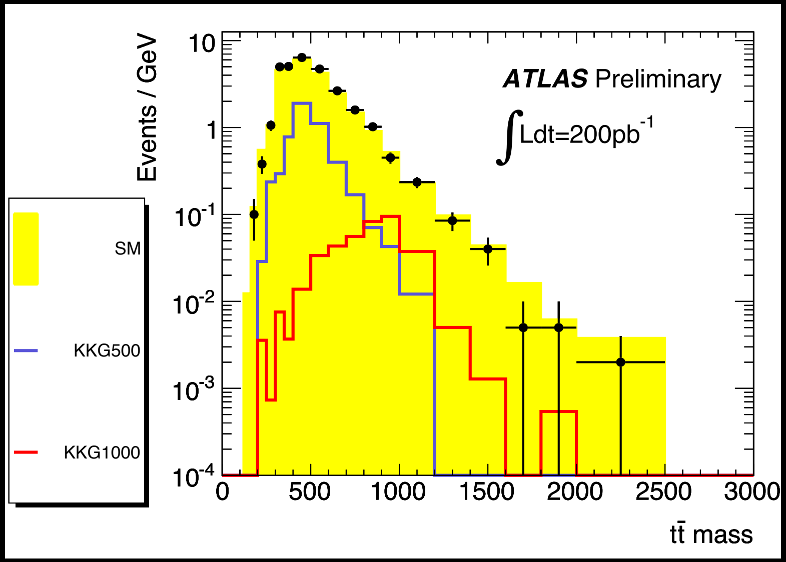
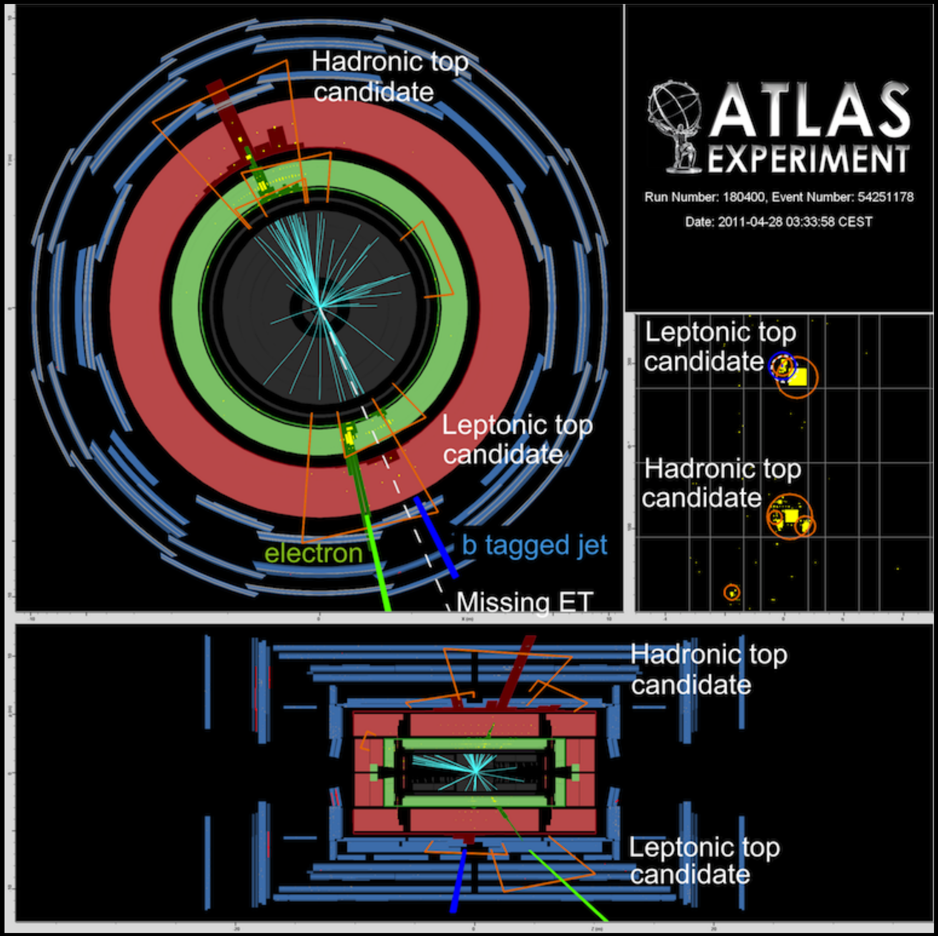
Tagging EW Boson-Jets and Top-Jets

Brock Tweedie
Boston University
23 July 2011
@ EPS 2011 Grenoble

Welcome to the TeV Scale



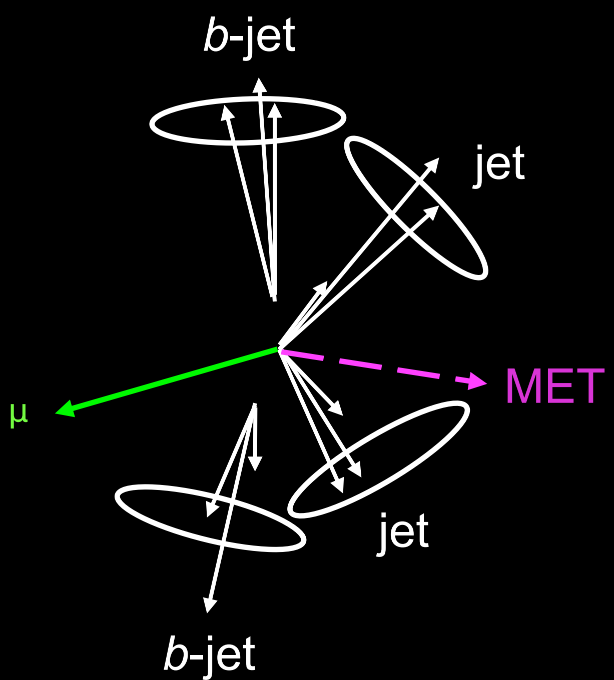
ATLAS-CONF-2011-087
(5 June 2011, 200 pb^{-1})



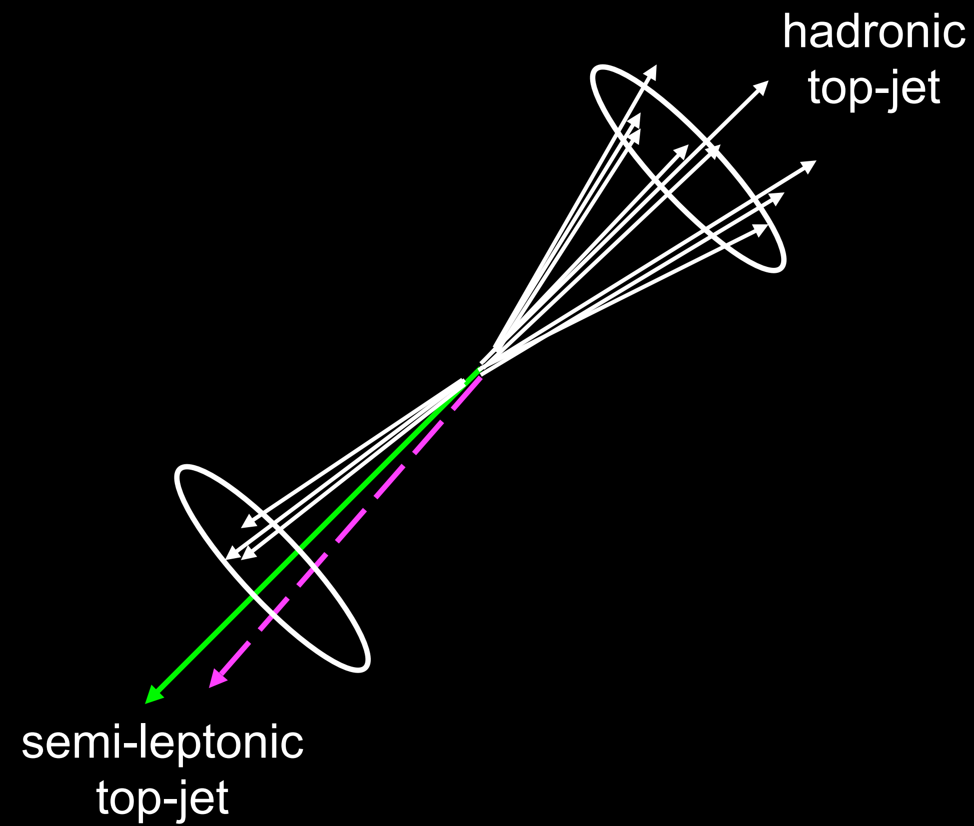
m_{tt} = 1602 GeV

TeV >> m_t

$E_{CM} \sim 2m_t$



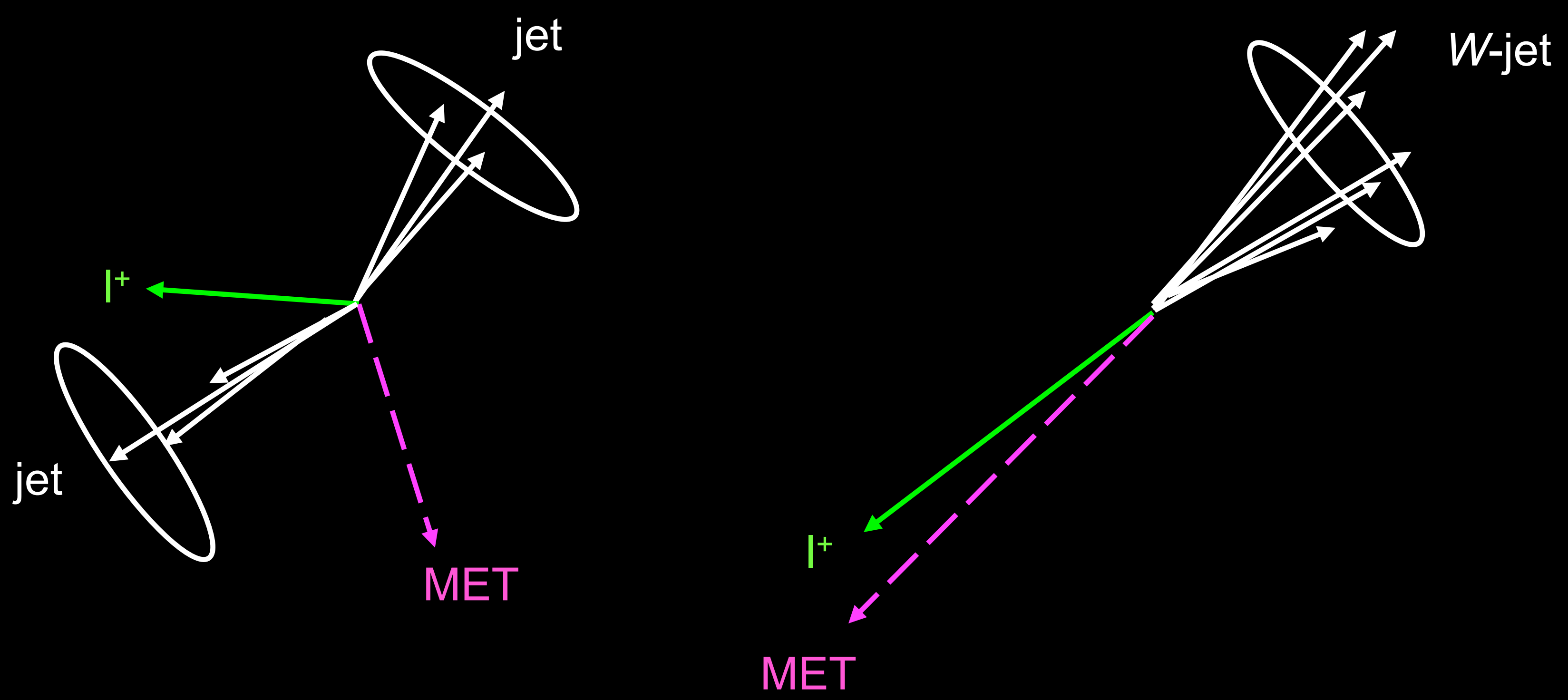
$E_{CM} >> 2m_t$



TeV >> m_W

$E_{CM} \sim 2m_W$

$E_{CM} >> 2m_W$

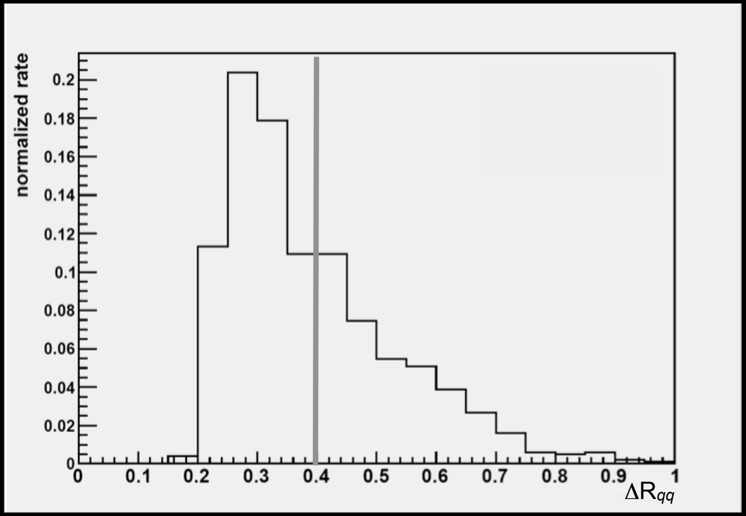


Bottom Line

- At high E, everything becomes a jet
  - analogous to conceptual transition we've made with tau, charm, bottom
  - e.g., *Z* -> $\tau\tau$ / *cc* / *bb* => *Z'*-> *tt* / *Zh* / *WW*
- This is a blessing
  - combinatorics become much easier
  - more complete radiation containment
  - dangerous backgrounds can become tamer due to PDFs and/or kinematics (cf., boosted Higgs search)
- And it is also a curse
  - normal jet reco merges decay products, losing kinematic info
  - large energy flow in core of jet => uncorrelated soft radiation at periphery affects mass reco ($\Delta m^2 \sim p_T * \rho_{UE+PU} * R^4$)
  - small angles -> new regime for detector fuzziness issues
  - how to model QCD backgrounds without tripping over logs?

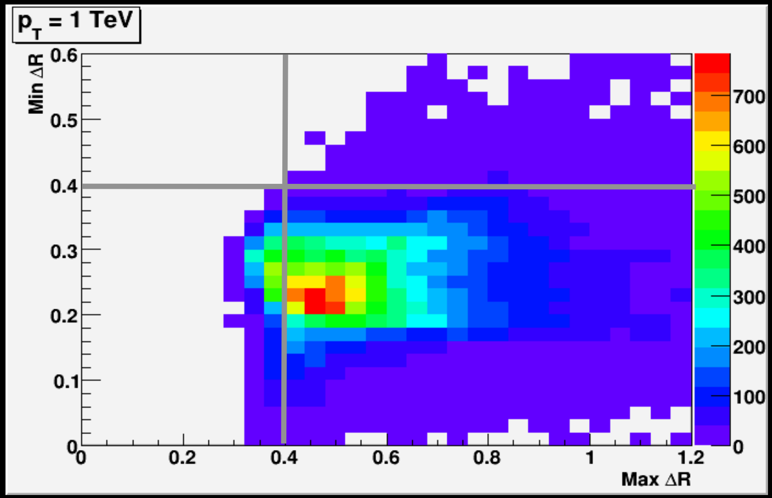
Angular Scales

- *W/Z/h -> qq*
  - $\Delta R > 2m / p_T$
  - *W/Z*: $\Delta R \sim 0.4$ at $p_T \sim 400$ GeV
  - *h*(120): $\Delta R \sim 0.4$ at $p_T \sim 600$ GeV
- Typical LHC jet size
  - $\Delta R \sim 0.4 – 0.7$
- HCAL cells
  - $\Delta R \sim 0.1$
- ECAL cells
  - $\Delta R \sim 0.02$
- Tracker
  - $\Delta R \sim 0.001$

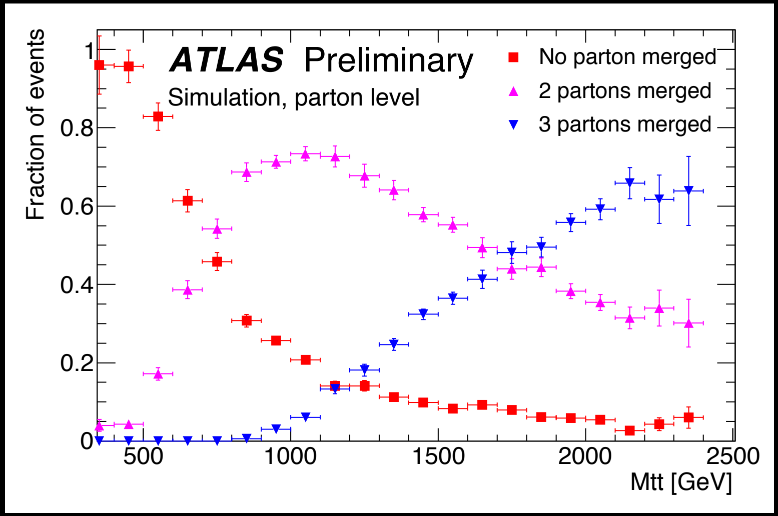


h->qq ΔR distribution in *Z*'(2 TeV) -> *Zh*

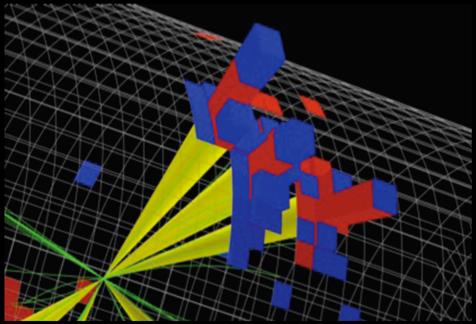
Angular Scales in Top Decay



p_T = 1 TeV, min/max ΔR_{ij}
probability distribution



ATL-PHYS-PUB-2010-008
R = 0.8 anti-kT



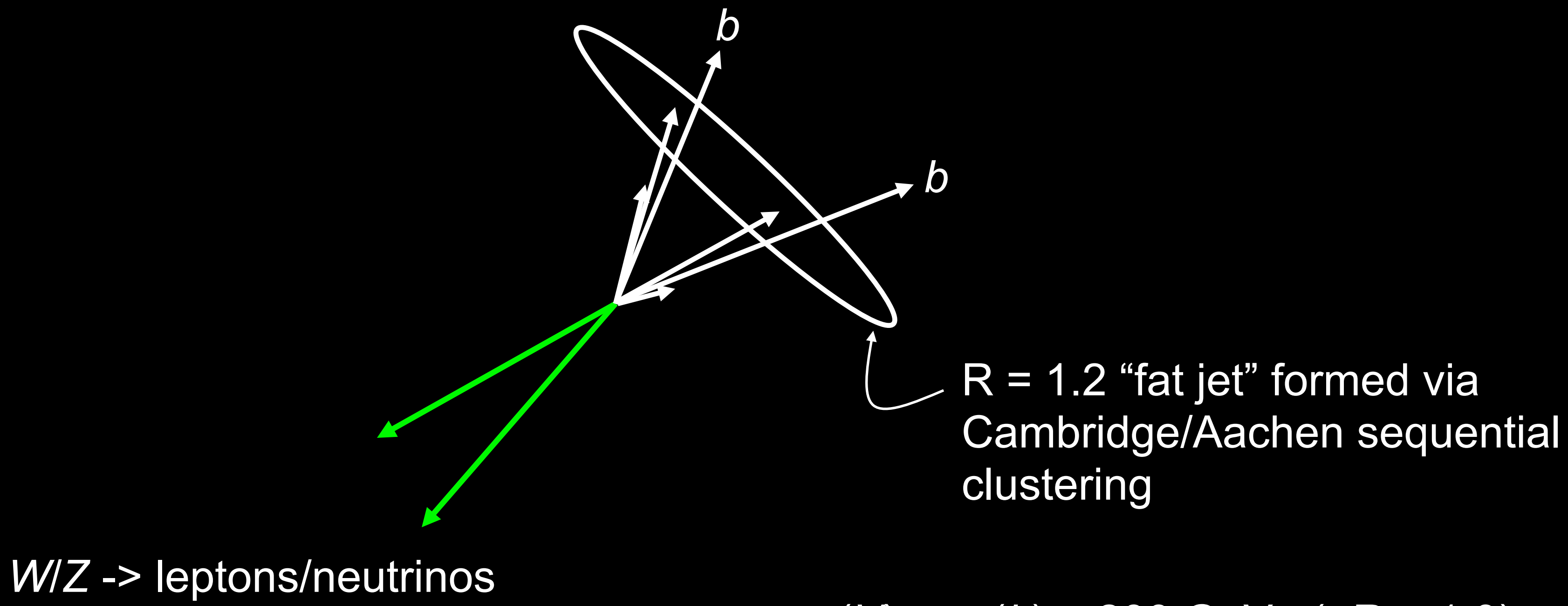
Jet Substructure for Tagging EW Bosons & Tops

- Figure out the relevant ΔR scales adaptively, instead of one-size-fits-all jet clustering
  - R = 0.4~0.7 jets -> variable-size ***subjets*** and/or jet-shapes
  - big-R ***fat-jet*** catches all decay products, substructure tells us where they're going (works for p_T ~ m and upwards)
- Discriminate against QCD parton splittings
  - multibody kinematics at arbitrarily small angles
  - potential to access more subtle aspects of radiation pattern
- Keep the radiation we want, toss the junk
  - ***jet grooming***

Brief History Sketch

- Classic Methods
  - Seymour (1991~1994): k_T-algorithm subjet-finding and HCAL cell-threshold jet-grooming for heavy $h \to WW \to (l\nu)(qq)$
  - Butterworth, Cox, Forshaw (2002); Butterworth, Ellis, Raklev (2007): k_T splitting scales inside *W*-jets in strong *WW* scattering or SUSY cascades
- Popular "Modern" Methods
  - Butterworth, Davison, Rubin, Salam (2008): Recursive, angle-based declustering into subjets with grooming ("filtering") for high-p_T $(W/Z)h$
  - Brooijmans; Kaplan, et al; Thaler & Wang (2008): Top-taggers for ttbar resonances (cluster-decluster, cluster-recluster)
  - Almeida et al (2008): Jet-shapes for tops and EW bosons
  - Ellis, Vermilion, Walsh (2009): "Pruning" reclustering method -- jet groomer and bottom-up substructure organizer
  - Krohn, Thaler, Wang (2009): "Trimming" dedicated jet groomer
  - and now many, many more new approaches, refinements, and applications...

Butterworth, Davison, Rubin, Salam (BDRS)

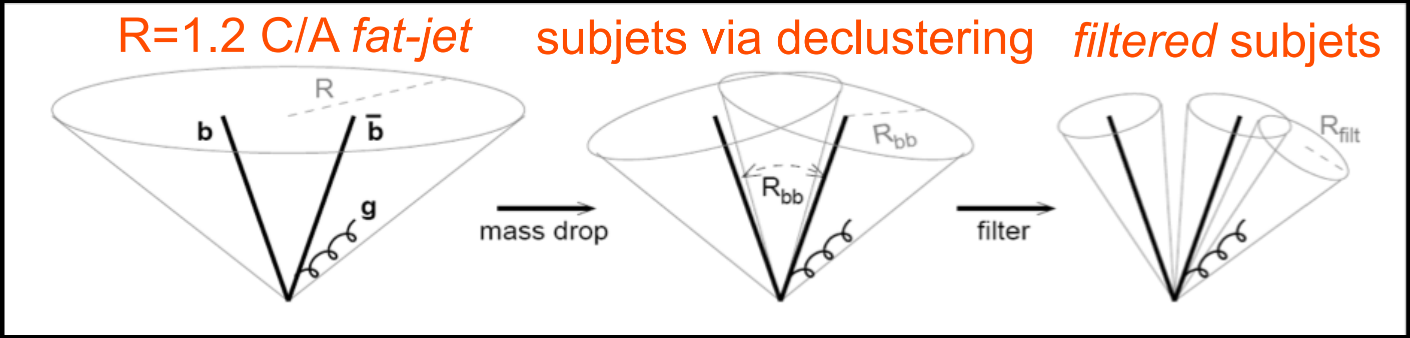


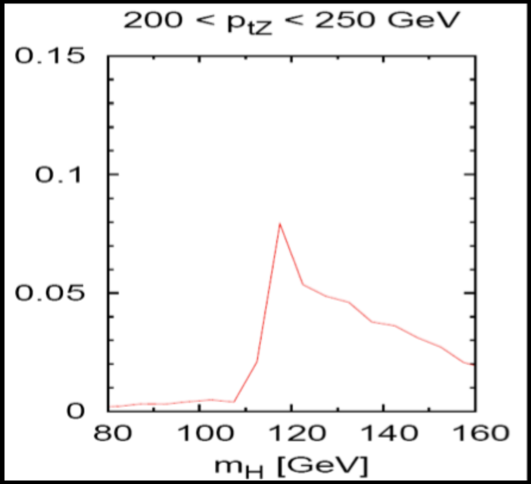
$p_T(V) \sim p_T(h) \sim 200$ GeV   ($\Delta R \sim 1.0$)

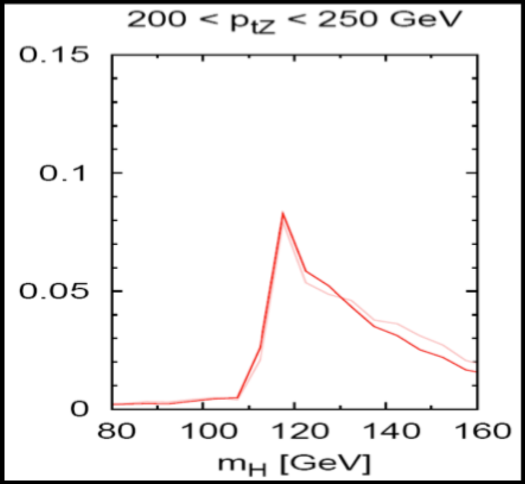
high-p_T kills backgrounds (esp. Zbb, ttbar) faster than signal

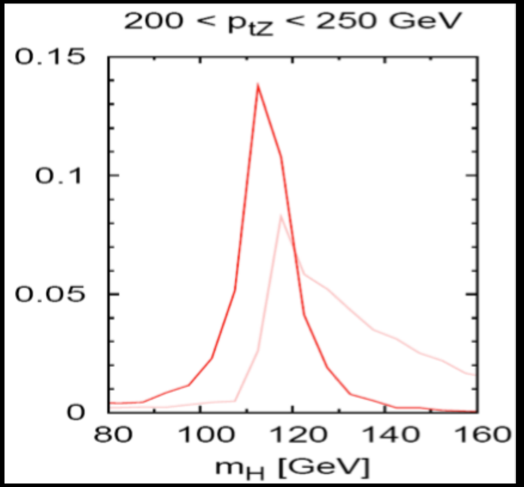
also: Agrawal, Bowser-Chao, Cheung, Dicus, DPF Conf.1994:488-492

Butterworth, Davison, Rubin, Salam (BDRS)





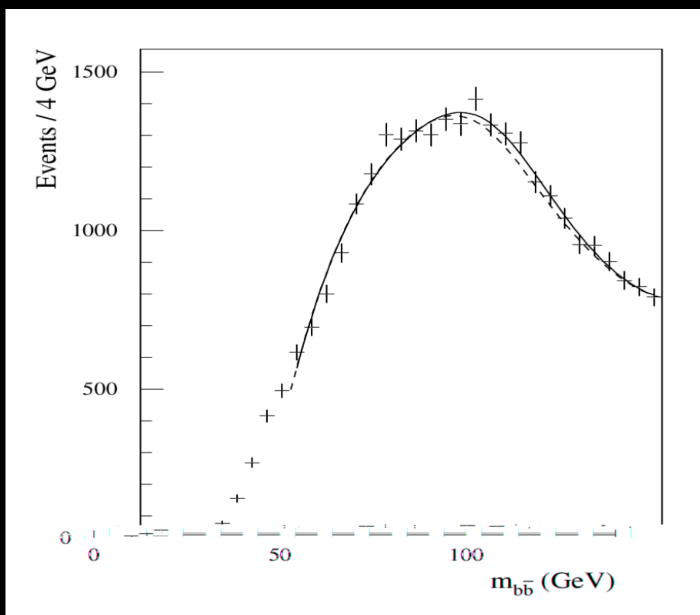




just UE (no pileup)

Butterworth, Davison, Rubin, Salam (BDRS)

Wh -> (*l*ν)(*bb*)

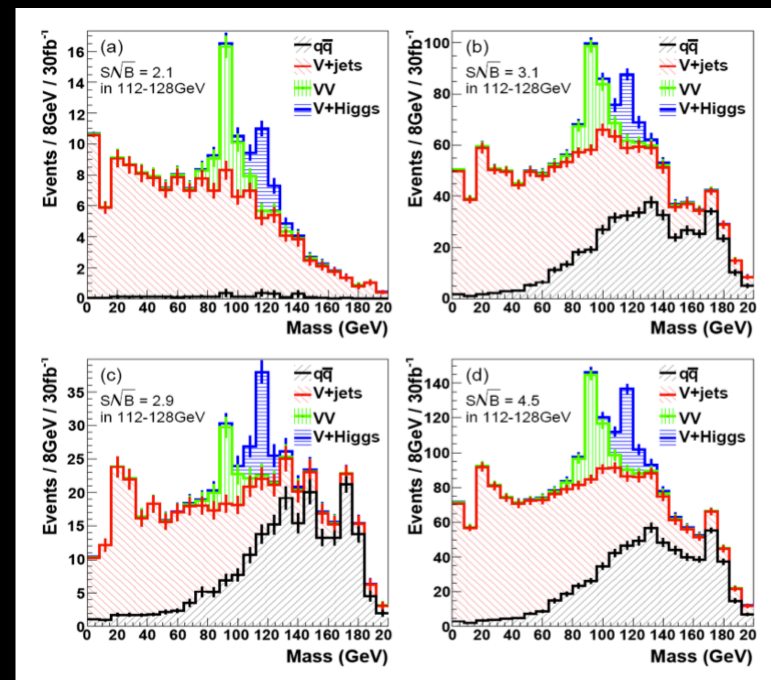


ATLAS TDR

30/fb, m_H = 100 GeV

High p_T →

Zh -> (l^+l^-)(*bb*)   *Zh* -> (νν)(*bb*)

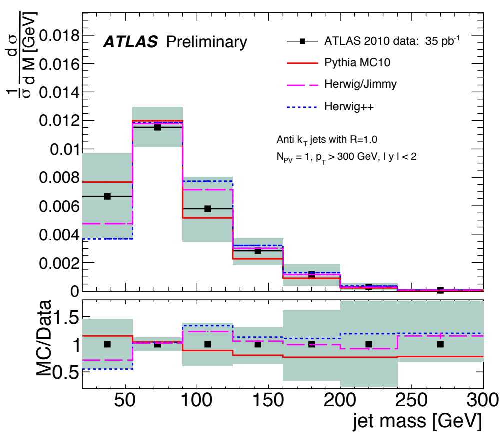


Wh -> (*l*ν)(*bb*)   combination

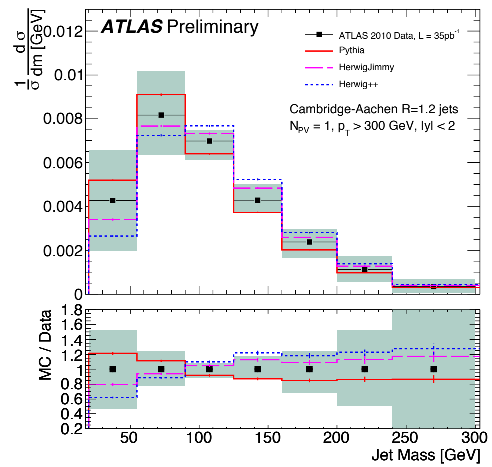
original claim: 4.5σ sensitivity at 30/fb LHC14

more detailed studies: 3σ and change, but still a hot topic for investigation

Proof of Principle in Data: Fat QCD Jets

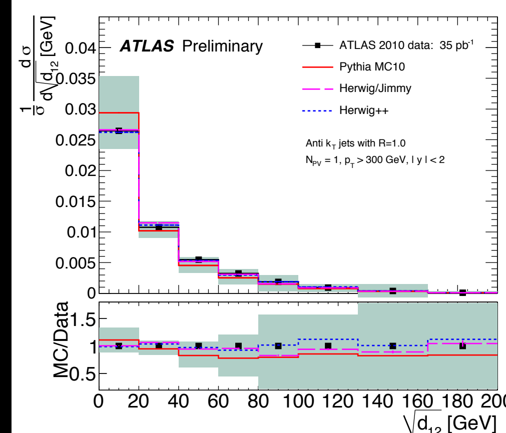


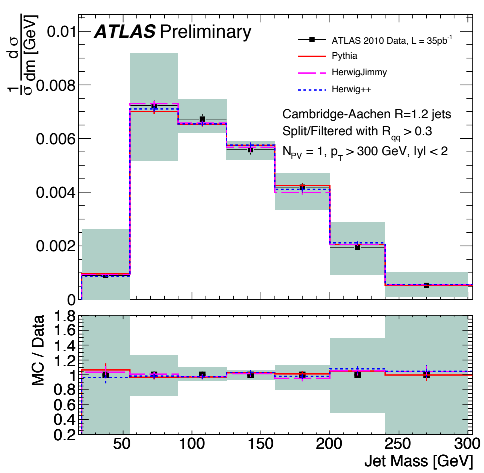
anti-k_t, $R = 1.0$ mass (35 pb^{-1})



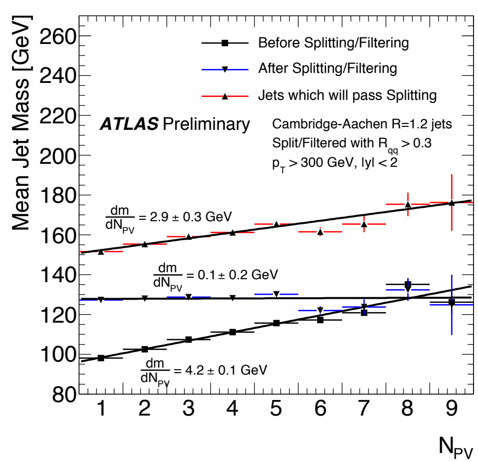
C/A, $R = 1.2$ mass (35 pb^{-1})

$$\sqrt{d_{12}} = \min(p_{T,1}, p_{T,2})\delta R_{12}$$





C/A, $R = 1.2$ (filtered) mass (35 pb^{-1})

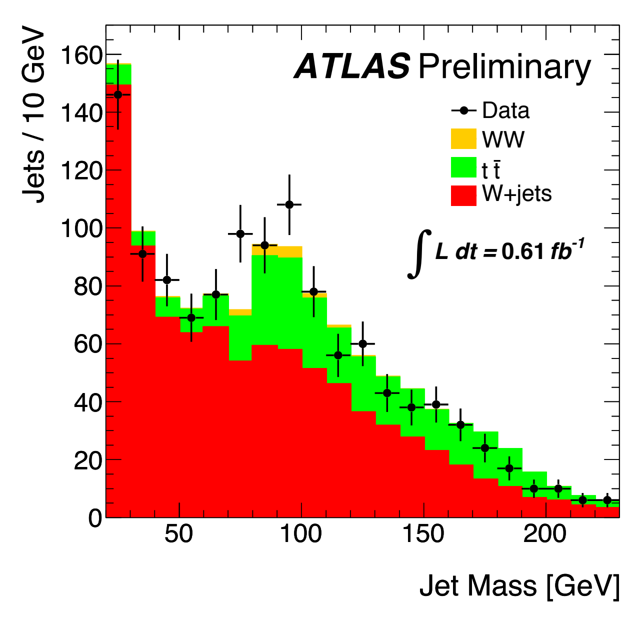


Impact of pile-up on mass w/ & w/o filtering

ATLAS-CONF-2011-073

Proof of Principle in Data: First Hint of Fat W-Jets

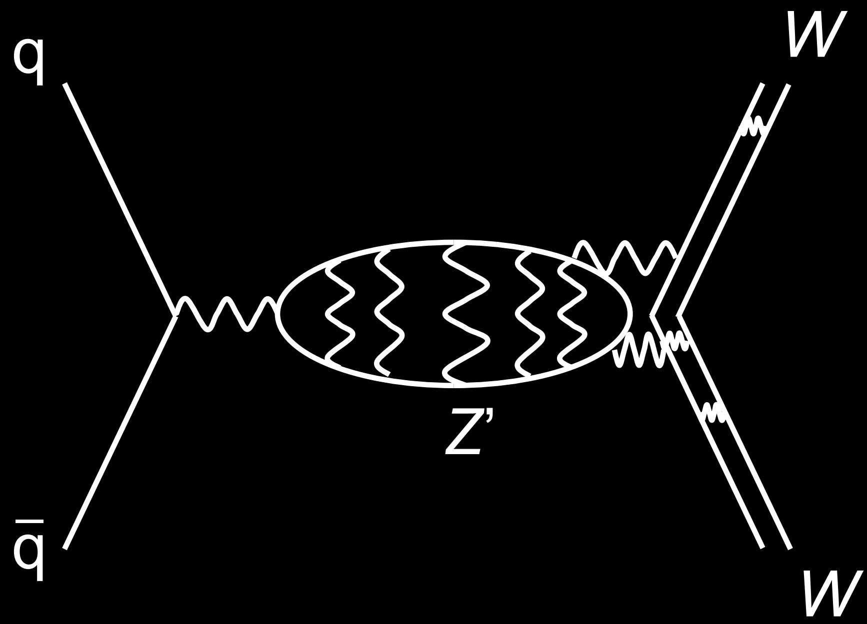
leptonic W + fat-jet events with pT > 180
BDRS procedure without b-tags



Uncorrected jet mass in $W \to l\nu$ events

From David Miller's talk on Thursday
(officially ATLAS-CONF-2011-103)

BDRS on Steroids:
Z' -> Electroweak Bosons



& *Zh*

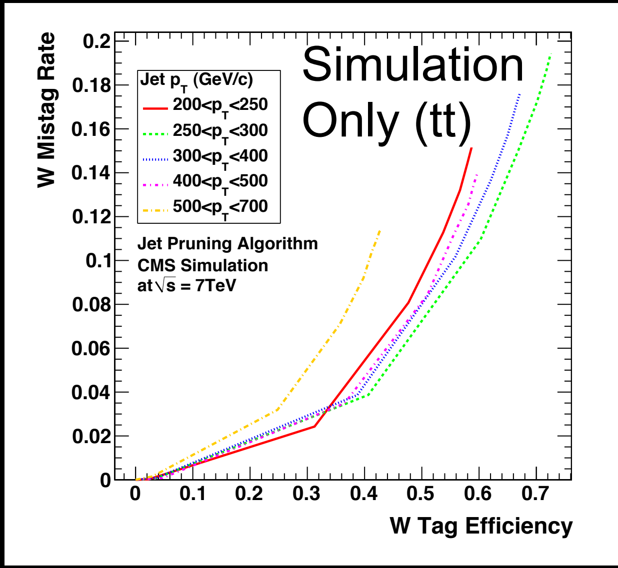
VBF: **Butterworth, Cox, Forshaw, hep-ph/0201098**

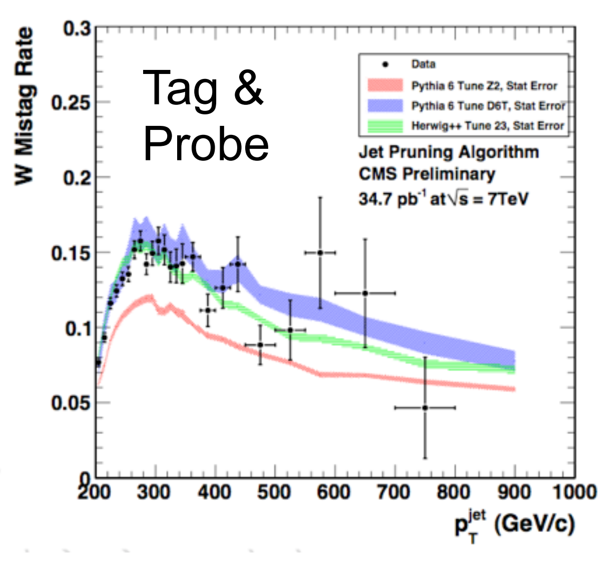
Direct qqbar: **Katz, Son, Tweedie, arXiv:1010.5253**

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

Data-Driven W Mistag (CMS)

- Our own (crude) theory estimate using BDRS method with a +/- 20% mass window suggests:
  - W/Z/h tag rates 60-80%, fairly independent of p_T (but need to fold in ECAL)
  - quark mistag 5-6%, gluon mistag 8-10%, decreasing with p_T due to FSR effects (primitive color discrimination)
- CMS Version uses pruning with BDRS-inspired parameters:



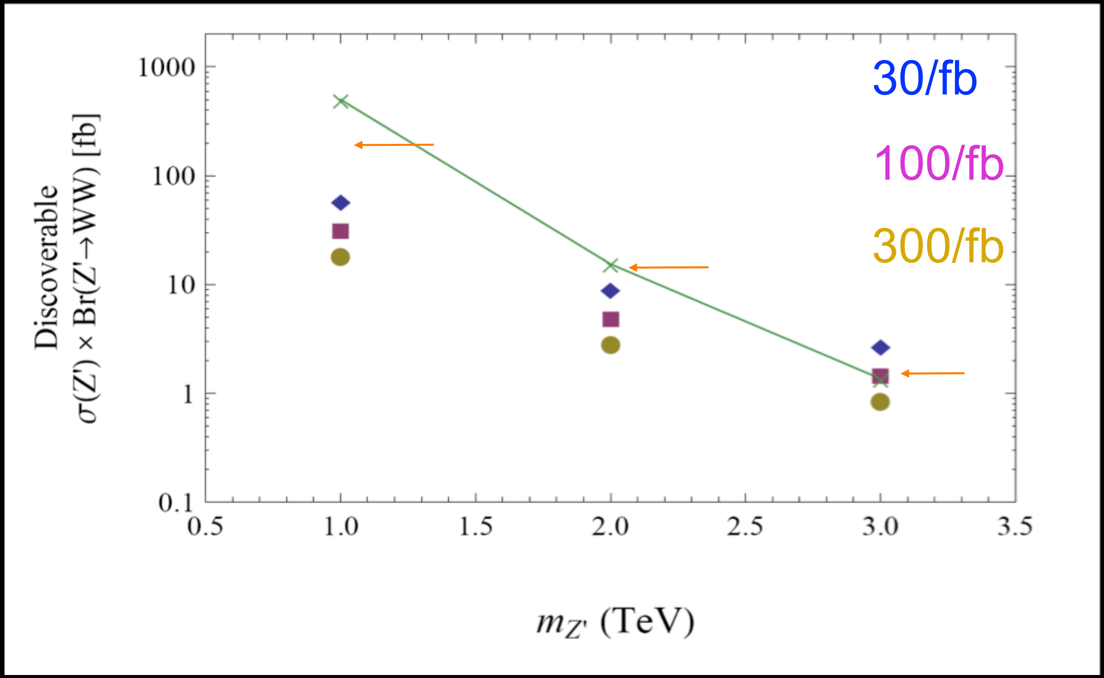


- Note that different quark and gluon mistag rates suggest care should be taken in interpreting “tag-and-probe”

CMS PAS JME-10-013

Z'->WW Discovery Reach (LHC 14)

(Simple counting on simple simulation at LO)



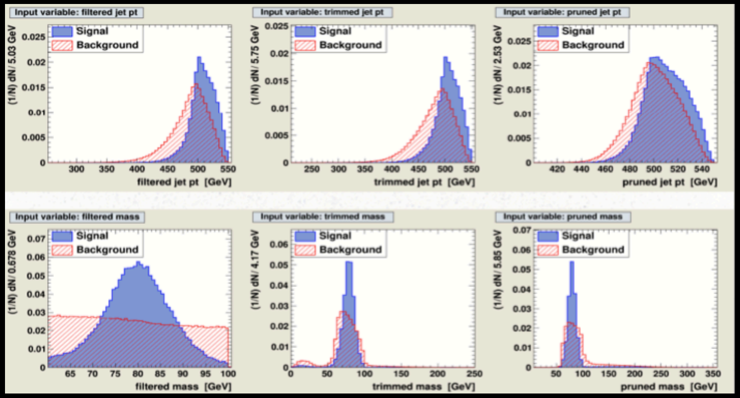
Arrows indicate custodial RS model OR sequential *Z'* model

Line indicates S/B=1

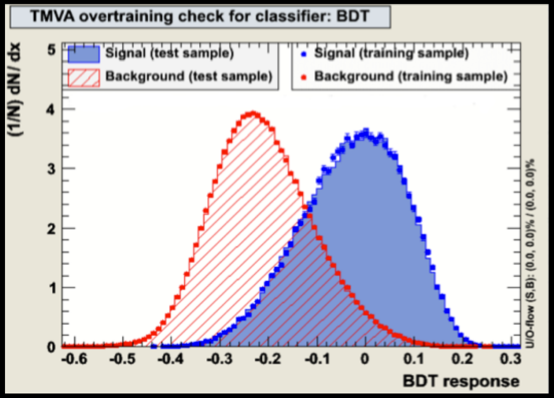
Earlier result (arXiv:0709.0007): Need 1000/fb to reach 3 TeV

Some Ideas for Improvements

- Standard techniques still mainly rely on recovering quasi-2-body kinematics, but radiation pattern for quark/gluon vs boosted EW boson are very different
- Tricky to see by eye or to achive good S/B separation using simple energy flow variables, but multivariate analysis sees something nontrivial (holds up in detector??)
  - e.g., if willing to sacrifice ~half of signal, can ~double statistical significance (i.e., B down by ~16)
- For BDRS boosted Higgs search, also finds basic kinematic cuts that are more powerful than the default hard p_T cut



→



Cui, Han, Schwartz, arXiv:1012.2077

Top Tagging

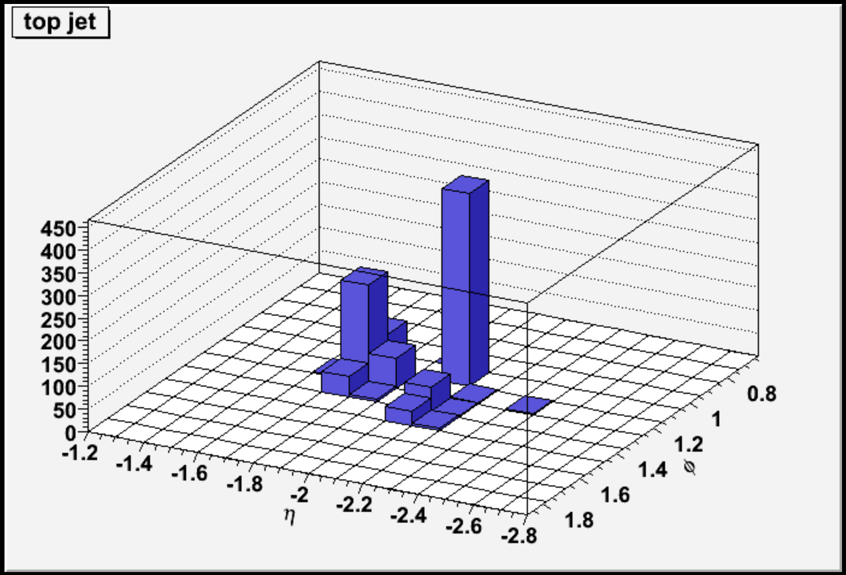
- Tear the jet down one more layer (or rebuild it from bottom-up, e.g. via pruning)
  - 3 or 4 subjets
- Full 3-body decay kinematics
  - subjet pairwise invariant masses (look for the W, veto small-mass pairs)
  - reconstruct top and W decay angles
- Groom as needed
- b-tags???
  - shown to be tricky at high-p_T in a crowded jet, still under investigation
  - muon-based tag is still perhaps an (inefficient) option

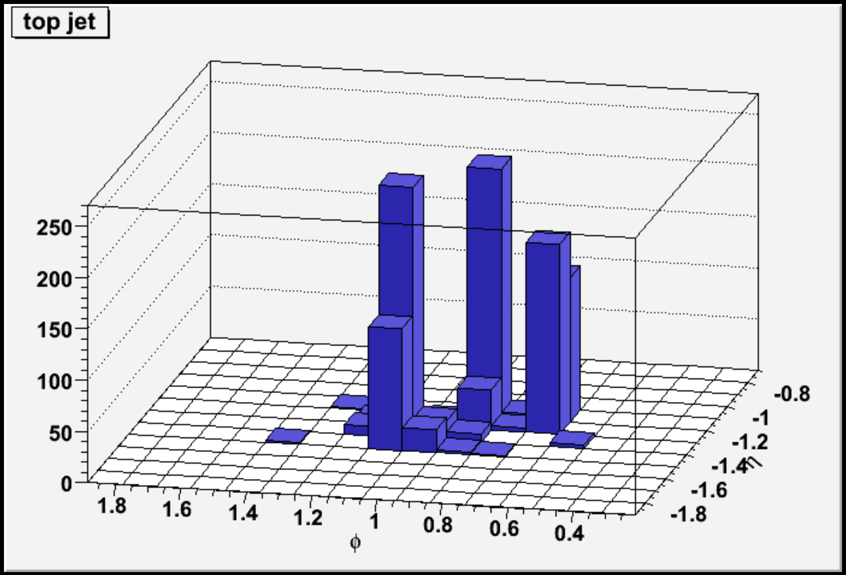
Some Top-Tag Tactics

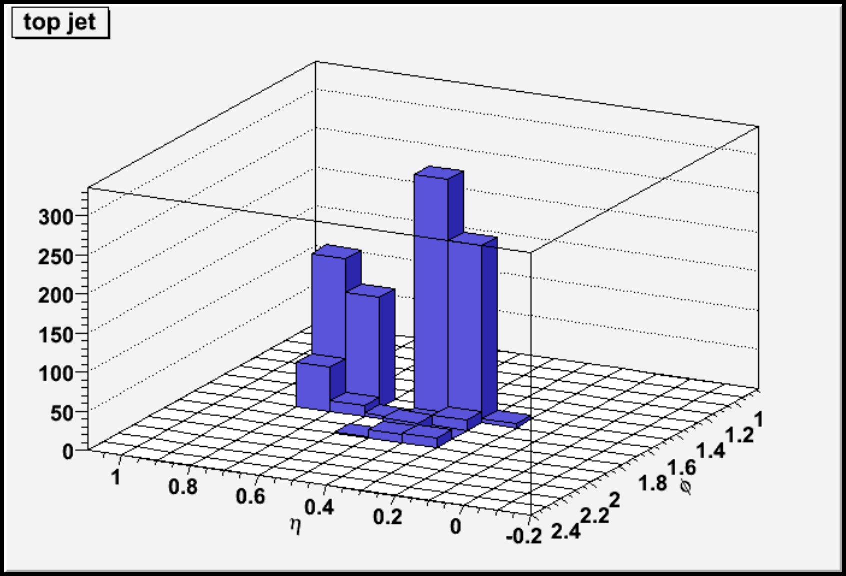
- ATLAS: Evolved from Brooijmans (2008)
  - cluster jets with k_T algorithm, decluster 2 or 3 stages and study mass/splitting scales
- Thaler & Wang
  - cluster jets with anti-k_T algorithm, exclusively recluster with k_T into 3 “subjets” and apply multibody kinematic cuts
- Hopkins/CMS: Evolved from Kaplan, Rehermann, Schwartz, Tweedie (2008)
  - cluster jets with Cambridge/Aachen algorithm, decluster recursively until 3 or 4 subjets are found and apply multibody kinematic cuts
- Jet Shapes: Almeida, Lee, Perez, Sterman, Sung, Virzi
  - angularities, planar flow, etc
- Pruning: Ellis, Vermilion, Walsh
  - selective jet clustering removes junk and self-organizes substructure simultaneously
- HEP Tagger: Plehn, Spannowsky, Takeuchi, Zerwas
  - decluster into arbitrary # subjets, sophisticated kinematic discrimination
  - works with for large top-jets with additional activity inside
- Template Overlap: Almeida, Lee, Perez, Sterman, Sung
  - calorimeter cell pattern -> multidimensional vector
  - check dot products with ensembles of template top-jets and QCD-jets
- N-Subjettiness: Thaler and Van Tilberg
  - continuous scores assigned for mono-subjet-like, di-subjet-like, tri-subjet-like, etc
- Dipolarity: Hook, Jankowiak, Wacker
  - improved discrimination using observables sensitive to color connections
- Correlation function lineshape: Jankowiak and Larkoski
  - look for sudden jumps in the multibody correlator wrt angle

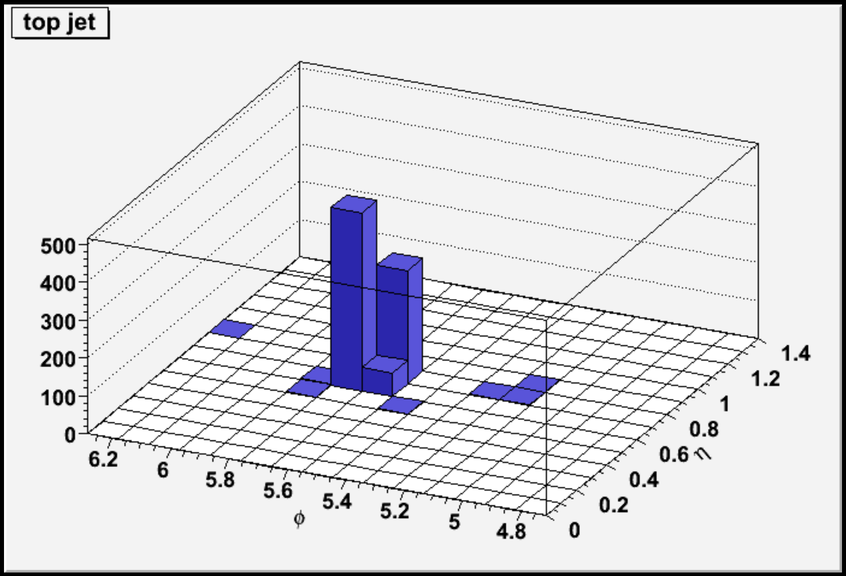
* idealized 0.1 x 0.1 calorimeter

1 TeV Top-Jet Gallery

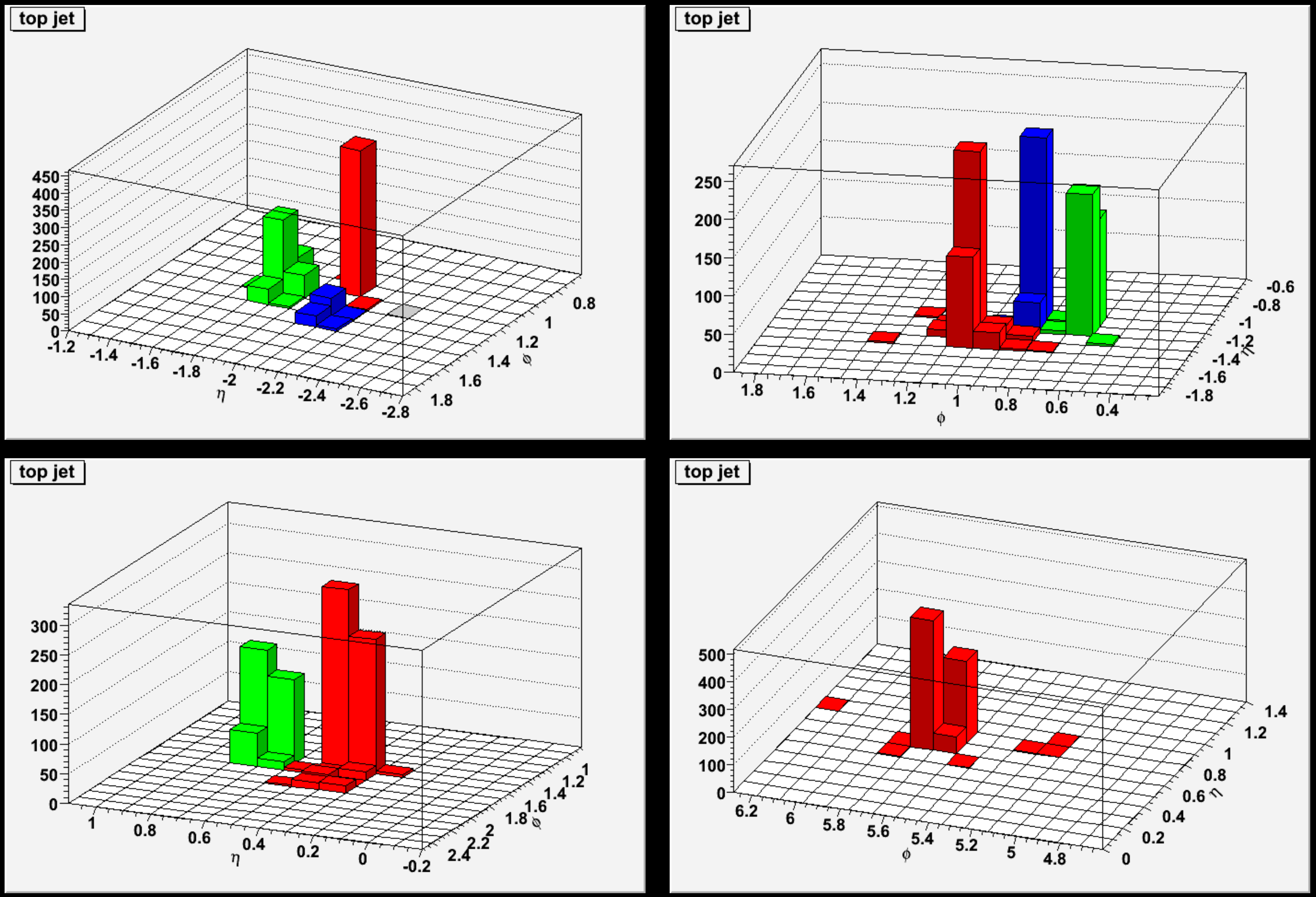




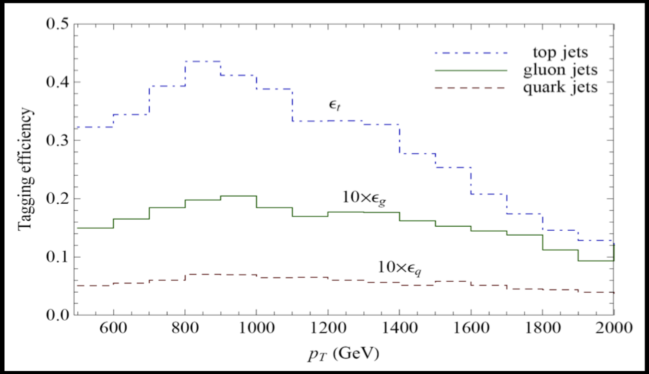




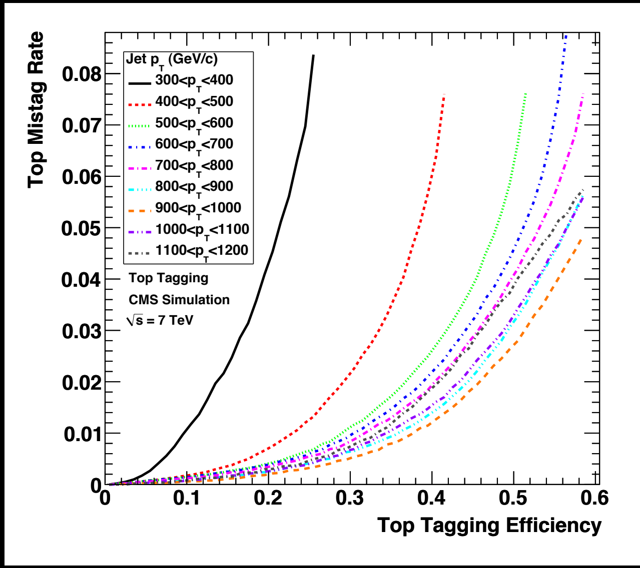
1 TeV Top-Jet Gallery



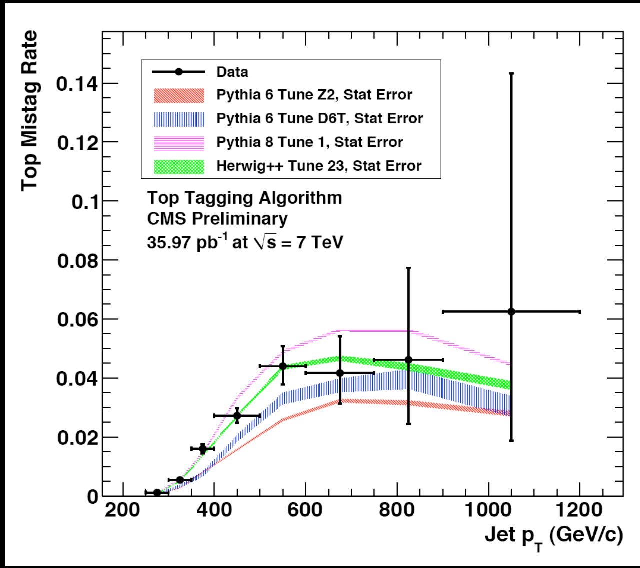
Tag/Mistag Rates



Hopkins top-tagger on our simple theorists' simulation



CMS tag/mistag on full simulation

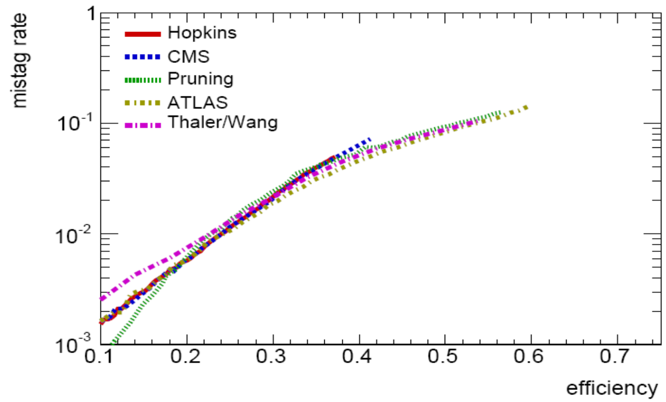


CMS tagger on data via tag & probe

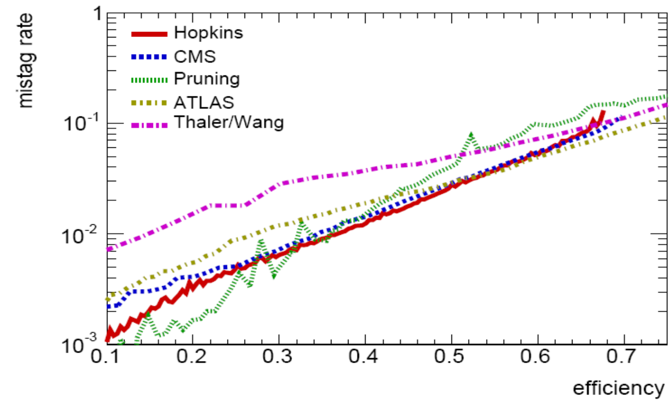
CMS PAS JME-10-013

Performance Comparison

BOOST 2010, arXiv:1012.5412

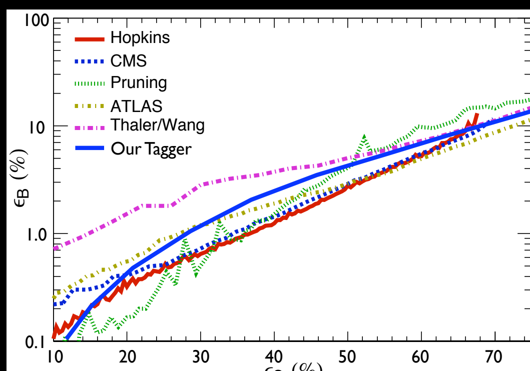


(c) 300–400 GeV

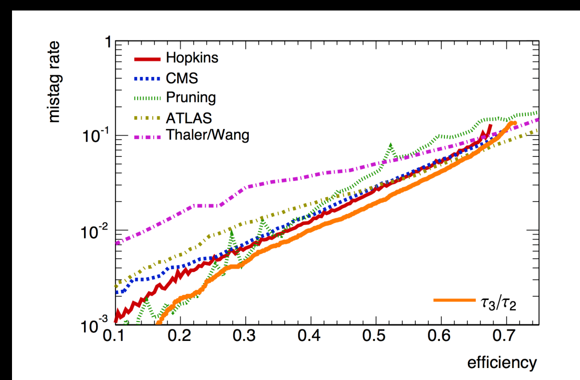


(d) 500–600 GeV

(back to perfect 0.1x0.1 calorimeter)



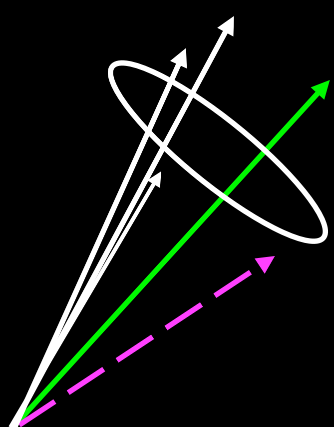
correlator lineshape

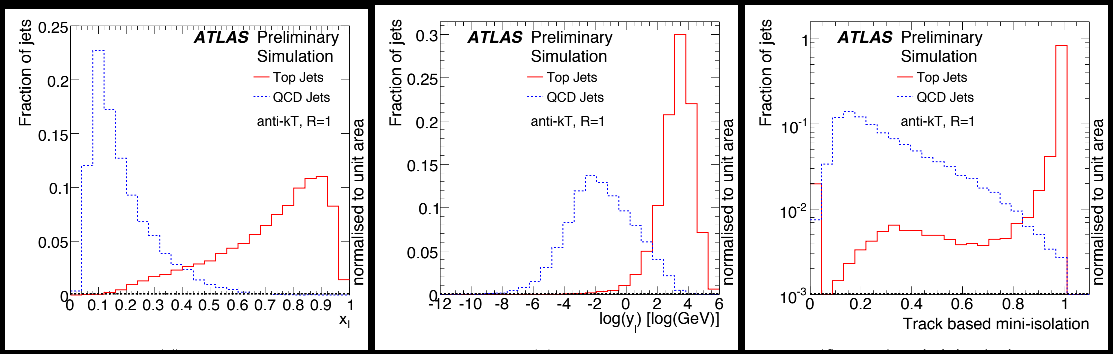


N-subjettiness

The Semileptonic Option

- Lepton-inside-of-Jet is in principle much cleaner than Jets-inside-of-Jet!
- Studies by theorists and experimentalists indicate that backgrounds can still be powerfully rejected
  - QCD becomes far subdominant in tt spectrum, even with no b-tag
- Combined with hadronic top-tagging, ultimate LHC14 sensitivity to RS g’ maybe up to ~5 TeV





ATL-PHYS-PUB-2010-008

Summary

- Lots of ideas, lots of progress for turning boosted EW bosons and tops into taggable objects for TeV-scale new physics searches, as well as the light Higgs search
- Basic elements of tags are now being validated *on data*
- Theory remains a bit floppy (lots of work in progress that I didn't talk about), but so far experimental philosophy has been in-situ calibration of taggers
  - detailed kinematic distributions for BDRS-type search look remarkably good
  - I'll re-emphasize that quark and gluon jets can behave very differently
- Looking forward to lots of interesting searches in the near-term and long-term future!!