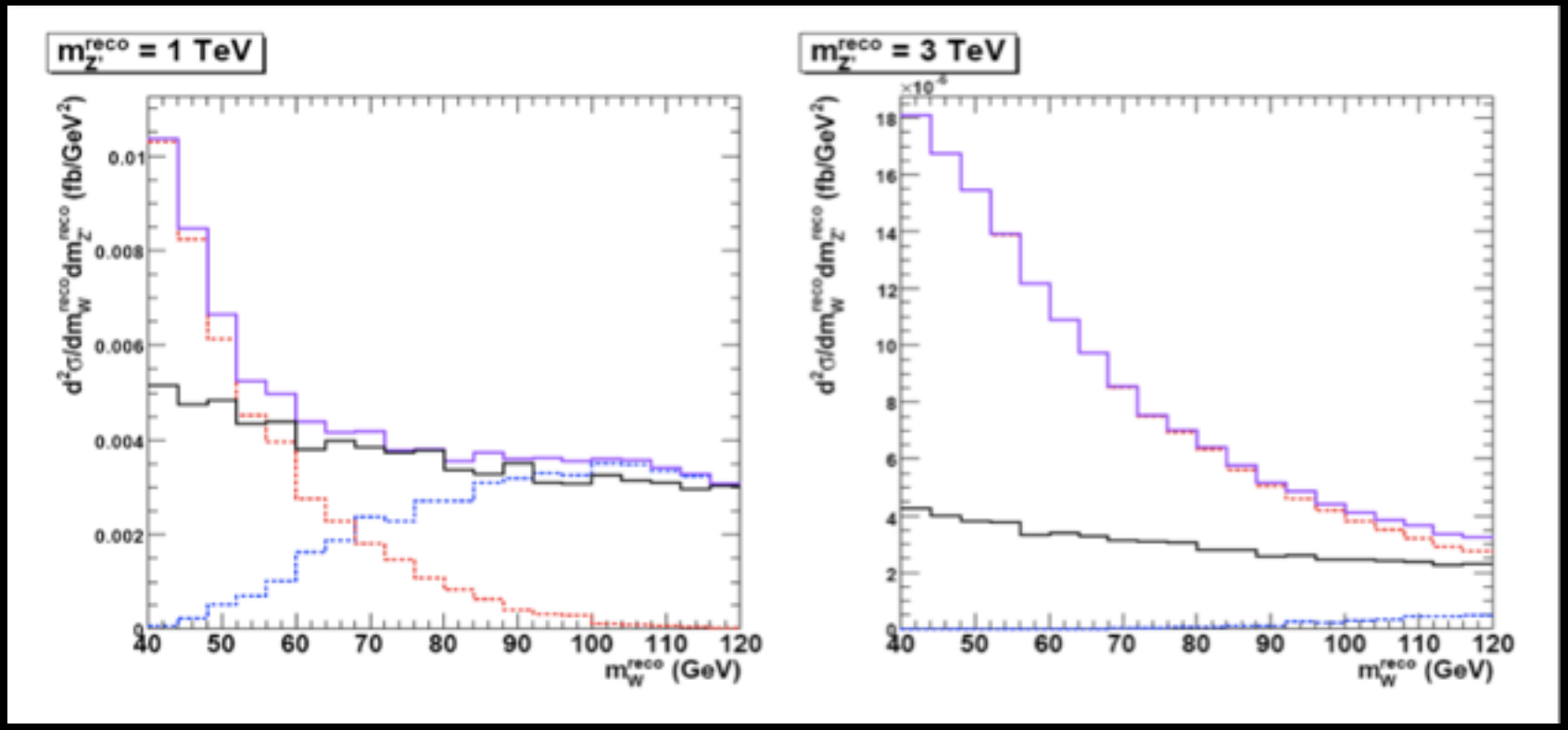
BDRS

traditional dijet

traditional monojet

traditional combined

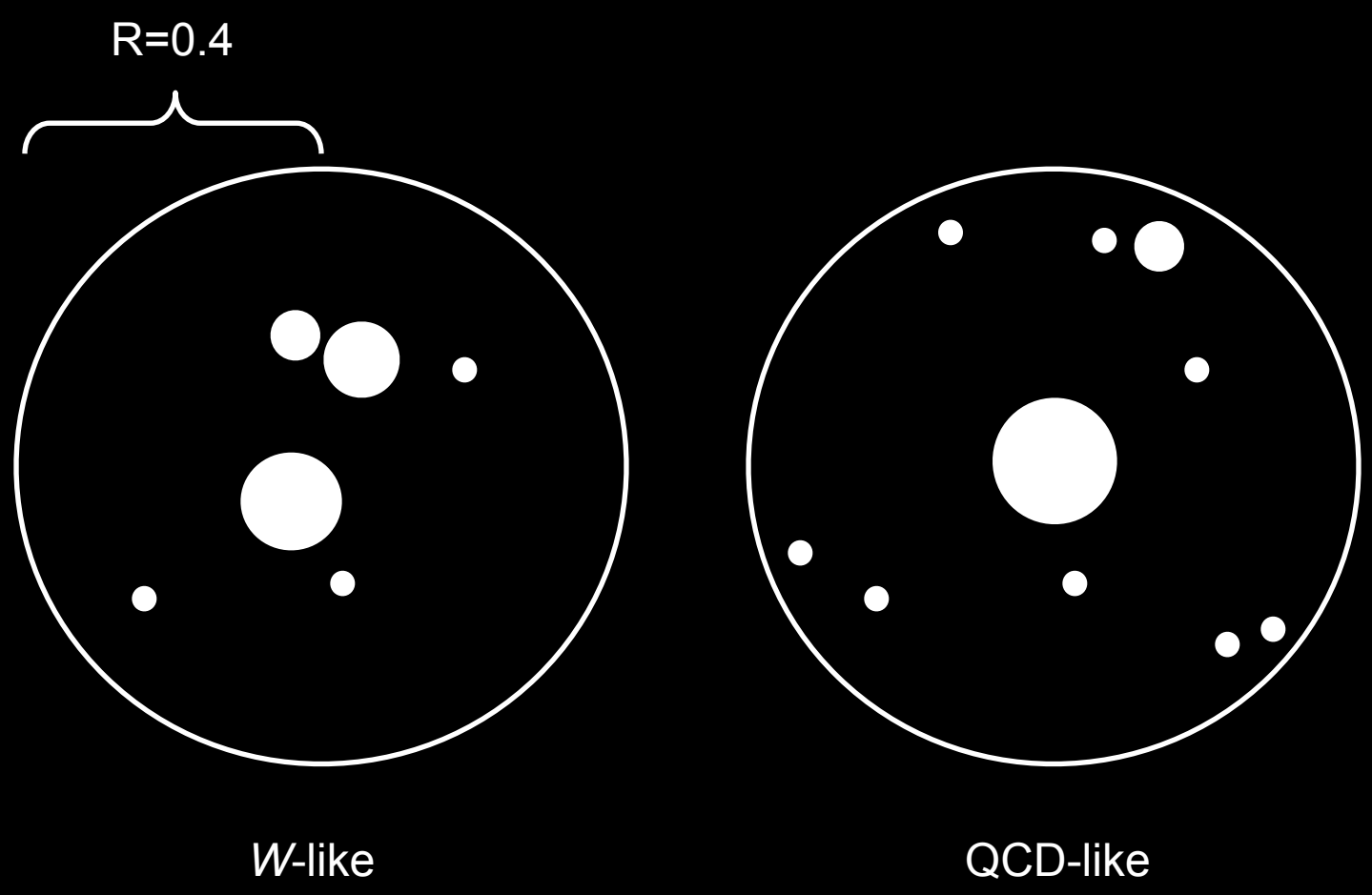
Substructure vs Jets: Background



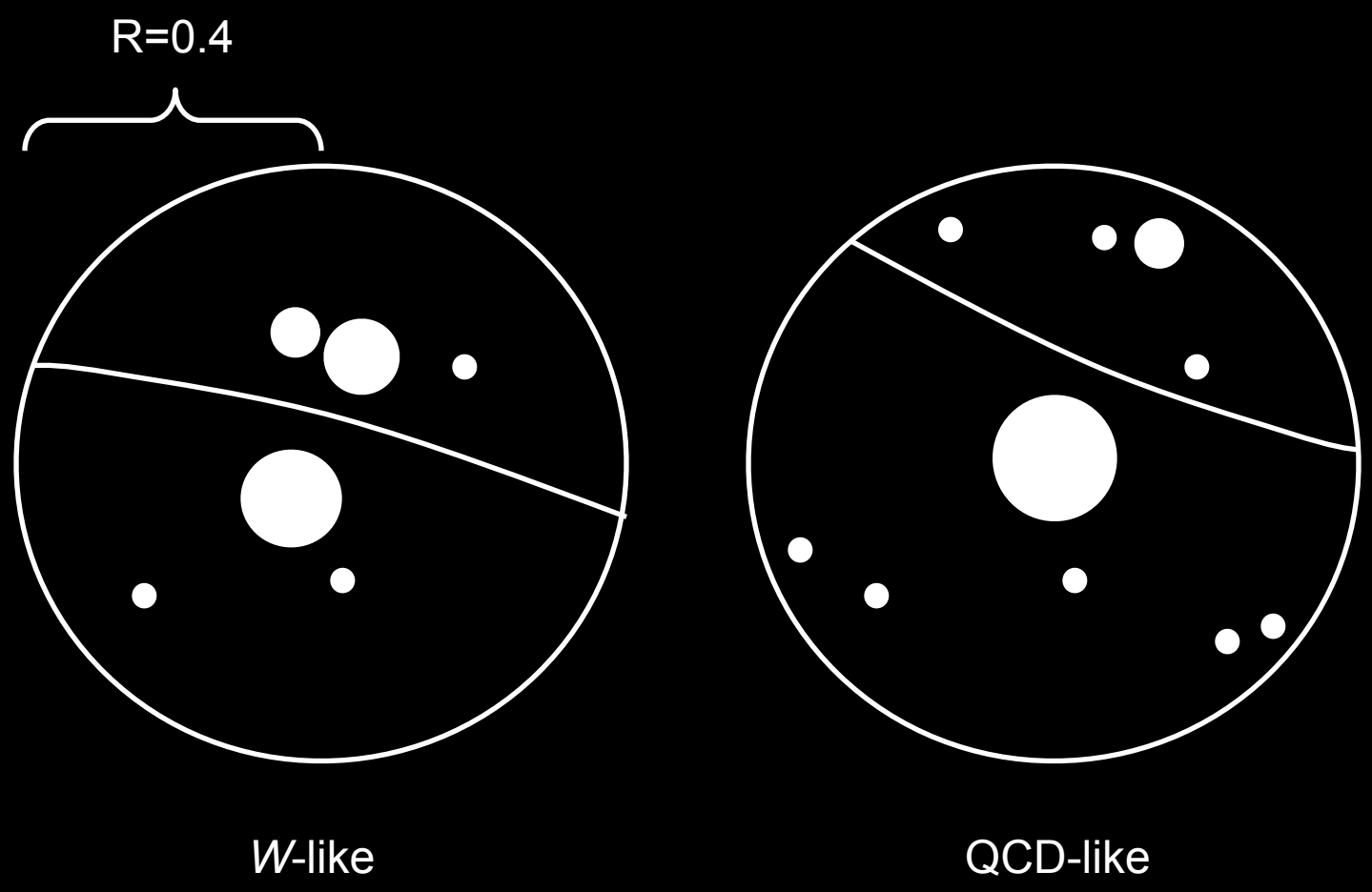
———	BDRS
— — —	traditional dijet
— — —	traditional monojet
———	traditional combined

PYTHIA 6.4 *Wq*/*Wg*, showered

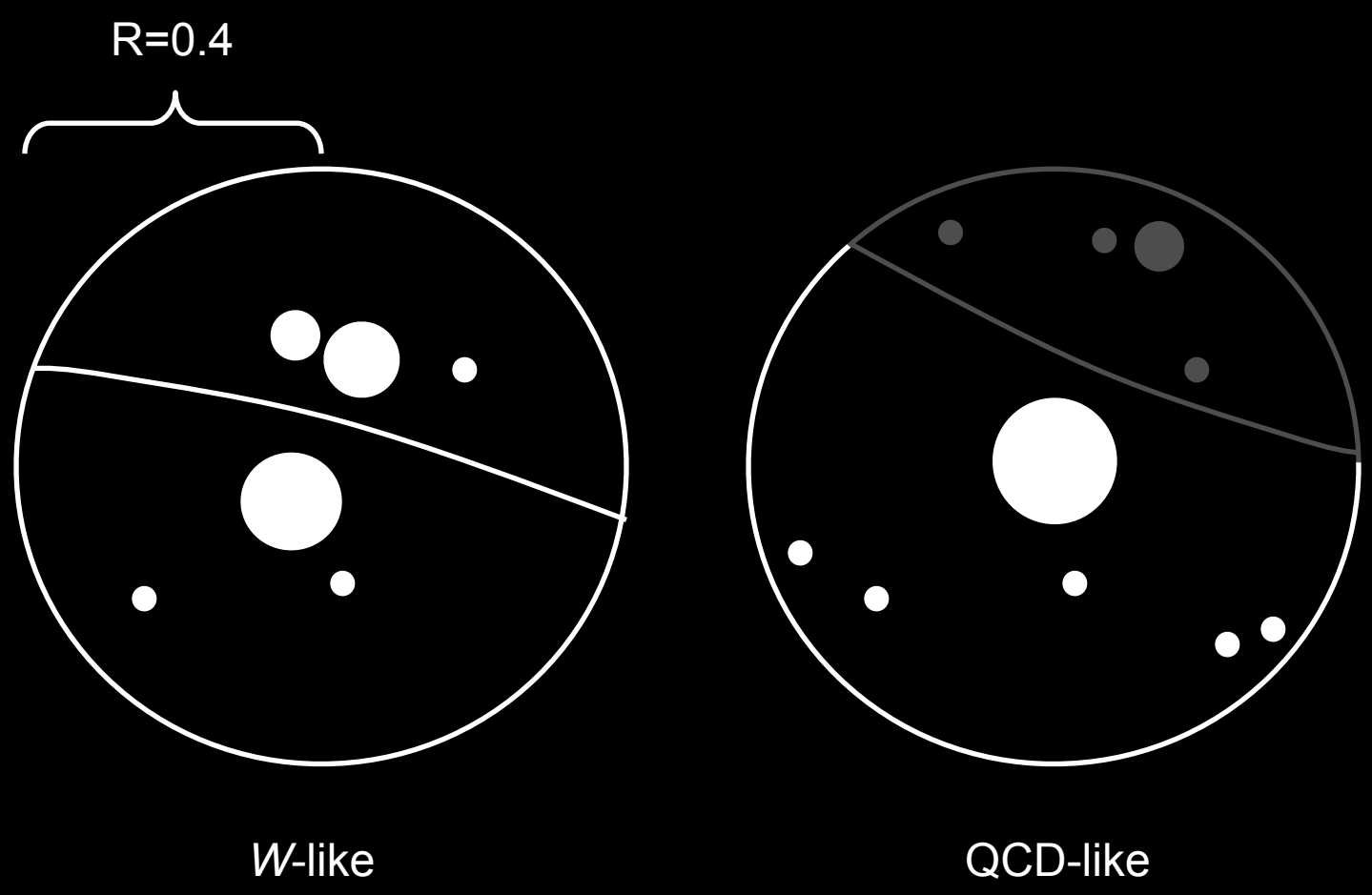
Internal Kinematics



Internal Kinematics



Internal Kinematics



W/Higgs Tag Rates from Substructure

SIGNAL

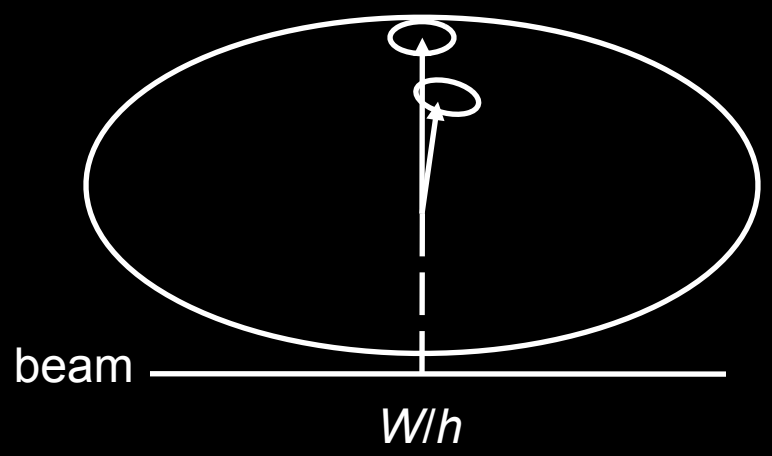
	$p_T \simeq 500$ GeV	$p_T \simeq 1000$ GeV	$p_T \simeq 1500$ GeV
W	76%	77	72
h	59	61	62

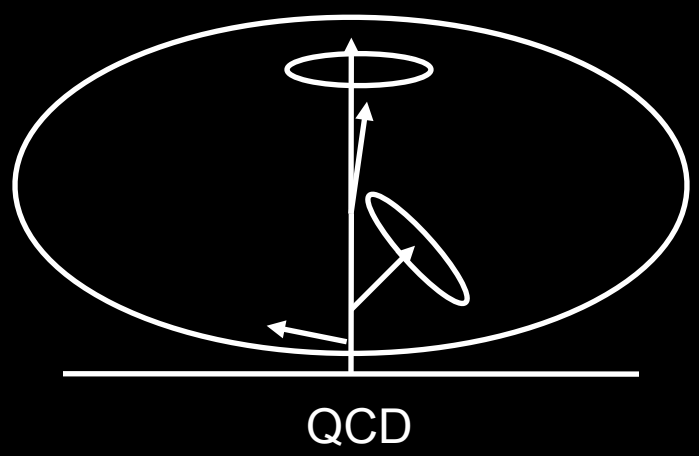
BACKGROUND

	$p_T \simeq 500$ GeV	$p_T \simeq 1000$ GeV	$p_T \simeq 1500$ GeV
quark $\rightarrow W$	6.5%	6.5	5.9
quark $\rightarrow h$	6.8	5.6	5.8
gluon $\rightarrow W$	10.4	8.3	7.4
gluon $\rightarrow h$	10.5	8.8	7.4

Counting events in windows [65,95] GeV for *W* and [100,140] GeV for Higgs (simple detector model applied)

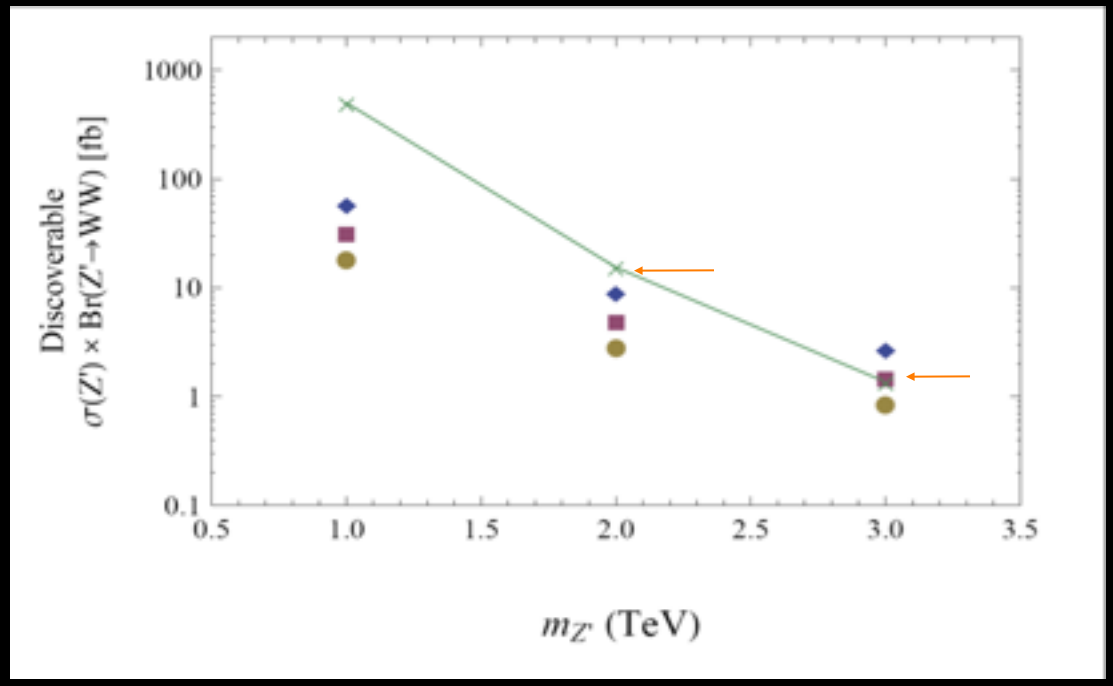
Why are the Mistag Rates Decreasing with p_T?





a hint of sensitivity to global color connections

Model-Independent Z'->WW Discovery Reach (LHC 14)



30/fb

100/fb

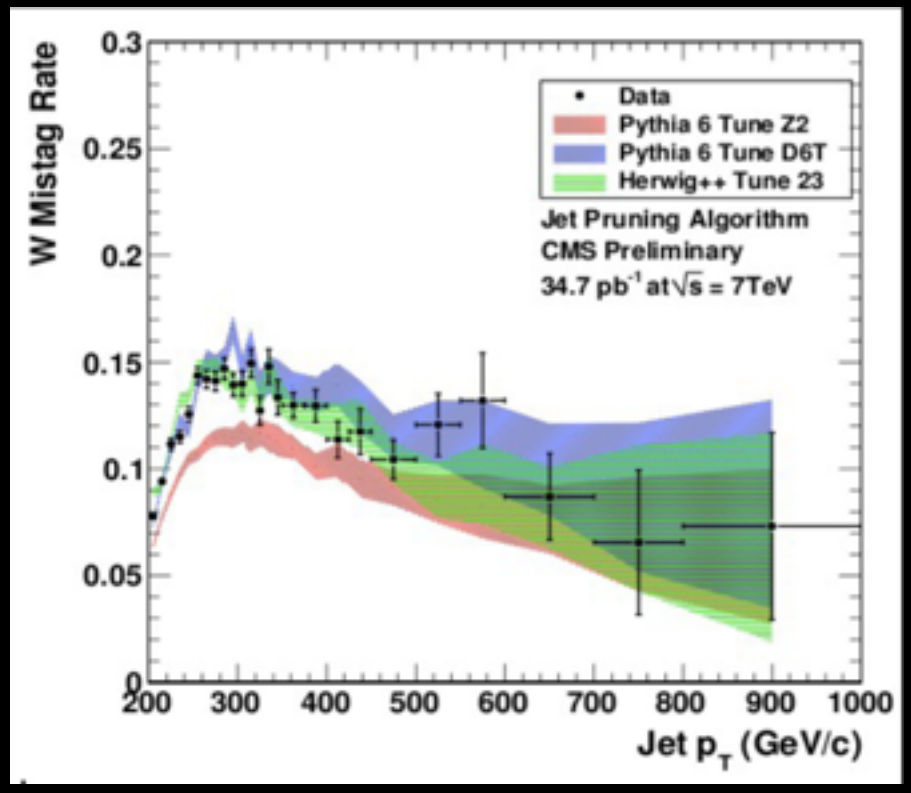
300/fb

Arrows indicate custodial RS model

Line indicates S/B=1

Earlier result (arXiv:0709.0007): Need 1000/fb

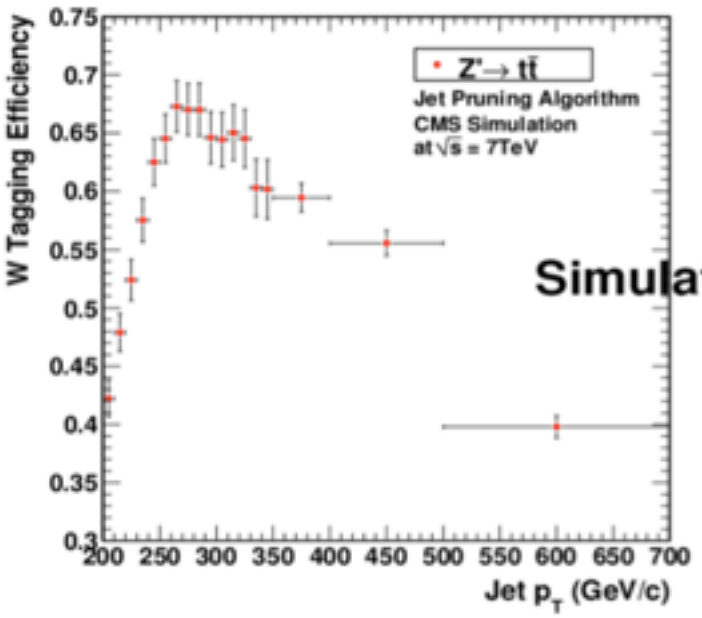
Data-Driven W Mistag (CMS)



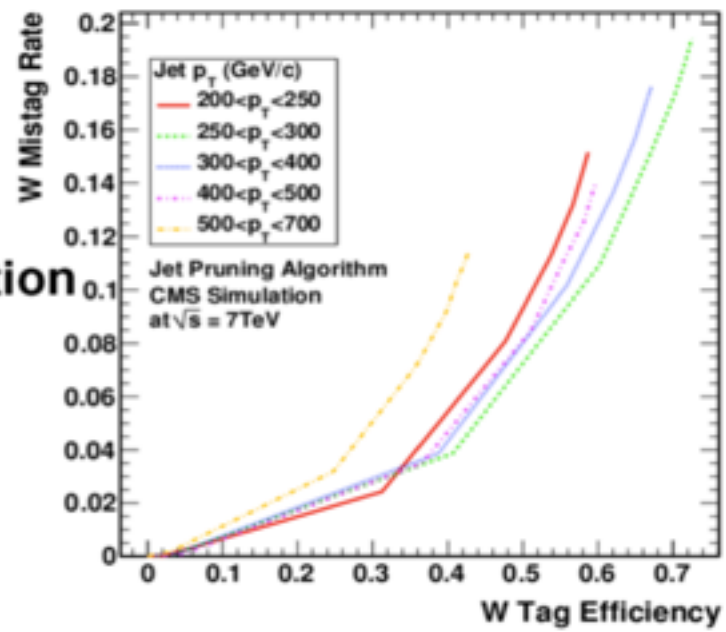
using pruning + BDRS style cuts on the final merging

Tag Rate and Tag vs Mistag

Monte Carlo signal efficiency for W tagging

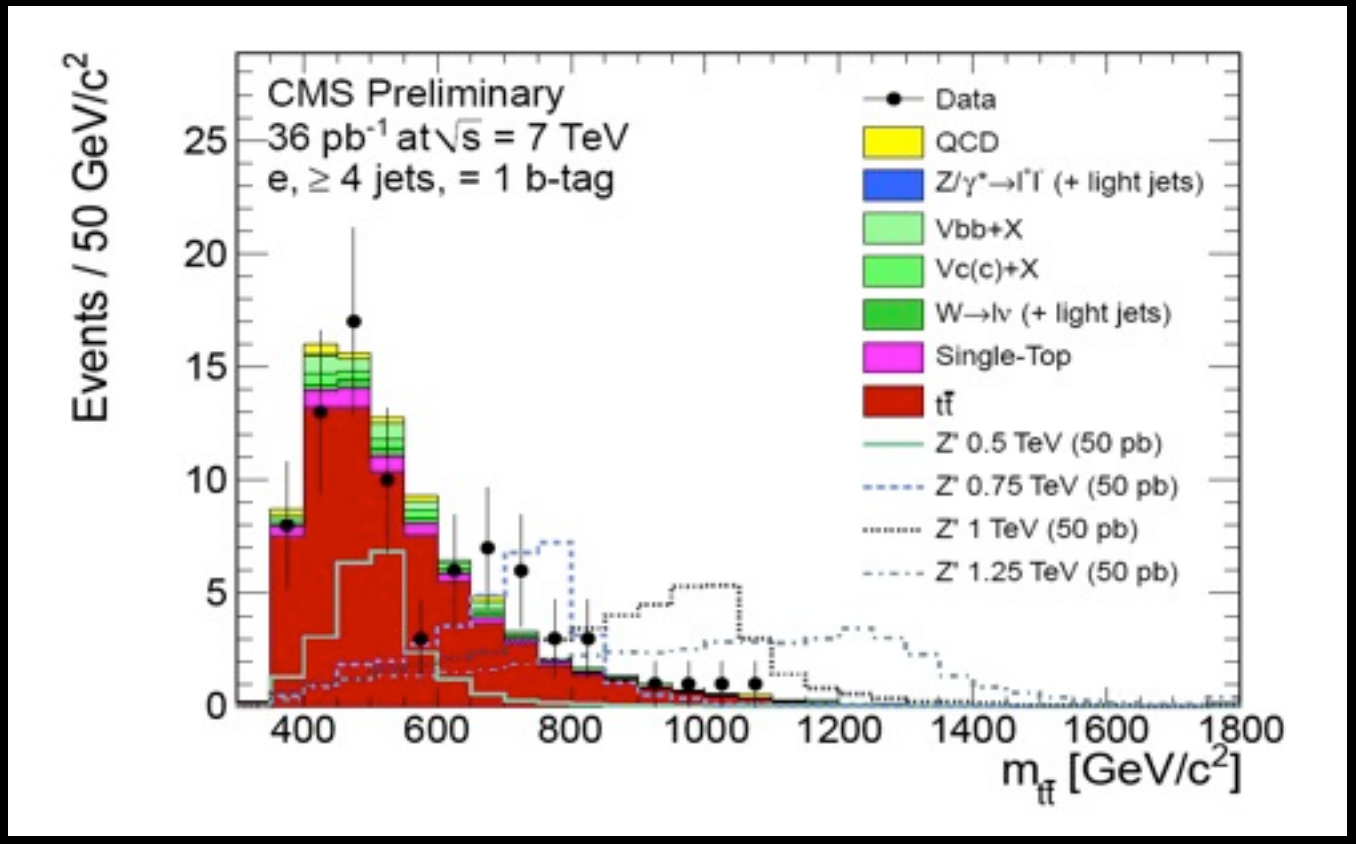


Monte Carlo mistag rate vs tag efficiency



Simulation

Boosted Tops

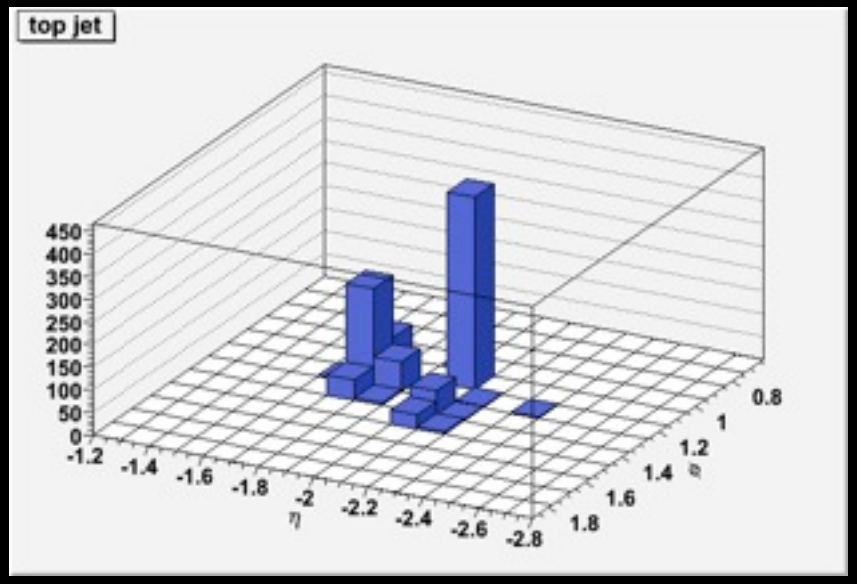


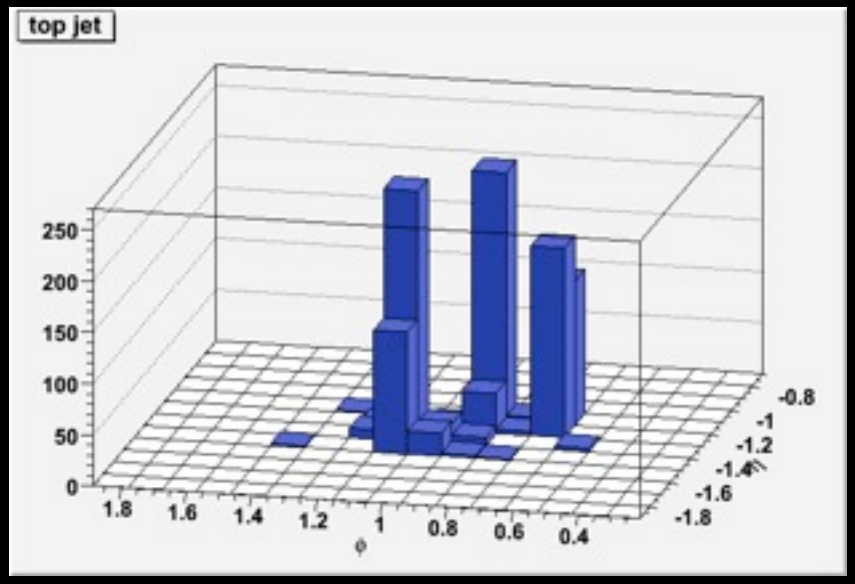
CMS PAS TOP-10-007

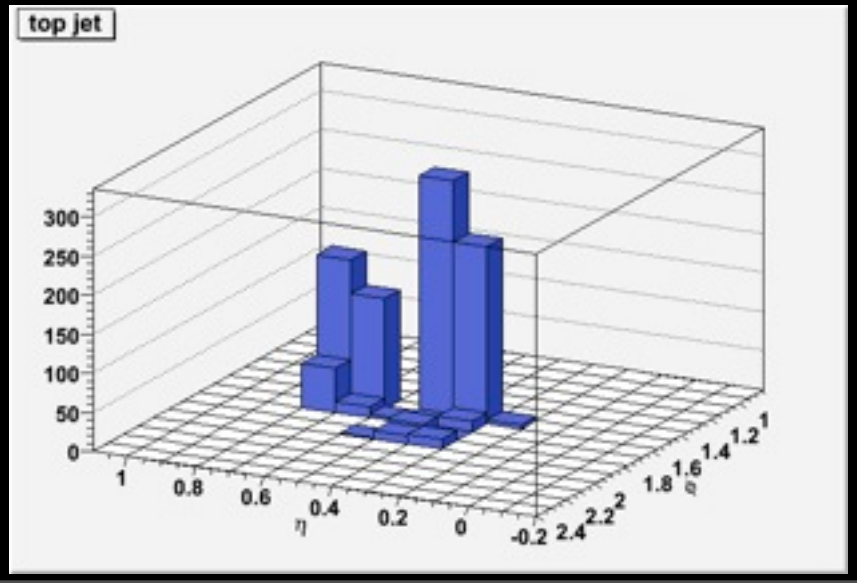
Top Tagging

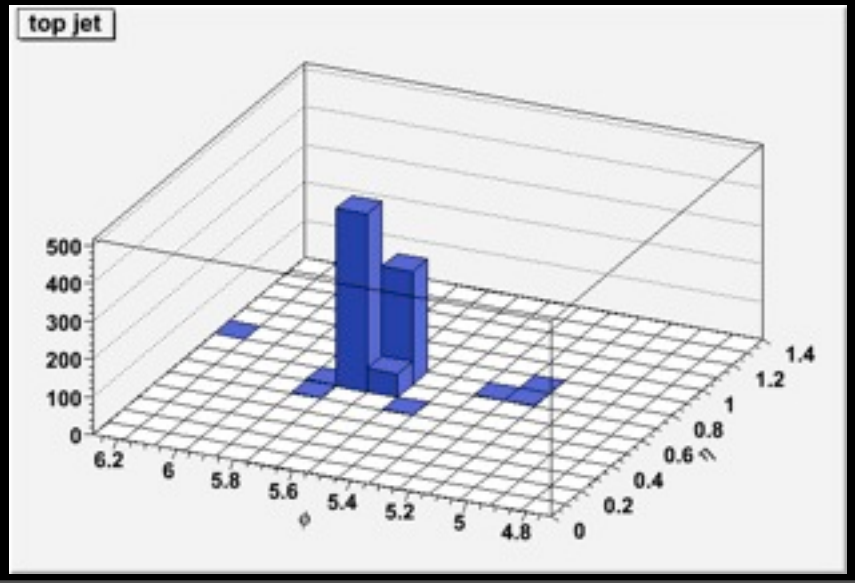
- Tear the jet down one more layer (or rebuild it from bottom-up)
  - 3 or 4 subjets
- >2 body kinematics
  - subjet pairwise invariant masses (look for the W, veto small-mass pairs)
  - reconstruct top and W decay angles
- Groom as needed

1 TeV Top-Jet Gallery

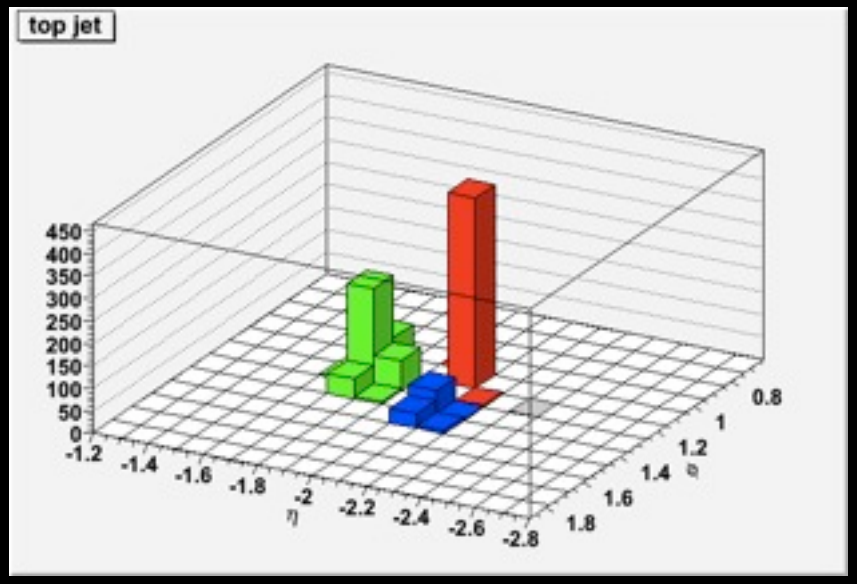


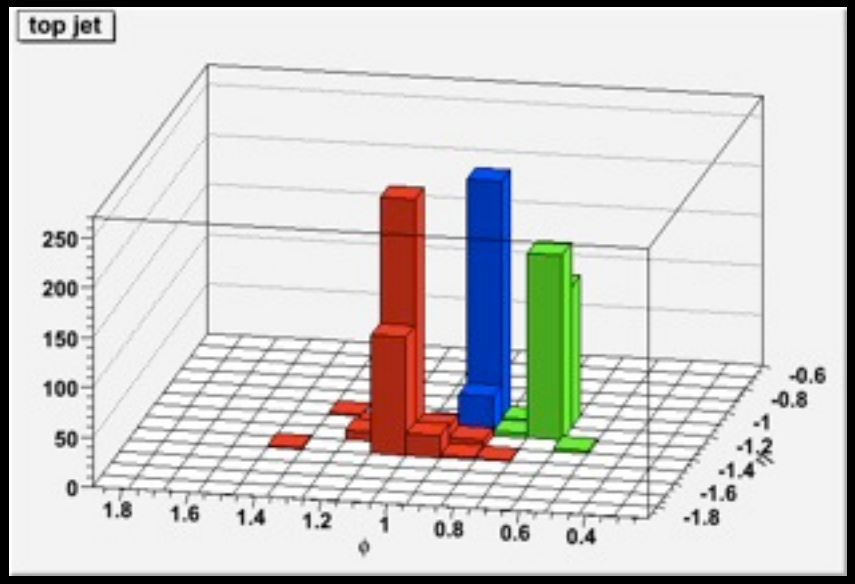


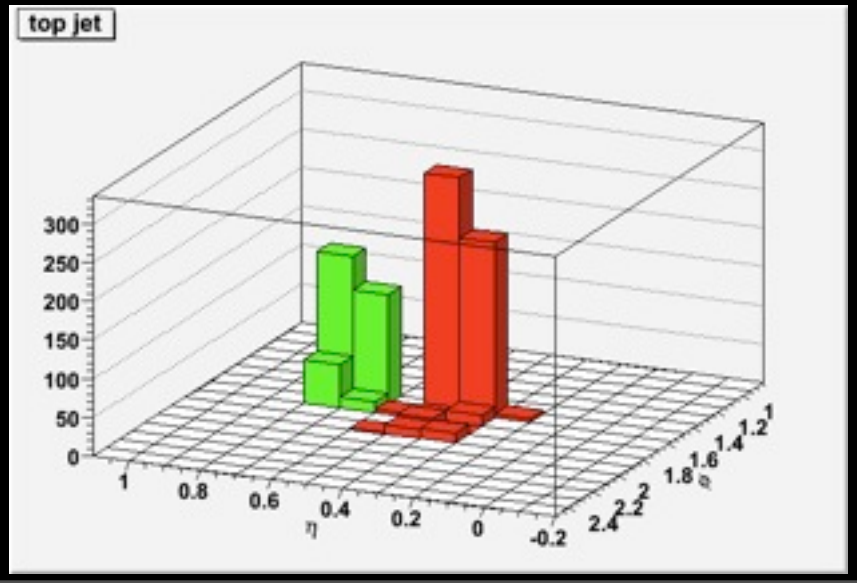


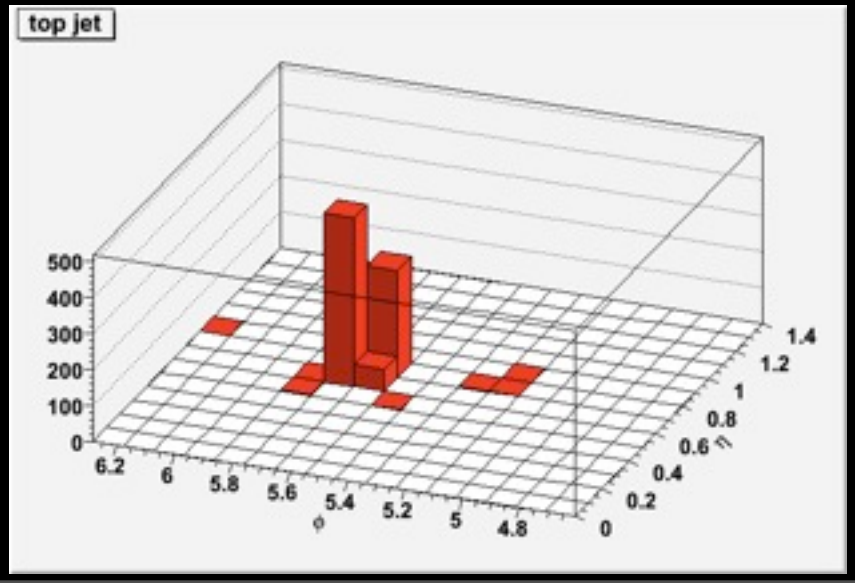


1 TeV Top-Jet Gallery

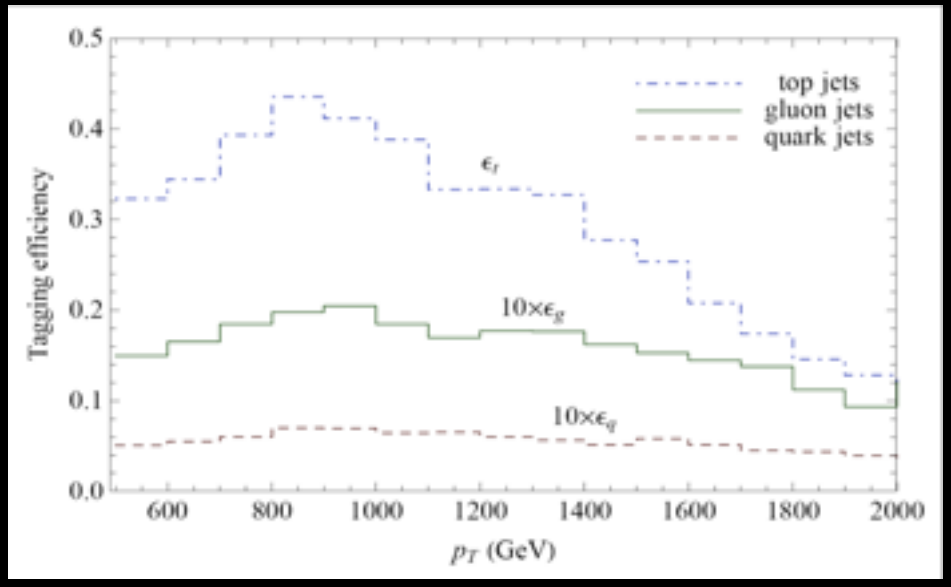




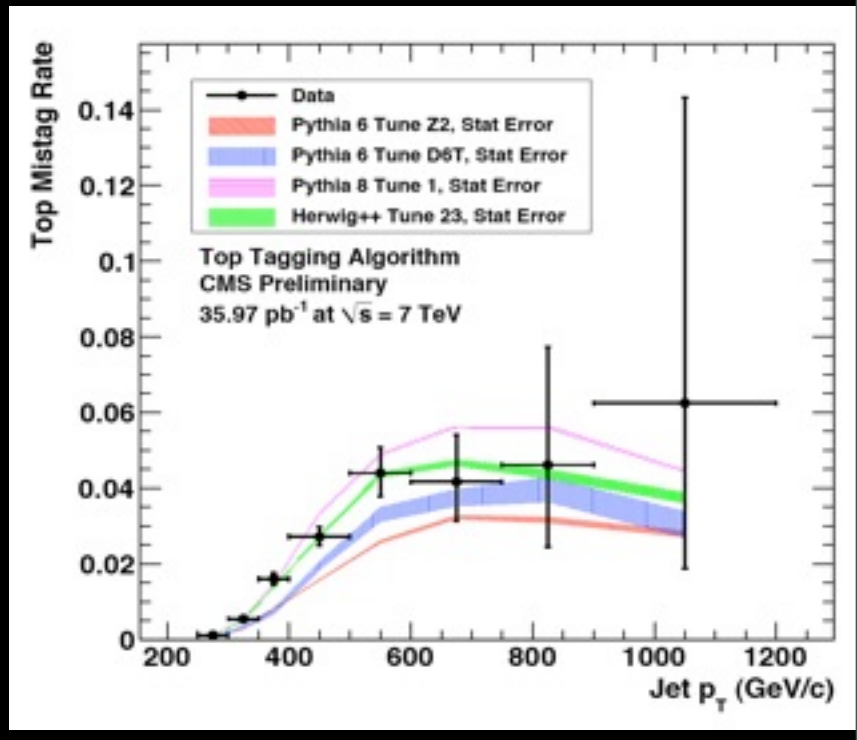




Data-Driven Top Mistag (CMS)

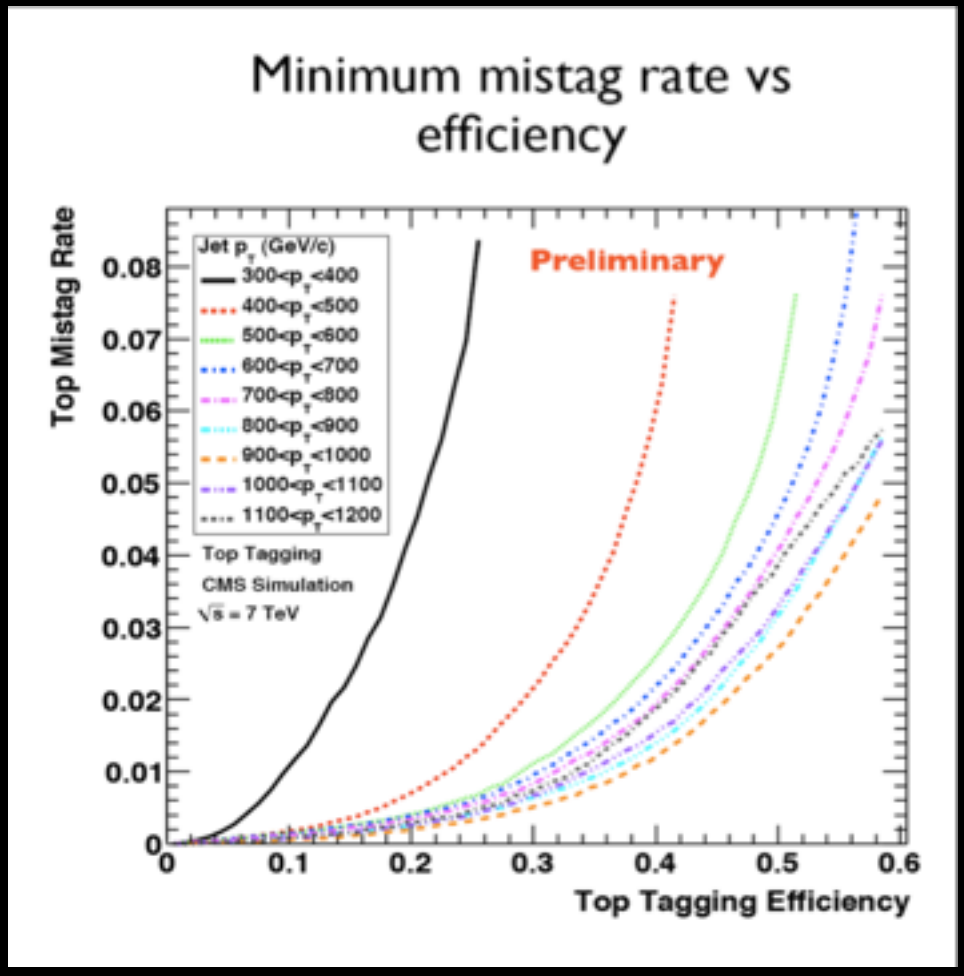


Hopkins top-tagger on (bad) simulalation



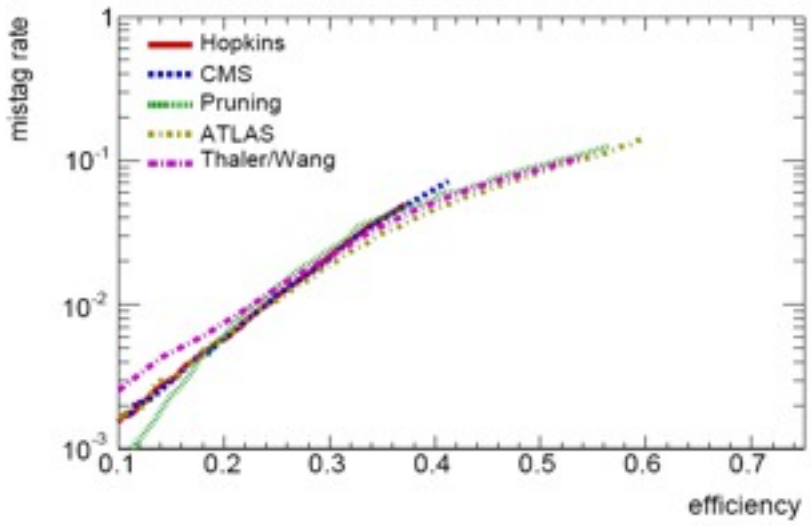
CMS top-tagger on data

CMS Tag vs Mistag

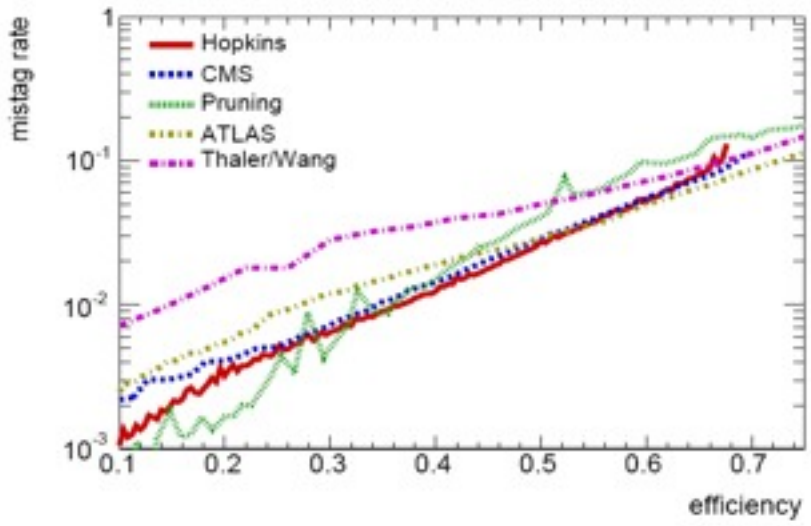


Simulation study

Many Other Techniques…



(c) 300–400 GeV



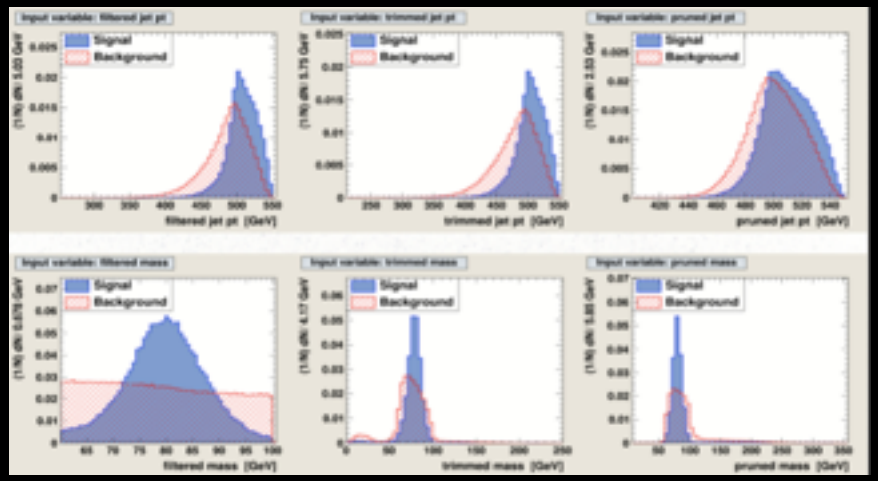
(d) 500–600 GeV

- ATLAS: Brooijmans
- Thaler/Wang: Thaler, Wang
- + HEPtagger: Plehn, Spannowsky, Takeuchi, Zerwas

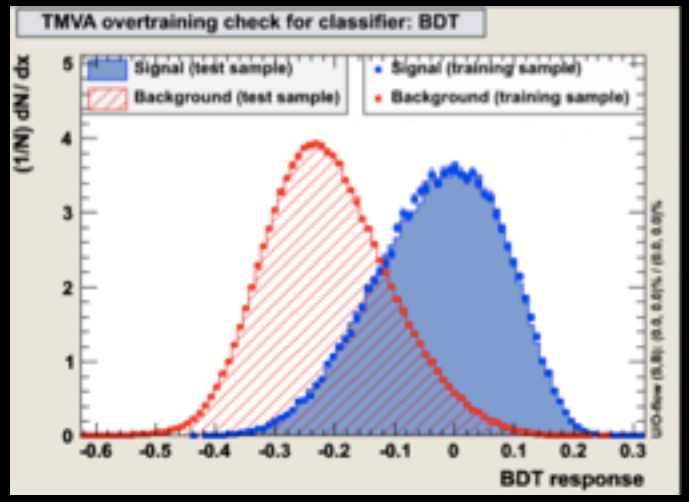
BOOST 2010, arXiv:1012.5412

Seeing W-Jets in Color…in Many Dimensions

Many weakly color-sensitive discriminants

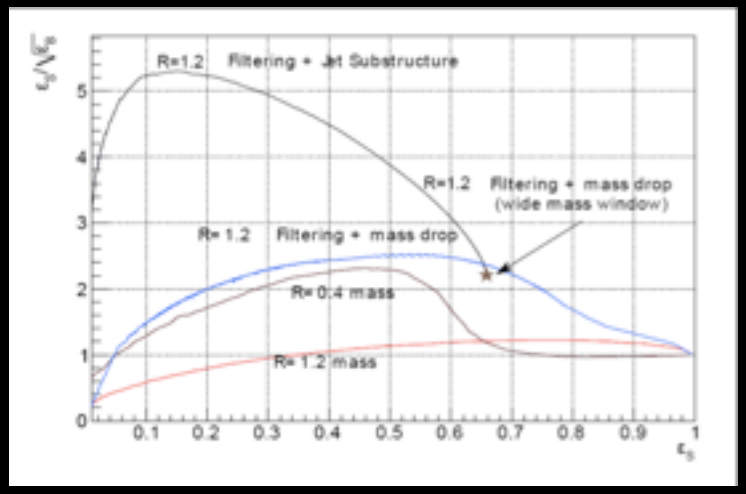


One powerful discriminant
(Boosted Decision Tree)

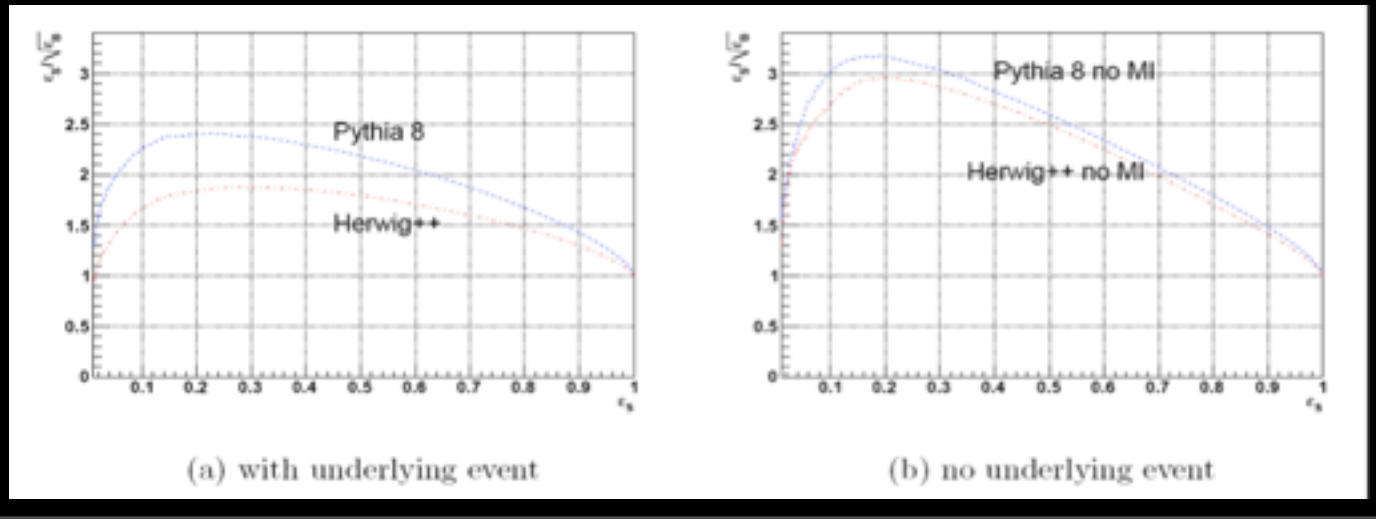


Cui, Han, Schwartz, arXiv:1012.2077

Seeing W-Jets in Color…in Many Dimensions

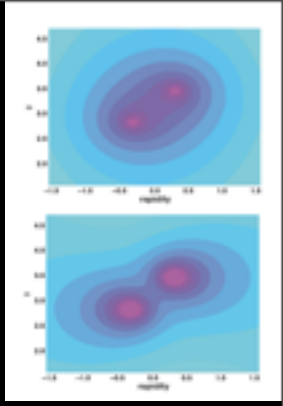


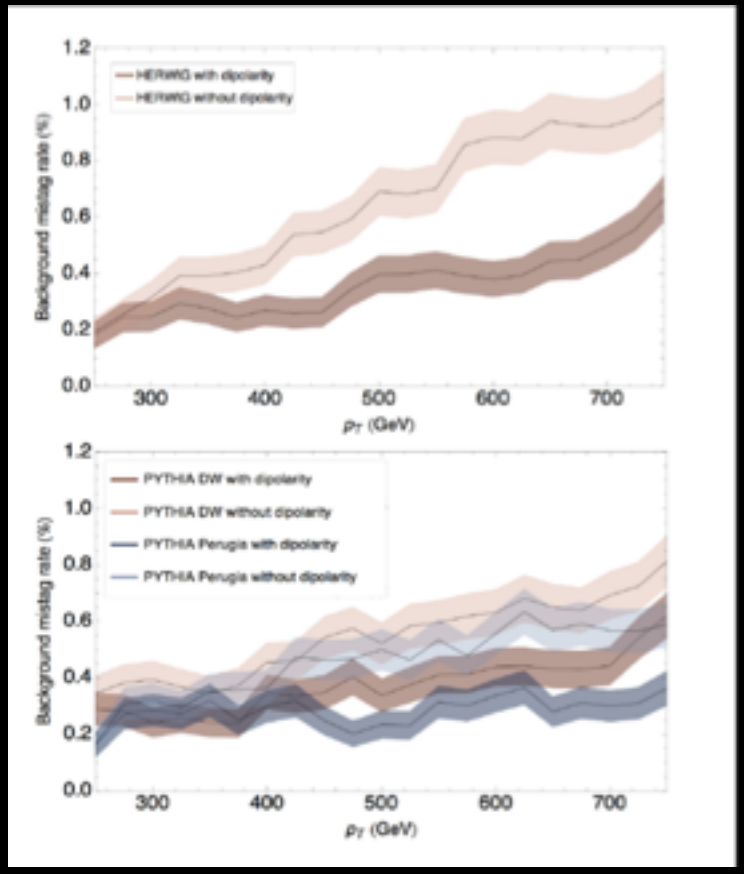
p_T ~ 500 GeV

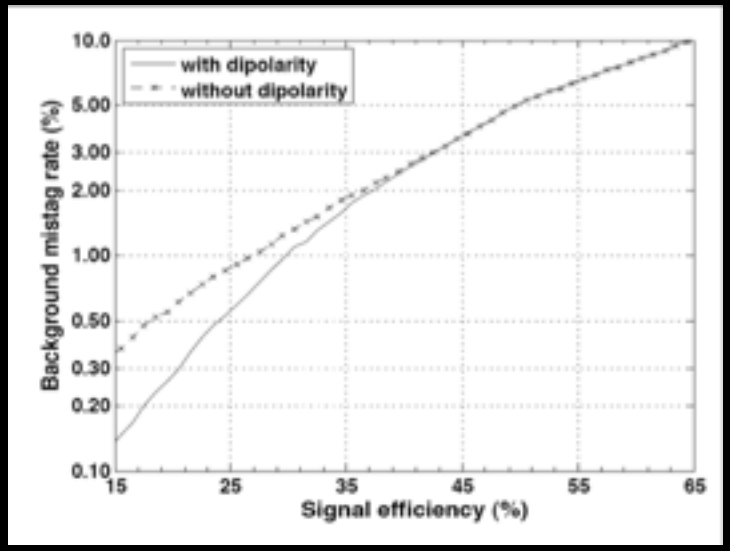


(a) with underlying event (b) no underlying event

Top-Tag with Dipolarity

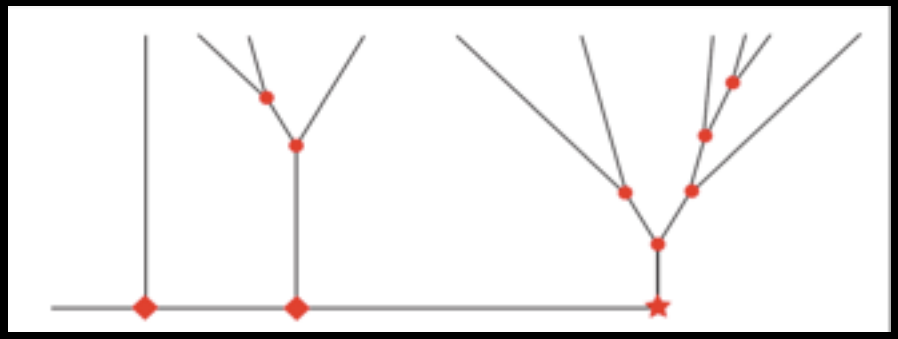


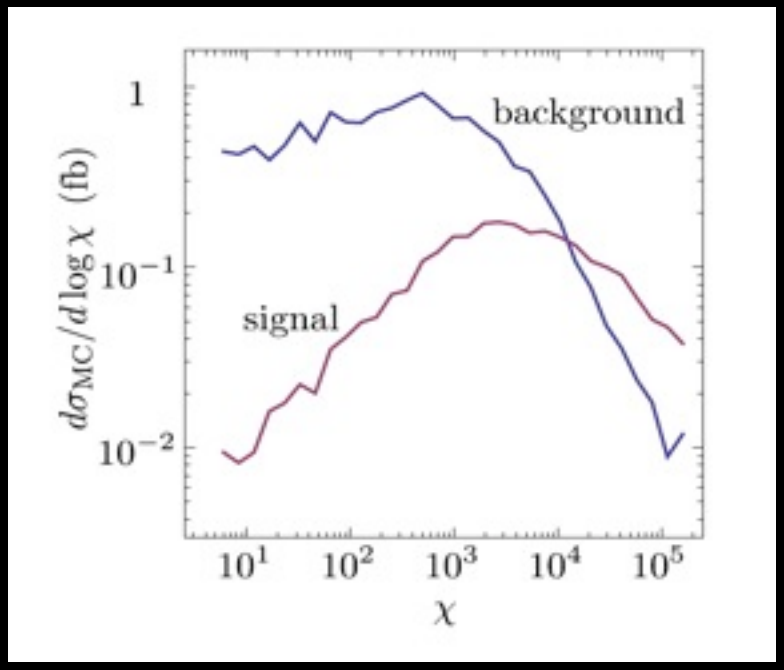




Hook, Jankowiak, Wacker, arXiv:1102.1012

Shower Deconstruction





Soper & Spannowsky, arXiv:1102.3480

Jet Shape Ideas

- N-subjettiness: Kim, Thaler & Van Tilberg
  - More amenable to PQCD calculations?
- Jet-angularities, planar flow, etc: Almeida, Lee, Perez, Sterman, Sung, Virzi
  - Ditto
- Template overlap: Almeida, Lee, Perez, Sterman, Sung

More Complicated Processes

- Top-Higgsstrahlung: Plehn, Spannowsky, Salam
- Boosted Higgs in SUSY: Adam Martin + UO

New Physics Substructure

- RPV neutralino: Butterworth, Ellis, Raklev, Salam
  - an alternative 3-quark final-state
- Unburied Higgs: Falkowski, Krohn, Shelton, Thalapillil, Wang; Chen, Nojiri, Sreethawong
  - h->aa->(gg)(gg)
  - exploit color isolation (analogous to taus)

Jets With Embedded Leptons

- Semileptonic boosted tops: Thaler & Wang, Rehermann & Tweedie
- Boosted h->$\tau\tau$: Katz, Son, Tweedie

Substructure at ILC or CLIC?

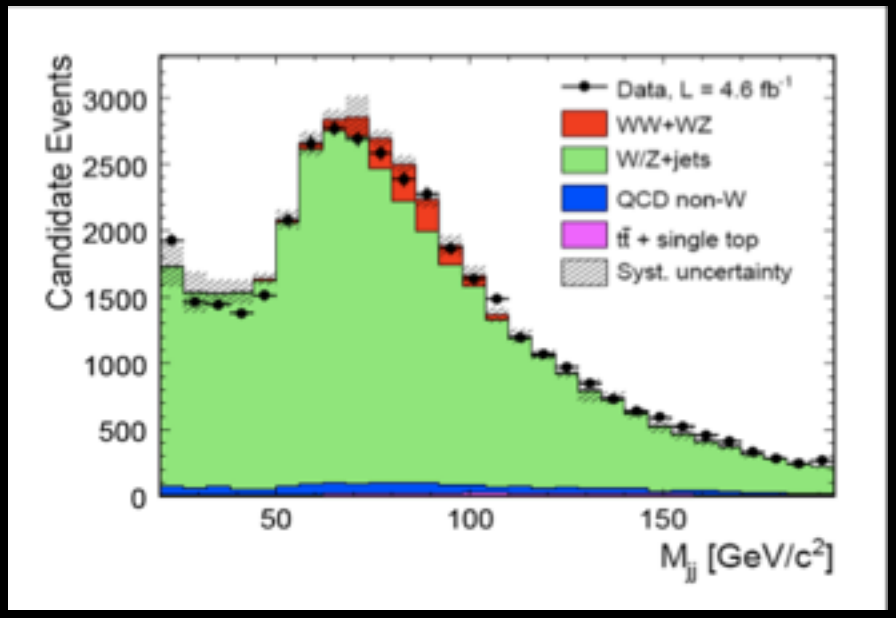
- No UE, no ISR
  - no trimming, no pruning, no filtering (?)
- Color singlet machine => “lepton+jets” backgrounds are of a very different quality
  - e.g., Wq, Wg absent
  - W+jets still there, but *much* lower rate
- Clearly useful for all-hadronic search channels (q-qbar backgrounds)
- Boosted SUSY?

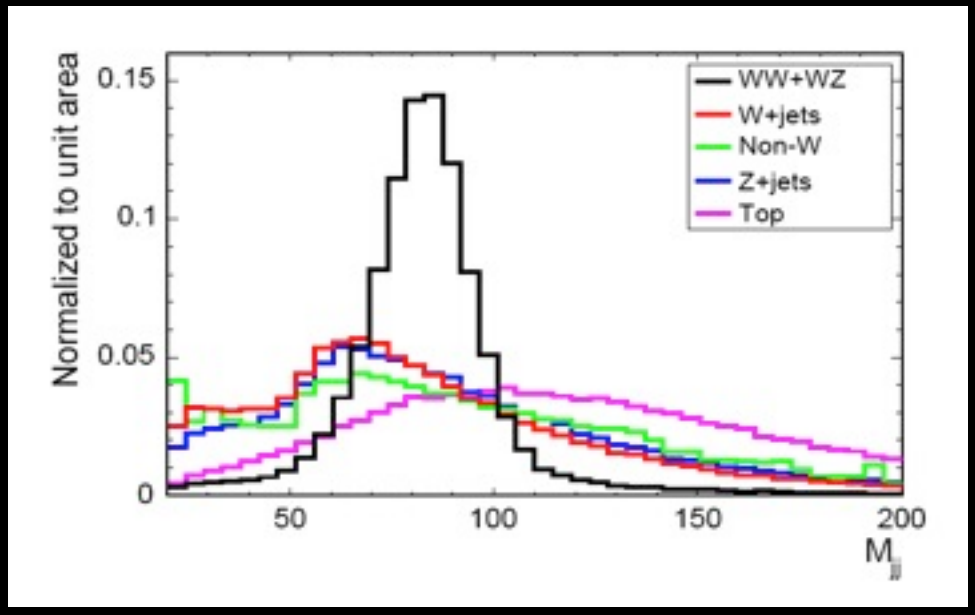
Summary

- Lots of ideas to beat QCD and identify EW boson jets & top jets, and all kinds of other good stuff
- So far mistag rates look sane in data
  - Stick around for Marcel's talk for an actual experimentalists' perspective on all of this
- Relevance for ILC1000 or CLIC?

Extras

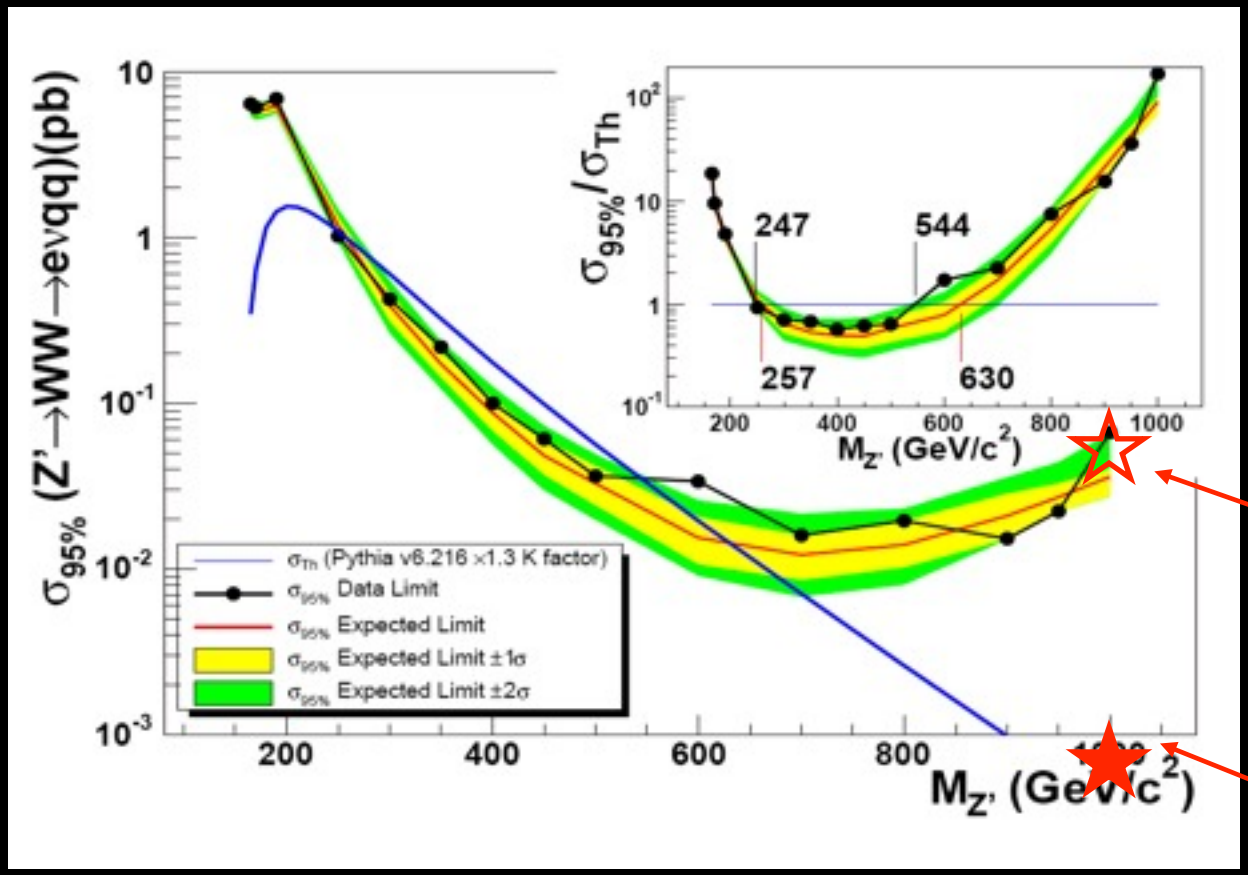
CDF Discovery of Semi-Leptonic WW/WZ





CDF arXiv:1008.4404 (4.6/fb of data)

CDF Exclusion of WW/WZ Resonance



(Warped KK Z' at *LHC* energy)

Warped KK Z' at Tevatron energy

CDF arXiv:1004.4946 (2.9/fb of data)