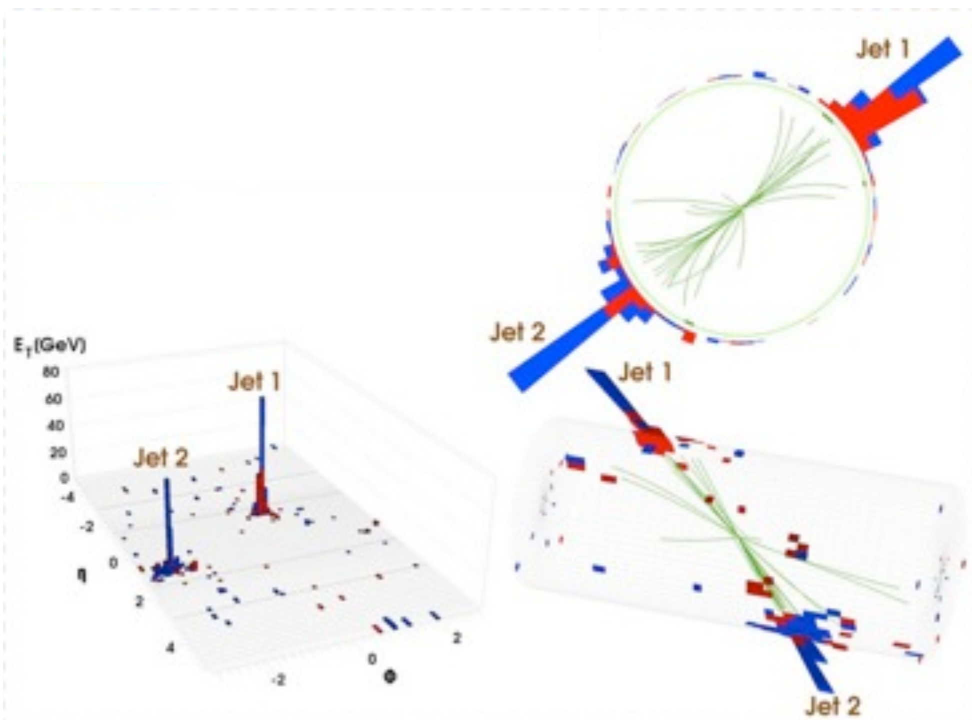
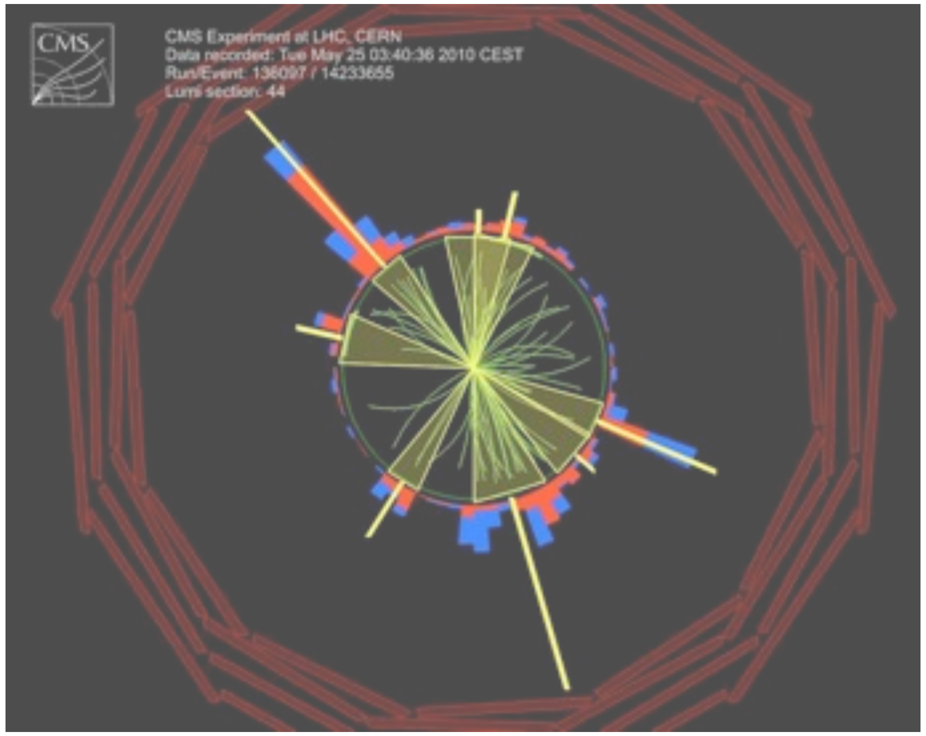
Jet Substructure at the LHC:
From BOOST to Boston and Beyond





Christopher Vermilion
Boston Jet Workshop
1/14/2011

Outline

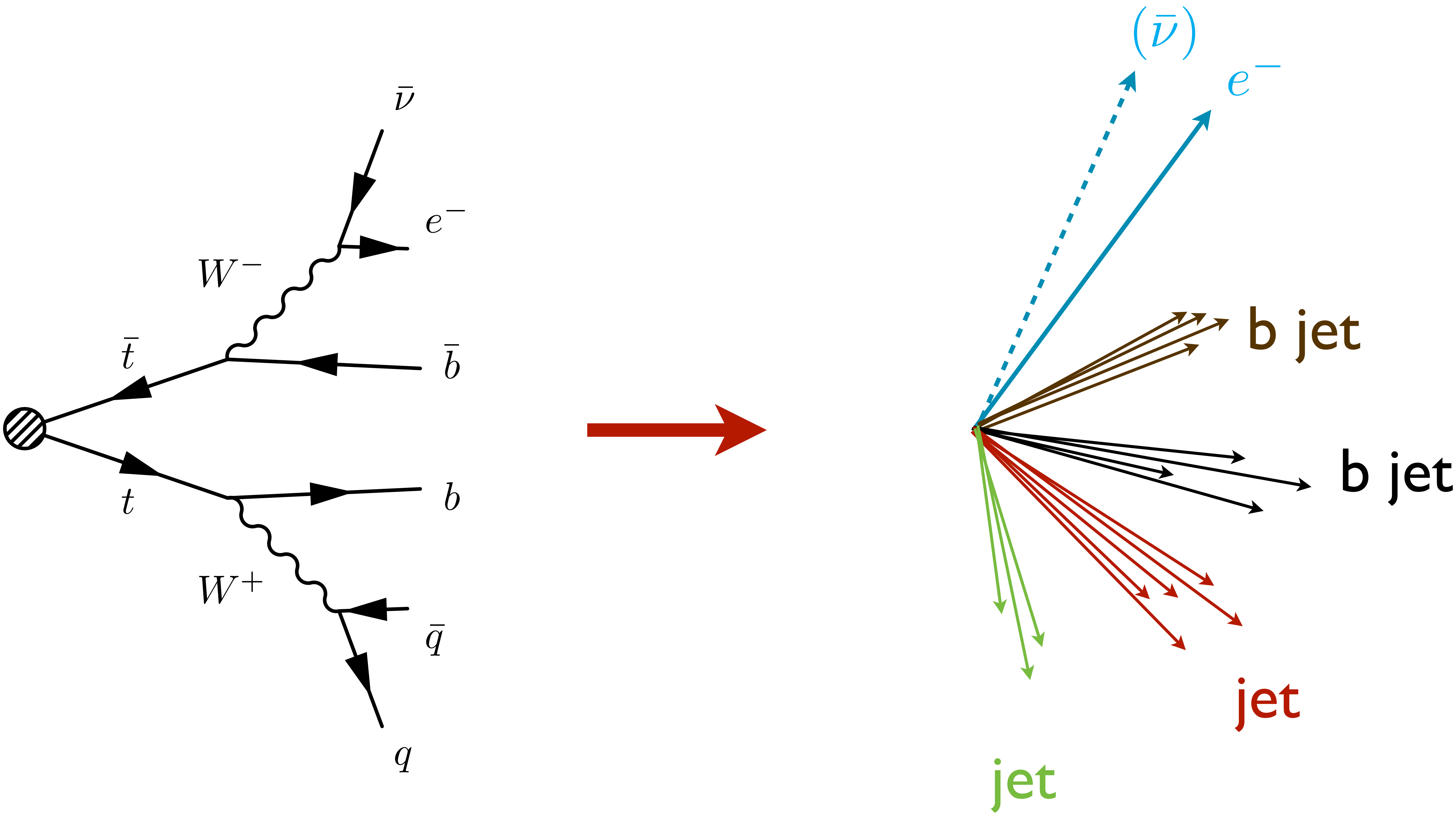
- A short history of jet substructure techniques
- Results from BOOST 2010
- What's next? (BOOST 2011 homework...)

Outline

- A short(ish) history of jet substructure techniques
- Results from BOOST 2010
- What's next? (BOOST 2011 homework...)

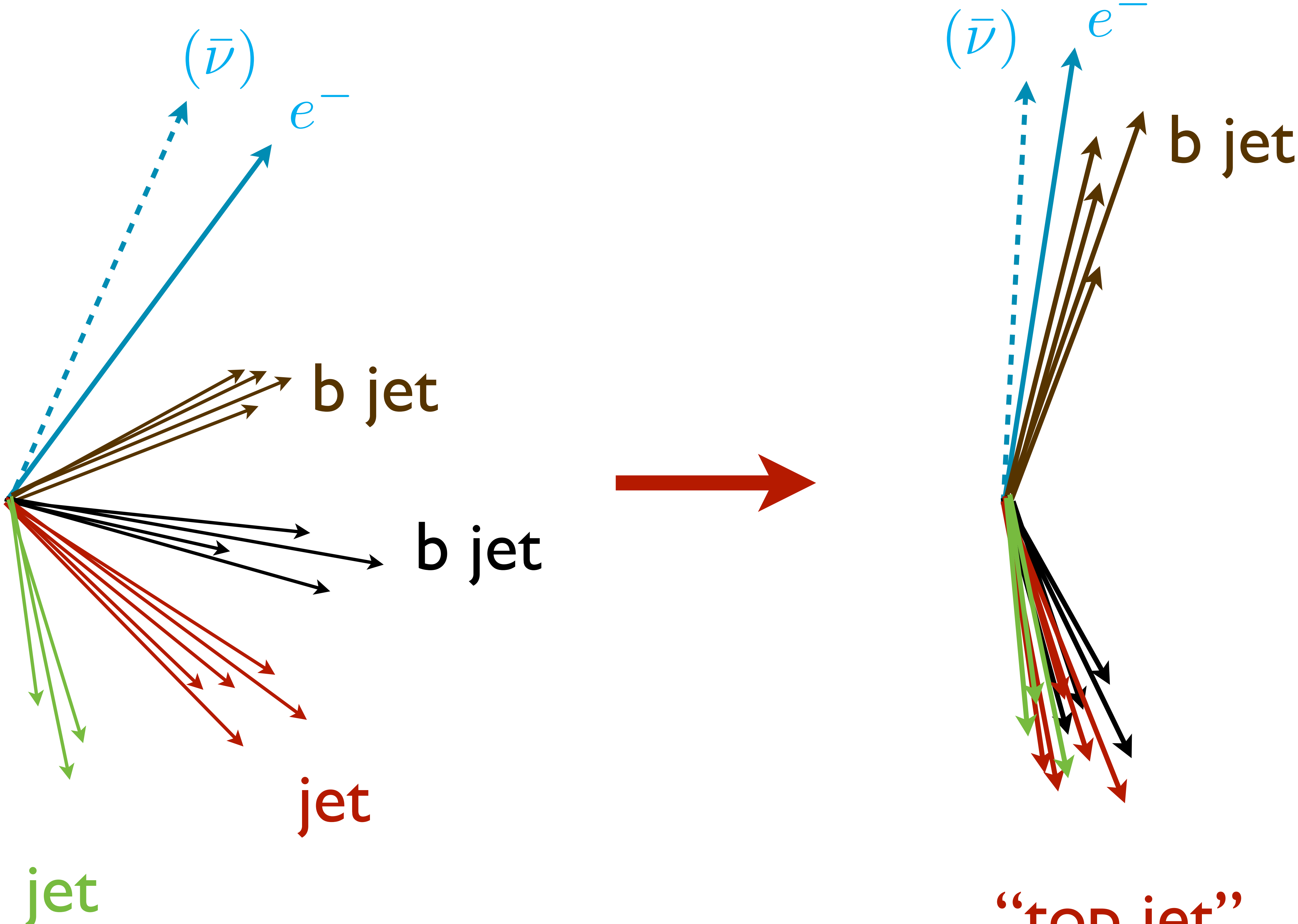
Why do we care about jets?

Anything that decays to quarks and gluons will produce jets!



Lots of interesting new, old particles have decays including jets!

High-p_T: heavy particles in single jets



How are QCD and heavy particle jets different?

- Jet mass
  - Heavy particles: $m_{jet} \sim m_X$
  - QCD: $m_{jet} \sim \# \, p_T$
- Subjet kinematics
  - Heavy particle: roughly symmetric subjets
  - QCD: typically hard core with diffuse radiation; asymmetric subjets
- Color flow
  - Heavy particle: if color singlet, decay products are color connected
  - QCD: often contains color connections to rest of event

Also: spin, charge

All jet substructure methods use some combination of these properties!

Problem 1: algorithm biases

- Jet algorithm shapes substructure, distorts kinematics you might expect from a decay
- One solution: don't use subjets!

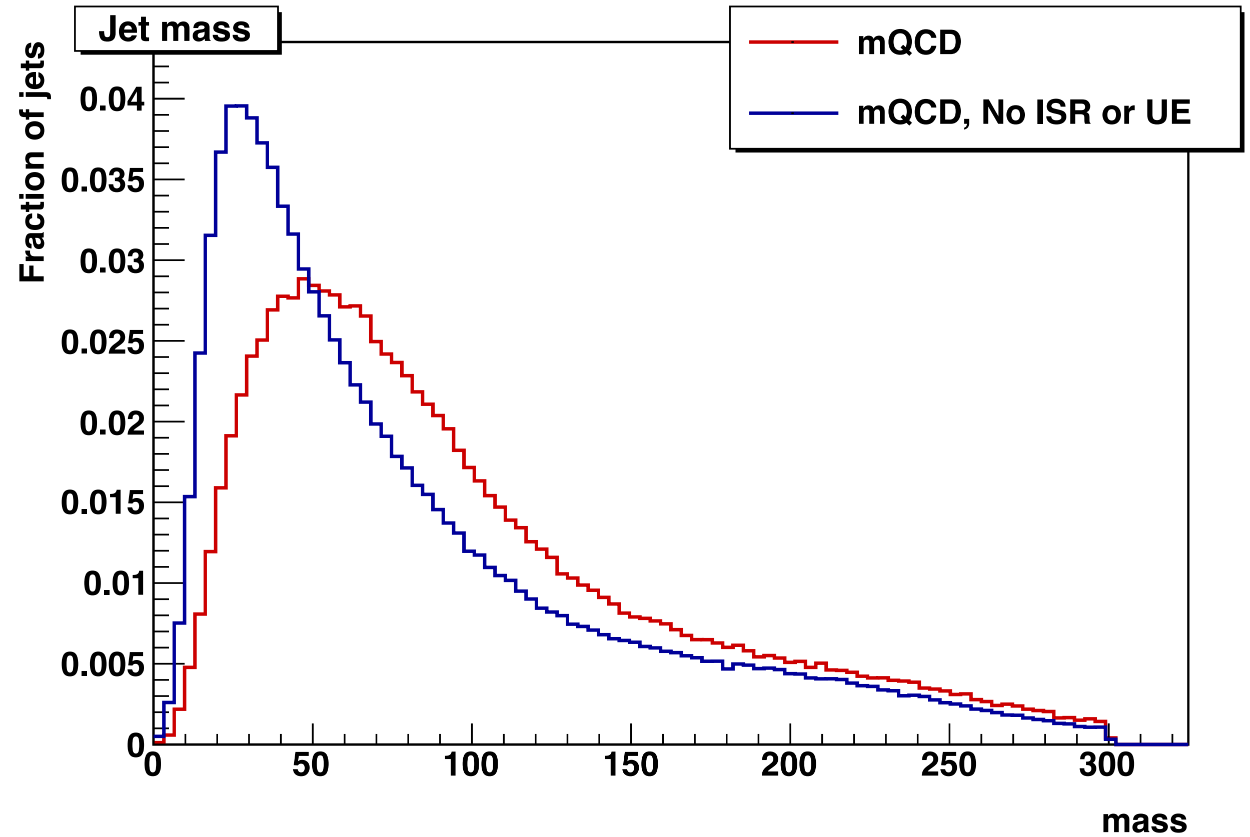
→ Jet shapes

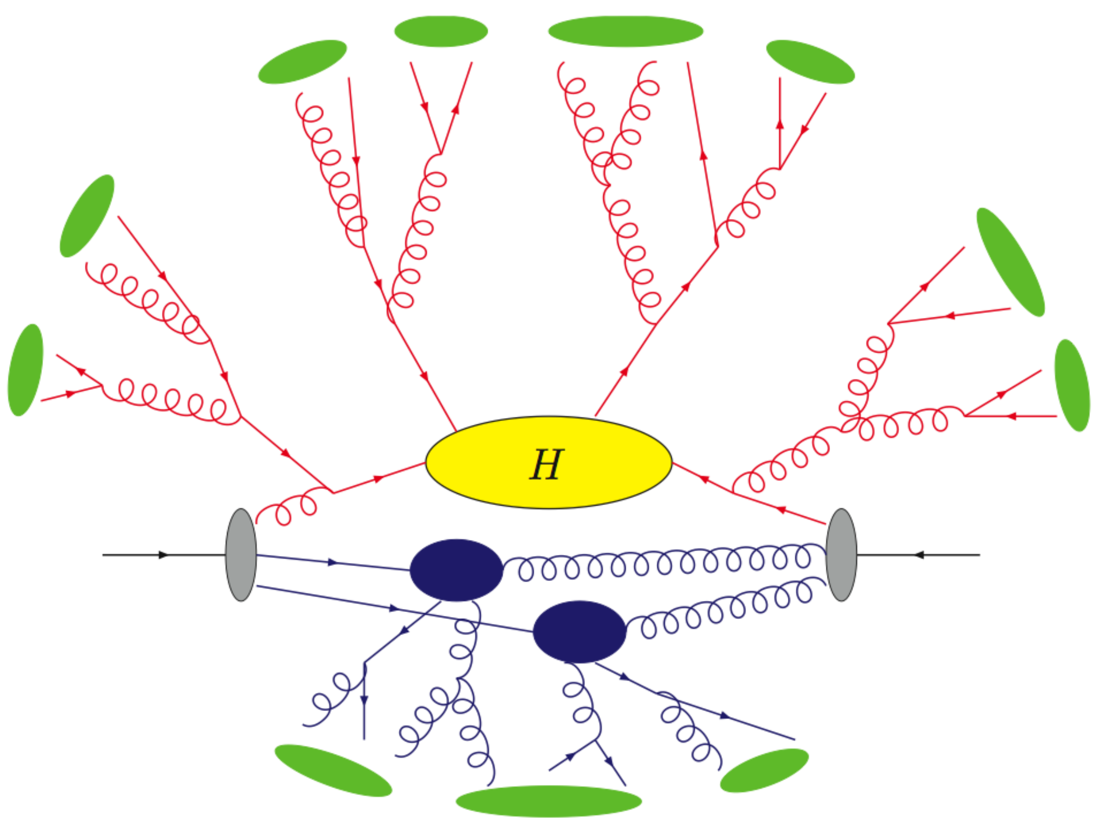
- Other solution: .....?

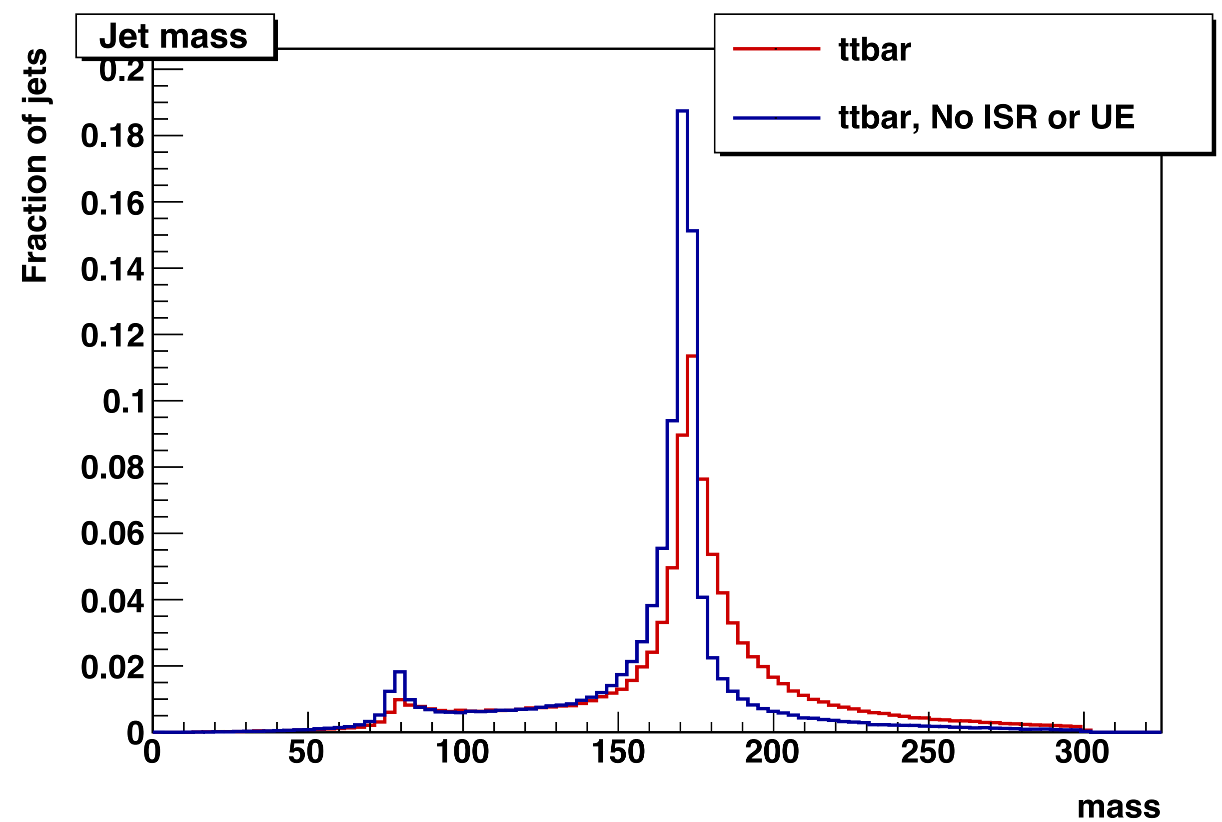
Problem 2: Splash-in distorts masses

pp collisions:
extra radiation from ISR, UE, PU

Shifts mass distributions upward, broadens peaks!







Jet substructure: "the early days"

- Seymour (Z. Phys. **C62** (1994) 127): boosted Ws
  - jet mass, subjet separation angle, filtering, jet areas, variable R
- Butterworth, et al. (hep-ph/0201098, hep-ph/0702150): boosted Ws
  - y_{cut} -- subjet separation in k_T
- Butterworth, Davison, Rubin, Salam (0802.2470): boosted Higgs
  - **Mass drop**: $m_{subjet} < \mu\ m_{jet}$ (if not, discard soft subjet and repeat)
  - **Filtering**: recluster with smaller R, keep 3 hardest subjets
  - Related: 0906.0728 (neutralinos)

Extension to 2-step decays (tops)

- Brooijmans (ATL-PHYS-CONF-2008-008): "**Y-Splitter**"
    - k_T measures for last three mergings in k_T jets
- Kaplan, Rehermann, Schwartz, Tweedie (0806.0848): "**Top tagging**"
    - Identify hard splittings by discarding soft, wide-angle branchings
    - Find W in subjets, then cut on top, W masses
    - CMS variant: 0909.4894* (*see Jim Dolen's talk yesterday)
- Thaler, Wang (0806.0023)
    - Several substructure variables, mostly energy sharing (z)

Generic methods: "Jet grooming"

- Ellis, CV, Walsh (0903.5081, 0912.0033): "**Pruning**"
    - Similar to first step in top-tagging, but bottom up
    - Remove soft, large angle mergings as you go
    - No attempt to find a specific number of subjets
- Krohn, Thaler, Wang (0912.1342): "**Trimming**"
    - Adaptive filtering
    - Recluster with small R, keep subjets with $p_T^i > f\, p_T^{jet}$

(Filtering can also be put in this category)

Jet shapes -- more general than subjets

Generically, jet shape = f($\{p_T{}^i\}$)

- Almeida, et al. (0807.0234, 0810.0934)
  - Several jet shapes for QCD, top jets
  - Mass, "**planar flow**"* (*see several talks this workshop)
- Chekanov, Proudfoot (+Levy, Yoshida) (1002.3982, 1009.2749)
  - Eccentricity and related geometrical measures
- Almeida, et al. (1006.2035): "**Template Overlap**"
  - Find some set of variables that characterizes signal (tops)
- Kim (1011.2268); Thaler, Van Tilburg (1011.2268): "**N-Subjettiness**"**
  - Smooth interpolation between N subjets

(**see talks by Kim, Van Tilburg)

Color flow

- Gallichio, Schwartz (1001.5027): "**Pull**"*

(*see Andy Haas's talk)

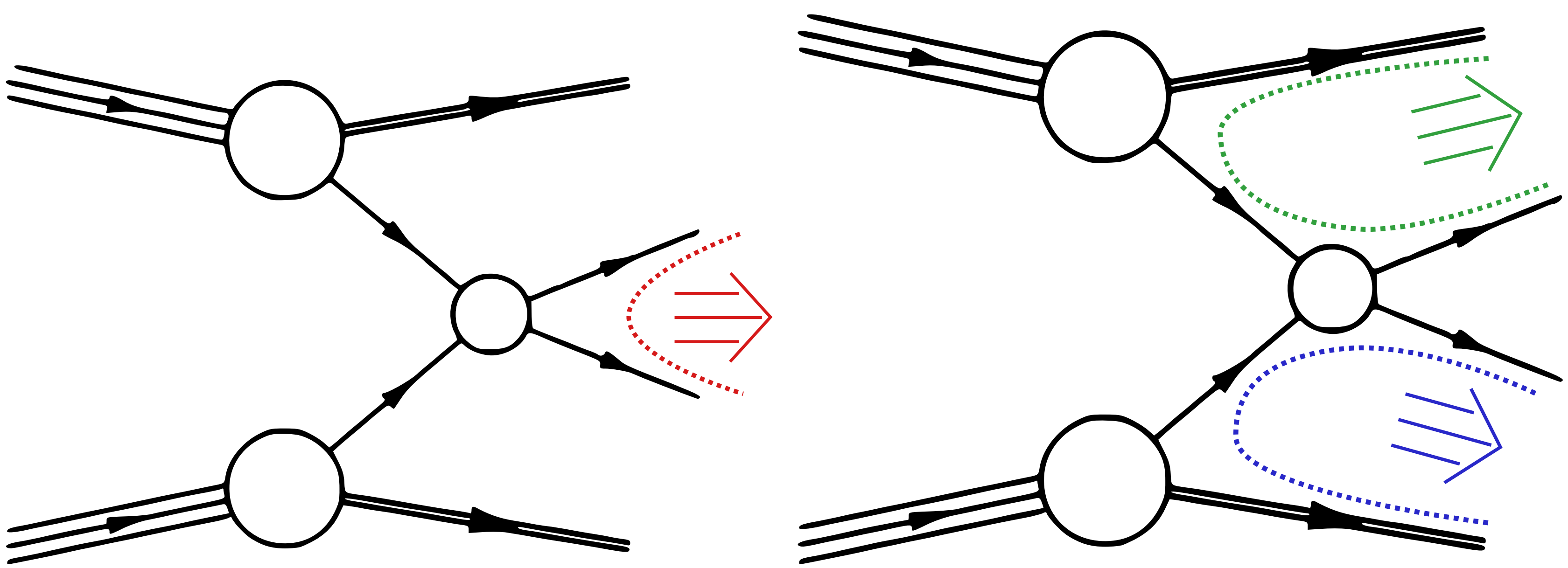


FIG. 1: Possible color connections for signal ($pp \to H \to b\bar{b}$) and for background ($pp \to g \to b\bar{b}$).

Applications

- 0909.3855: Polarization of boosted tops
- 0910.5472: *tth* with filtering+
- 0912.4731: Higgs searches with filtering+
- 1005.0417: Combining filtering, pruning, trimming on H/t
- 1006.1151: H -> 2X -> 4g
- 1006.1650: "Unburied Higgs"
- 1006.1656: MSSM Higgs
- 1006.2833: Stops with tops
- 1006.3213: Boosted tops from UED
- 1007.2221: Boosted semileptonic tops
- 1010.0676: Semileptonic ZZ
- 1008.2202: Higgs -> semileptonic ZZ
- 1010.3698: Multivariate H+W/Z
- 1010.5253: Z' -> W/Z
- 1011.4523: Ditau, boosted Higgs

(June 2010: BOOST)

More on multivariate later...

Multivariate approaches

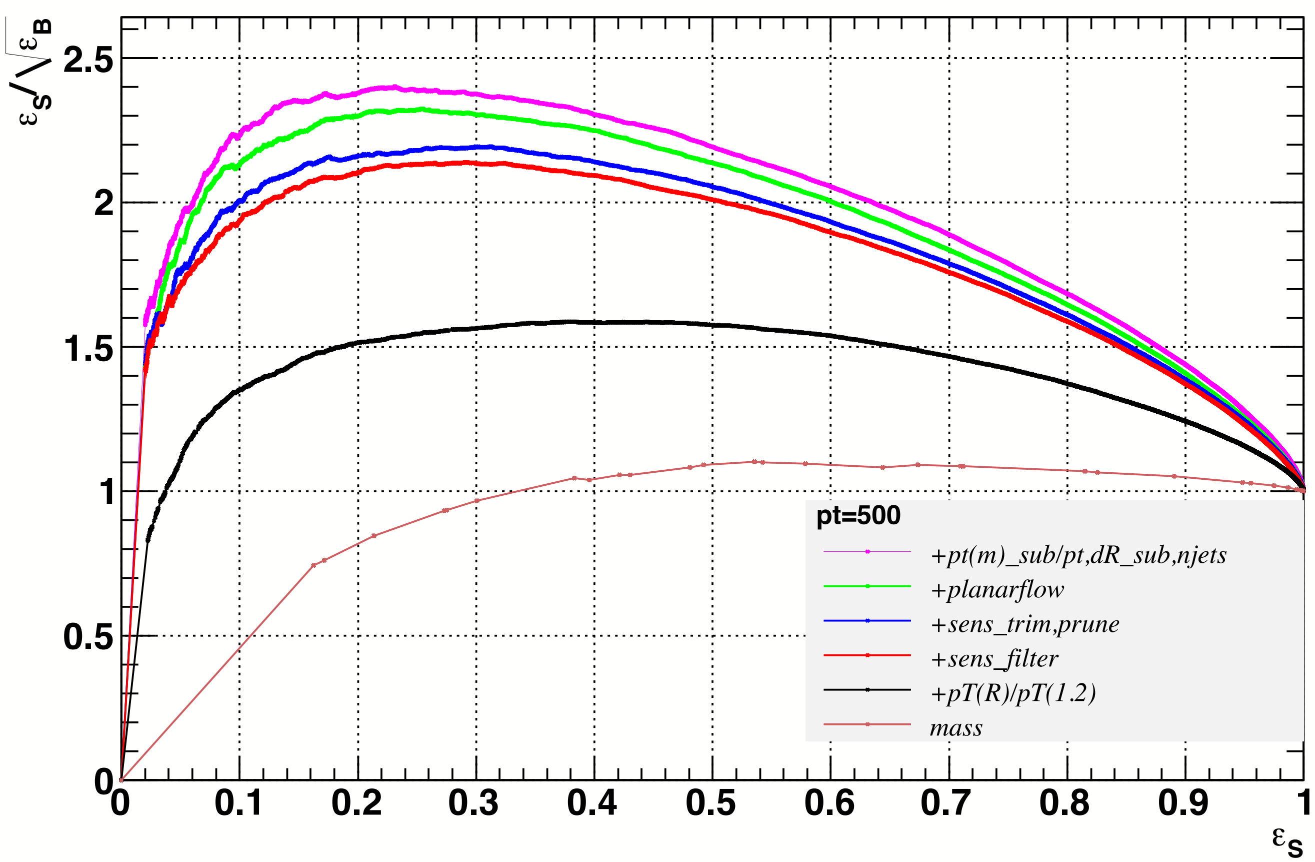
Cui, Han, Schwartz: 1012.2077 -- “W-jet Tagging” (builds on 1005.0417, 1010.3698)

Why not throw everything at the wall and see what sticks?

TMVA finds optimal set of **25 variables:**

$$m_{\rm jet},\ c_{p_T}(0.2-0.11),\ {\rm sens}^{m,p_T}_{\rm filt,trim,prun},\ P_f,\ P_f(0.4),\ \frac{p_T^{\rm sub1,sub2}}{p_T},\ \frac{m^{\rm sub1,sub2}}{m},\ \Delta R_{\rm sub},\ n_{\rm sub}$$

New: “R-Cores”



See talk by Zhenyu Han today!

Outline

- A short history of jet substructure techniques
- Results from BOOST 2010
- What's next? (BOOST 2011 homework...)

Proceedings: 1012.5412

First accomplishment: common samples

http://www.lpthe.jussieu.fr/~salam/projects/boost2010-events/ (and UW mirror)

Signal is all-hadronic ttbar; BG is dijet, generated with:

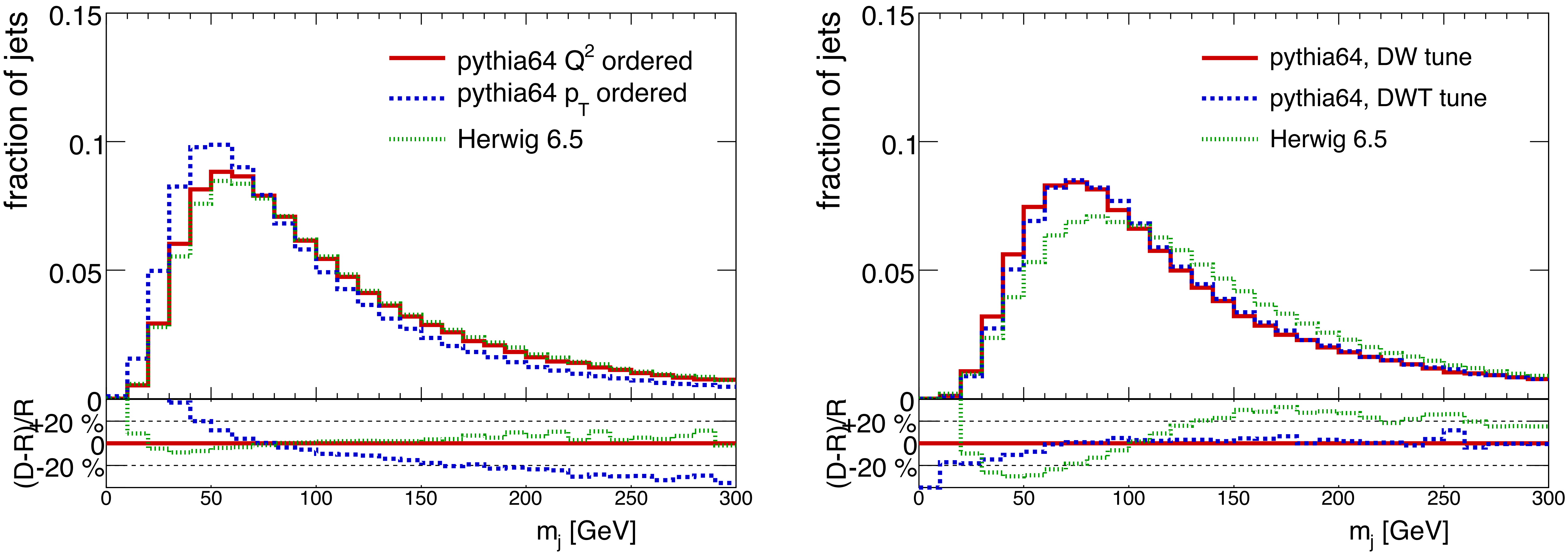
HERWIG 6.510 + JIMMY (ATLAS tune)
Pythia 6.4 x DW, DWT, Perugia0 tunes (Perugia0 is p_T-ordered)

Narrow p_T bins: {200-300, 300-400, ... 1500-1600} GeV

Even more remarkable -- common jet definition:

- anti-kT R=1.0 jets
- no detector
- at most 2 jets
- pT > 200

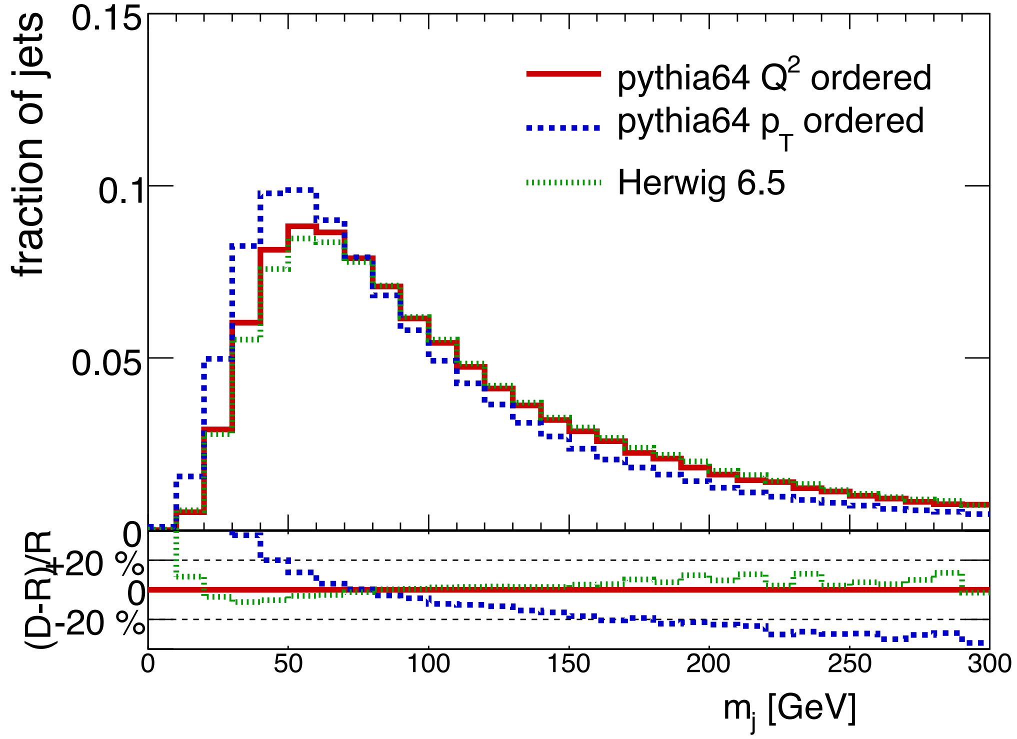
Monte Carlo sensitivity (dijets)



(a) jet mass - PS

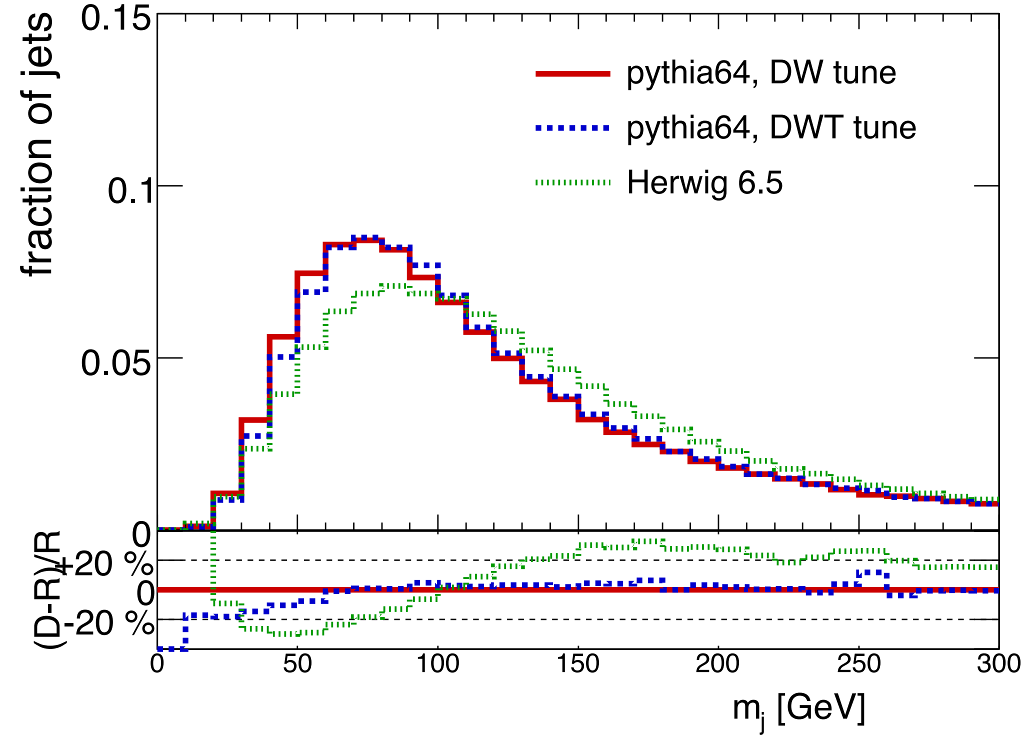
(b) jet mass - UE

no UE!



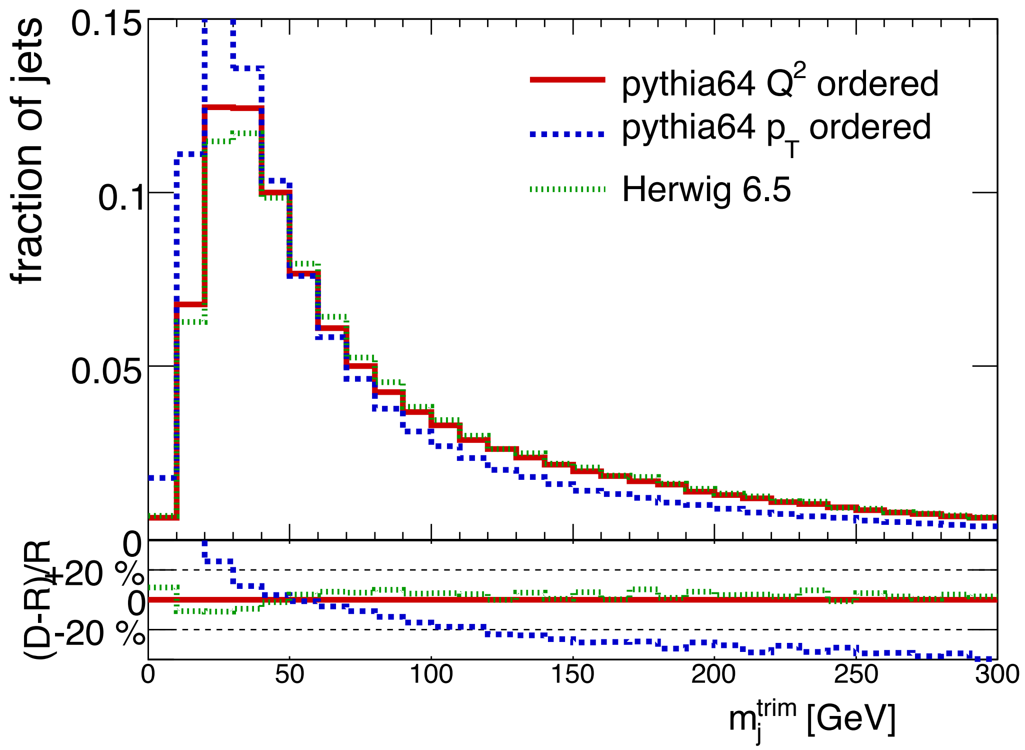
(a) jet mass - PS

GROOM

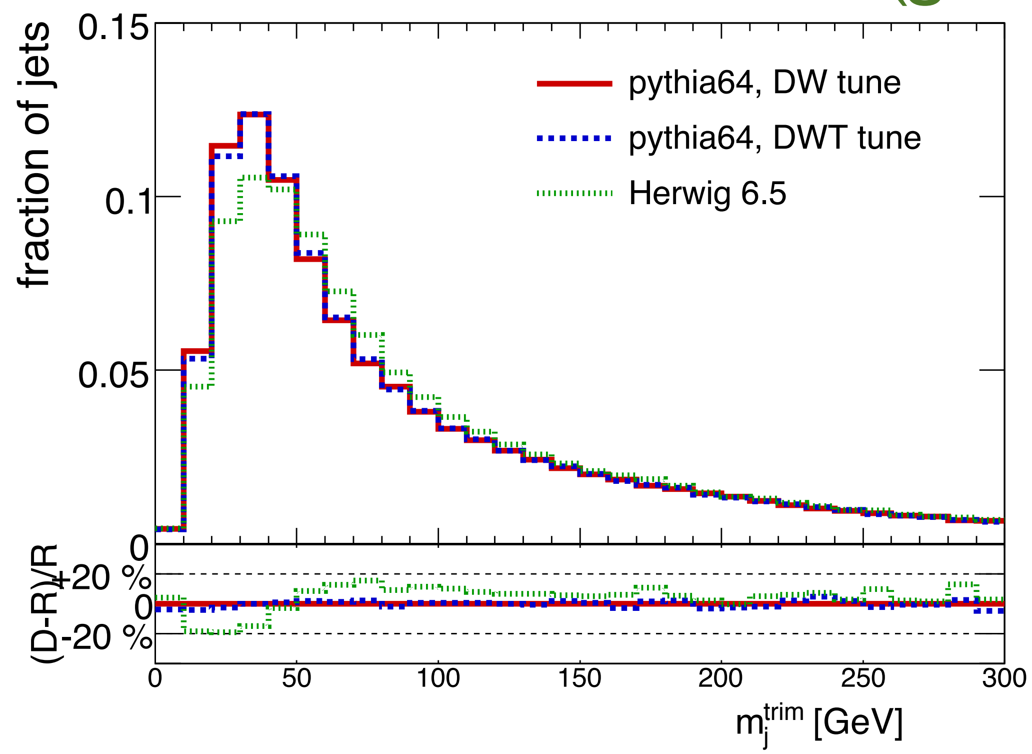


(b) jet mass - UE

(groom = trim)

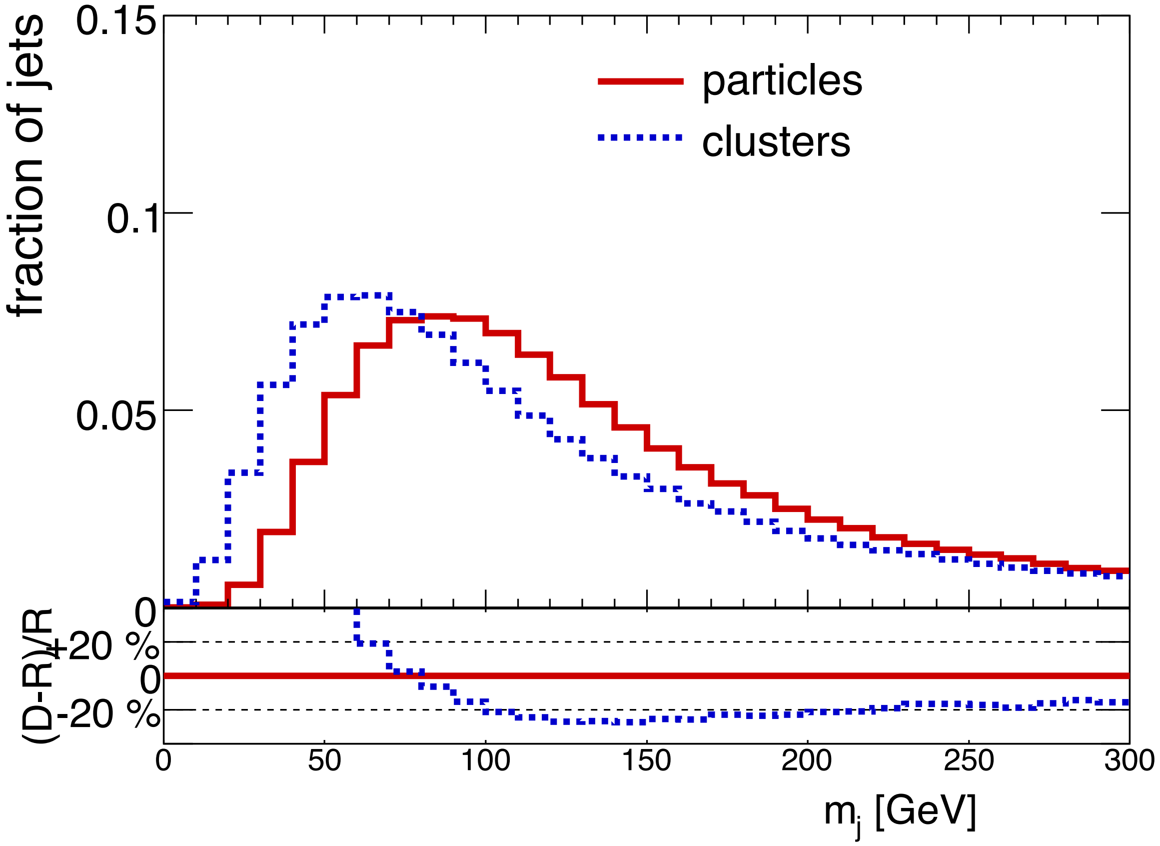


(d) groomed jets - PS

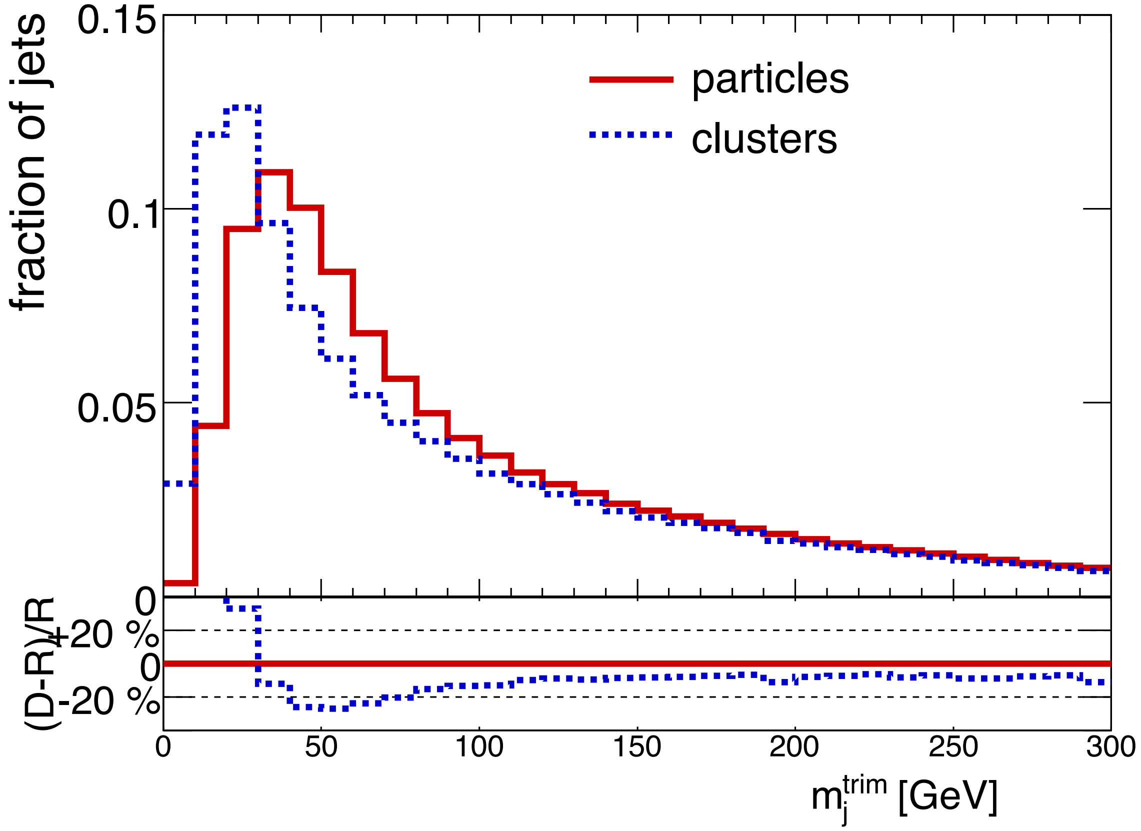


(e) groomed jets - UE

“Detector” effects



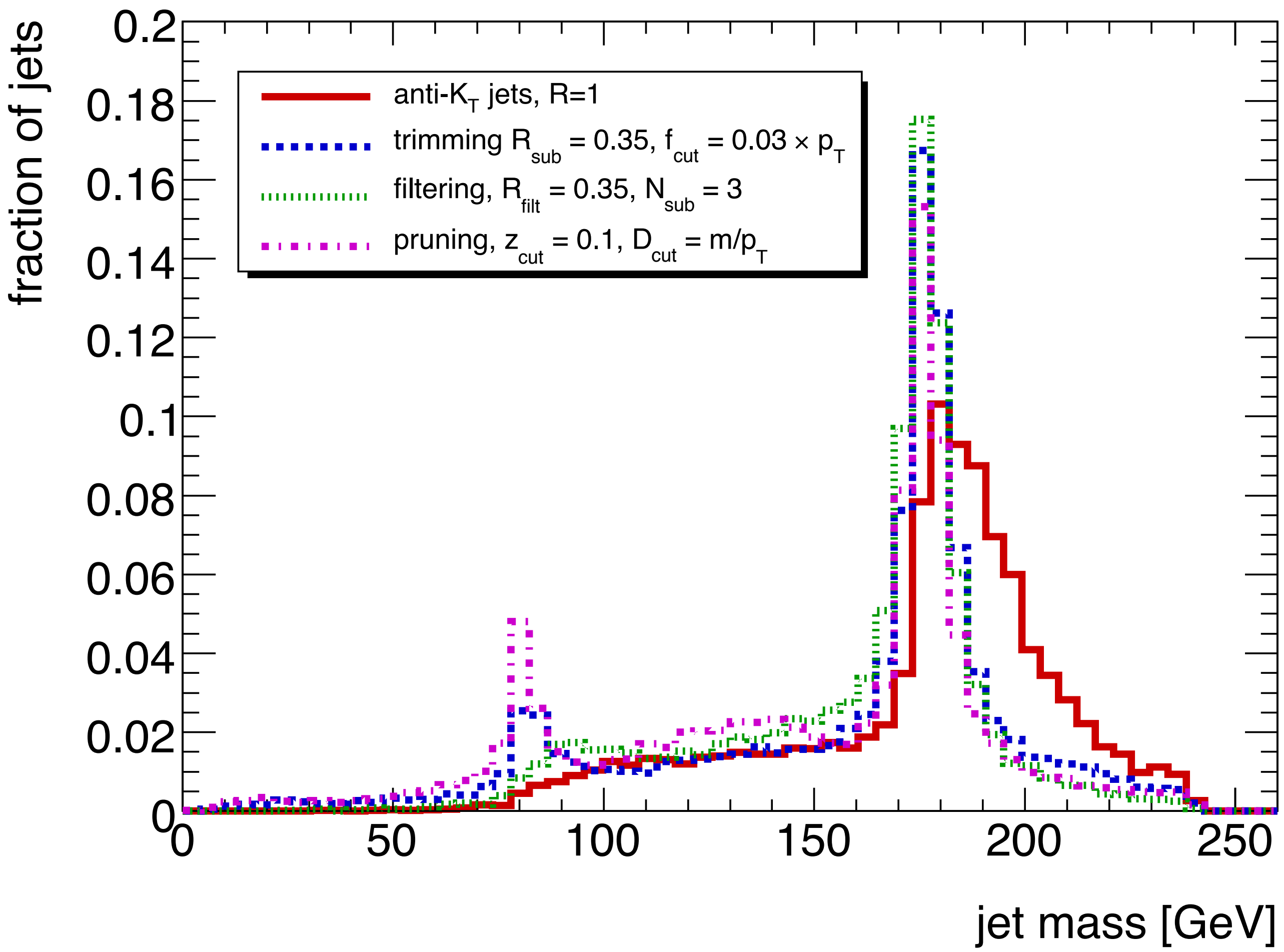
(c) jet mass - detector



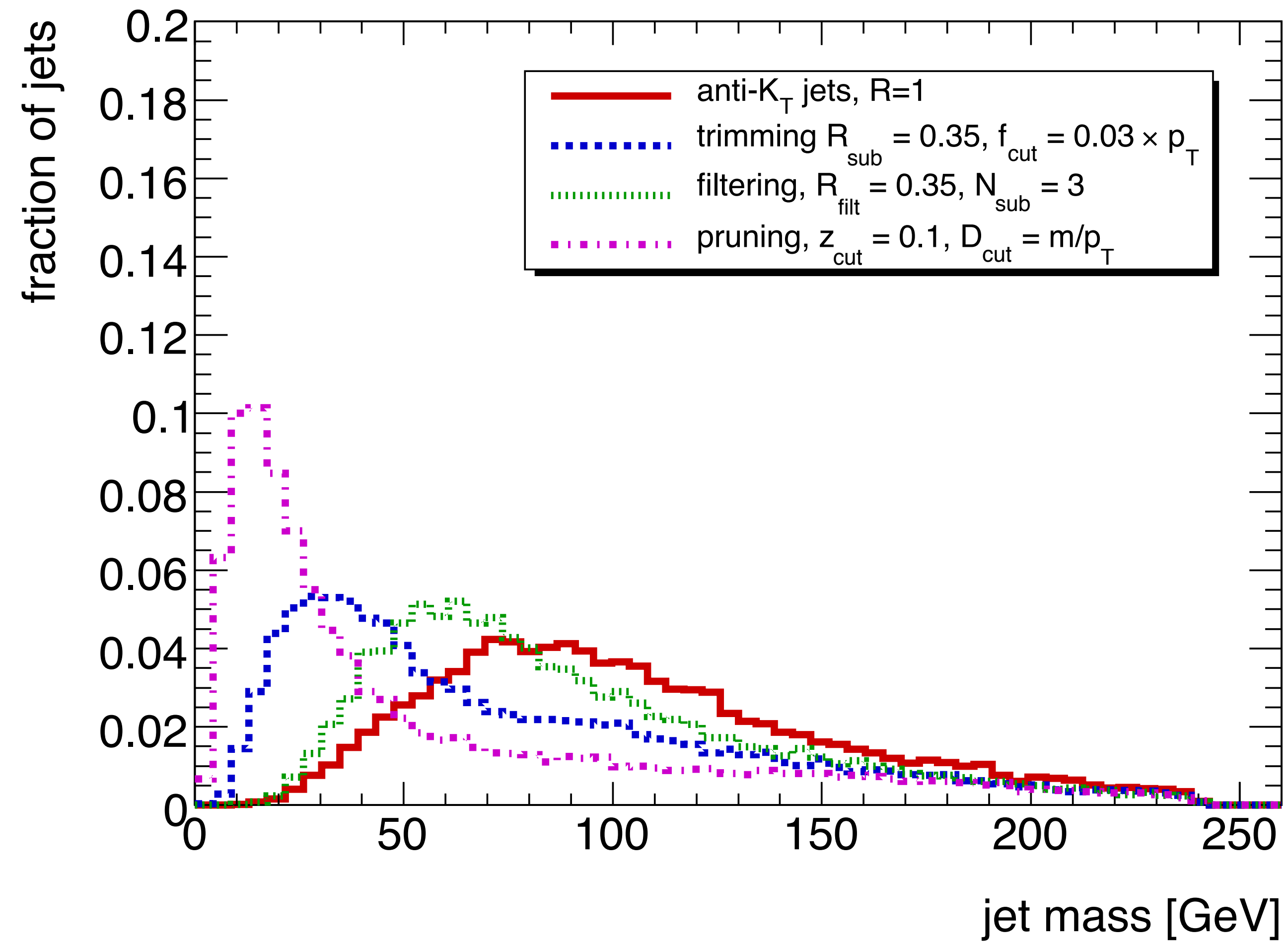
(f) groomed jets - detector

Groomed jets: easier to calibrate?

Comparing grooming methods



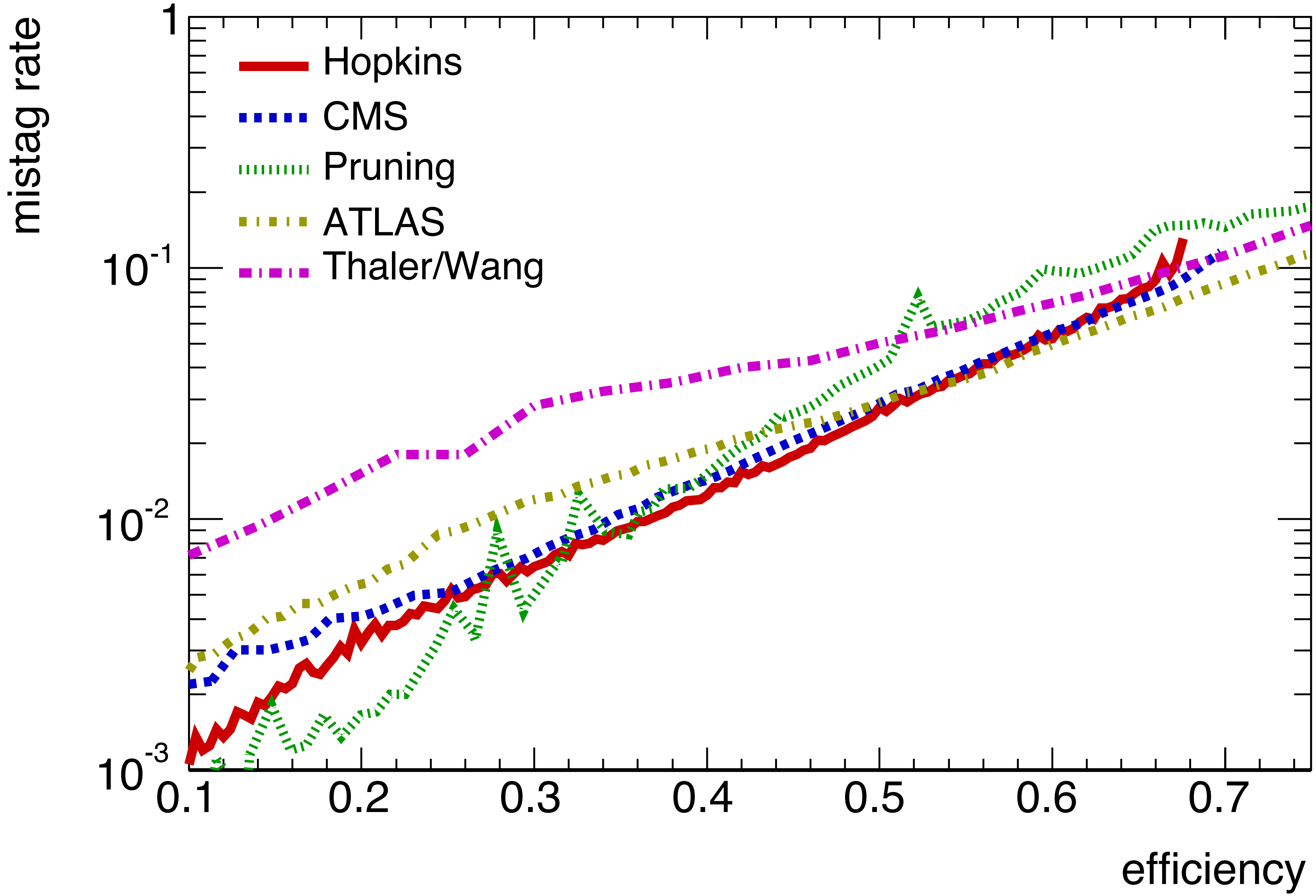
(c) $t\bar{t}$, 500–600 GeV



(d) dijets, 500–600 GeV

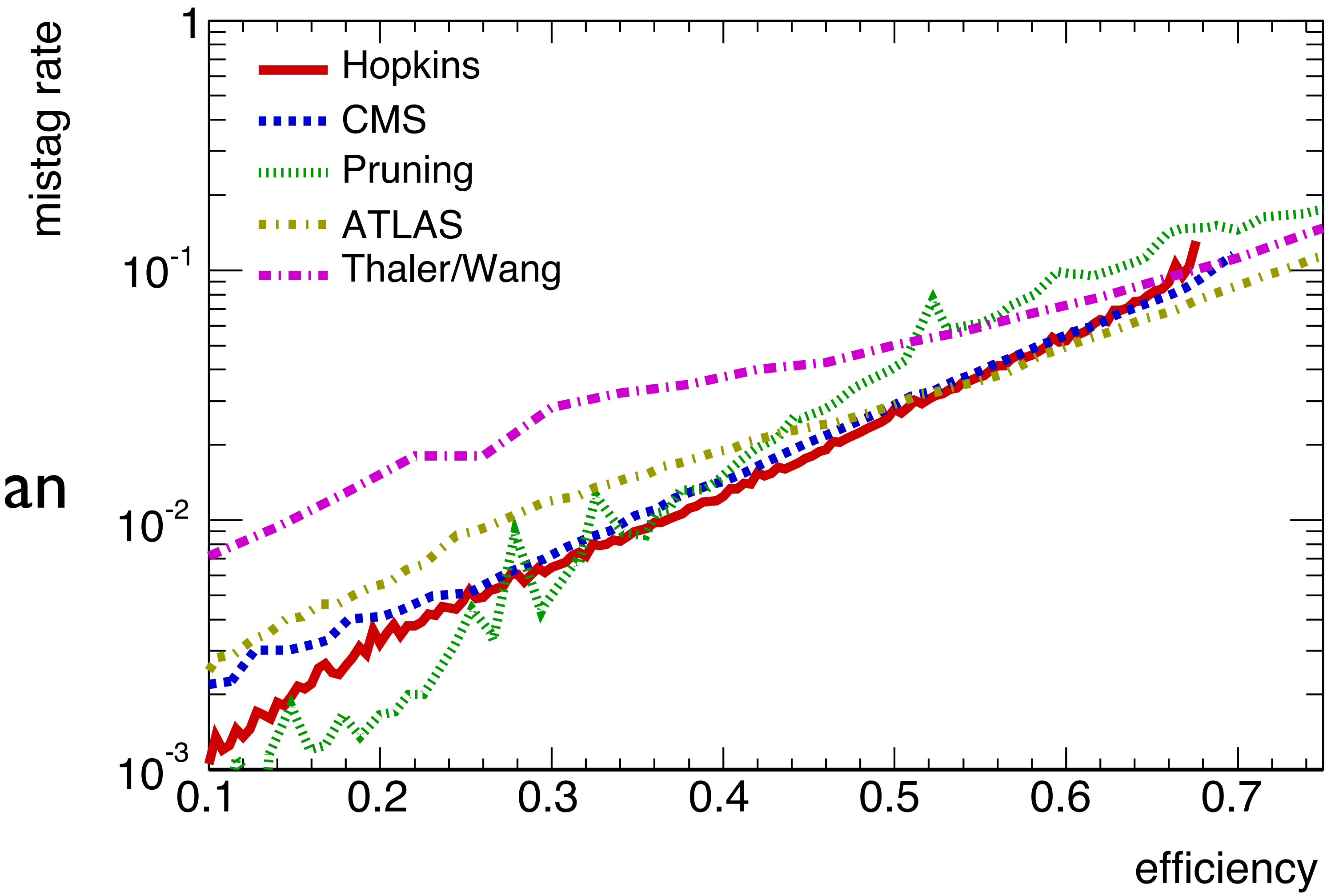
[Pruning: "standard" parameters; trimming and filtering: rough optimization for this sample]

Top-tagging comparison



Mistag vs. efficiency (tops over dijet): 500 GeV < pT < 600

Caveats!



* Not all taggers scanned all parameters (easier to scan cuts than parameters...)

* Not all ways to find tops represented (jet shapes?!?)

* BG is just dijets -- do we trust the Monte Carlo for this?

Still a lot to do to understand -- and believe -- the differences!

Outline

- A short history of jet substructure techniques
- Results from BOOST 2010
- What's next? (BOOST 2011 homework...)

What did we do? What's left?

- Compared *some* methods, for *one* signal/BG, with *one* MC
- *Started* to explore differences in MC models of substructure

- More complete set of techniques, especially jet shapes
- Other signals
- More consistent parameter scans

Aside: A tool to explore many jet tools would be useful! (See Brian's talk...)

(Some) outstanding issues in jet substructure

- What do these techniques look like in real data?
  - How do we calibrate all of these jet substructure variables?
  - What signals are good testbeds?
- Will theorists make better use of the detector?
  - (Detector != 0.1 x 0.1 HCAL...)
  - Can we enhance substructure techniques with ECAL resolution, tracking info?
- (How much) can we trust the Monte Carlos?
  - Study jet masses and jet substructure in early data and compare to MC
  - Does matching make things better? *Does matching make things worse?!*
  - Need a better understanding of how MCs differ on substructure

(Giant MC review: 1101.2599)

PS: More than you ever wanted to know about substructure: 1101.1335

arXiv:1101.1335v1 [hep-ph] 7 Jan 2011

Jet Substructure at the Large Hadron Collider:
Harder, Better, Faster, Stronger

Christopher K. Vermilion

A dissertation submitted in partial fulfillment
of the requirements for the degree of

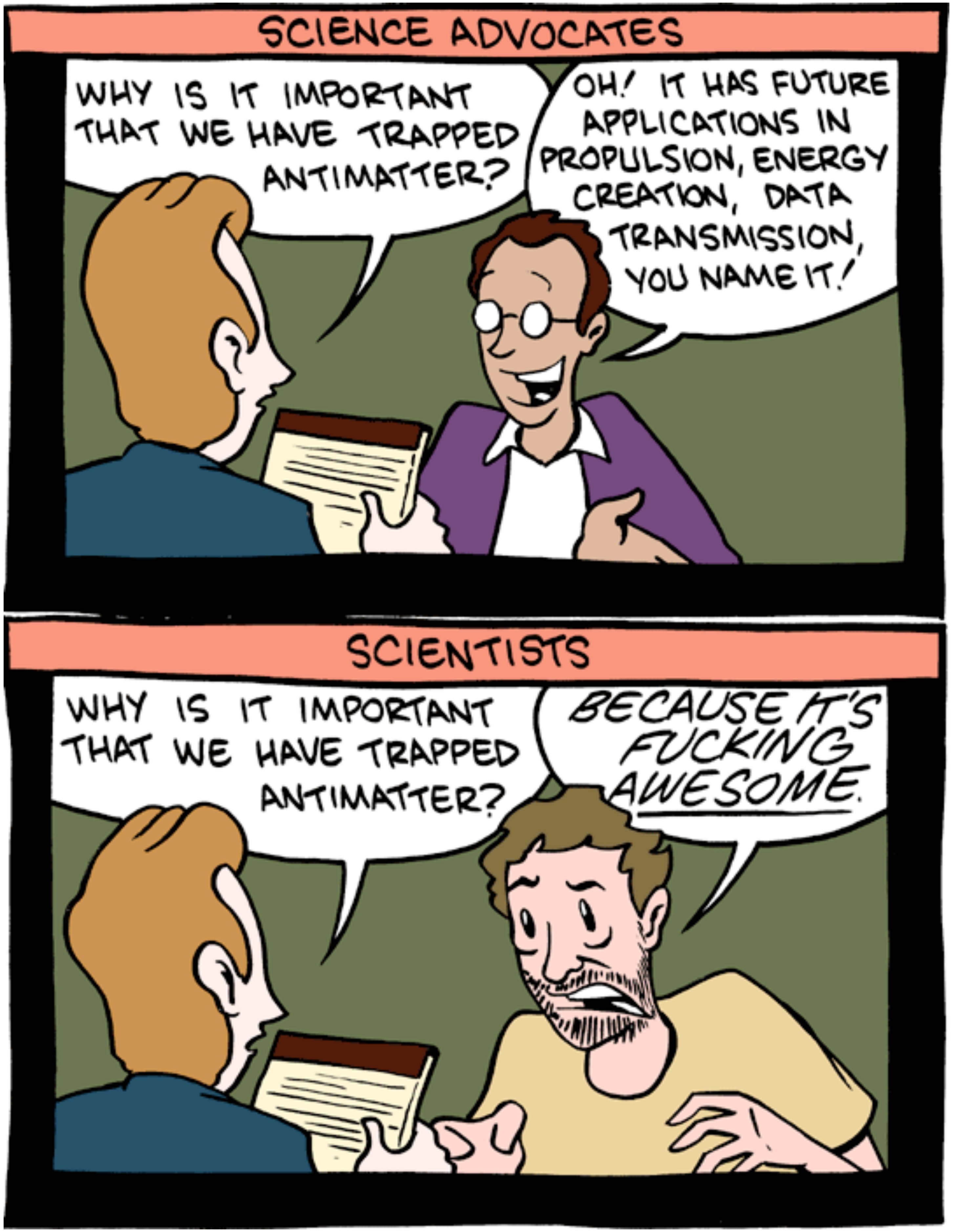
Doctor of Philosophy

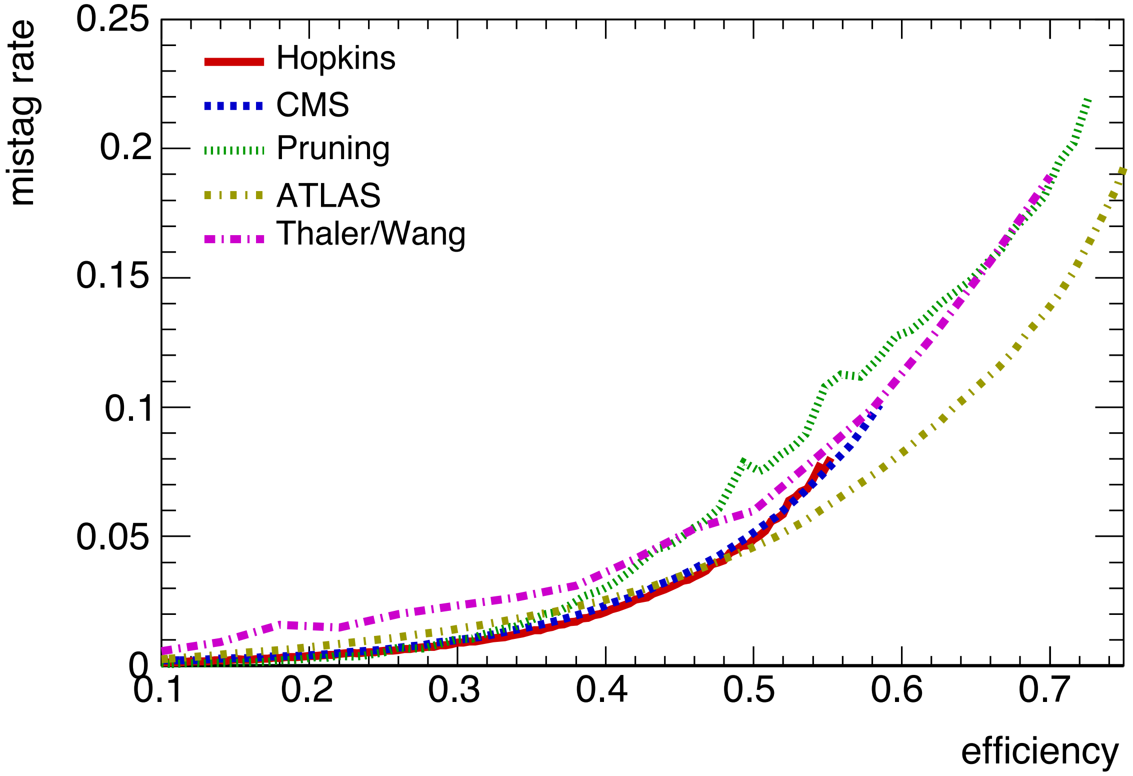
University of Washington

2010

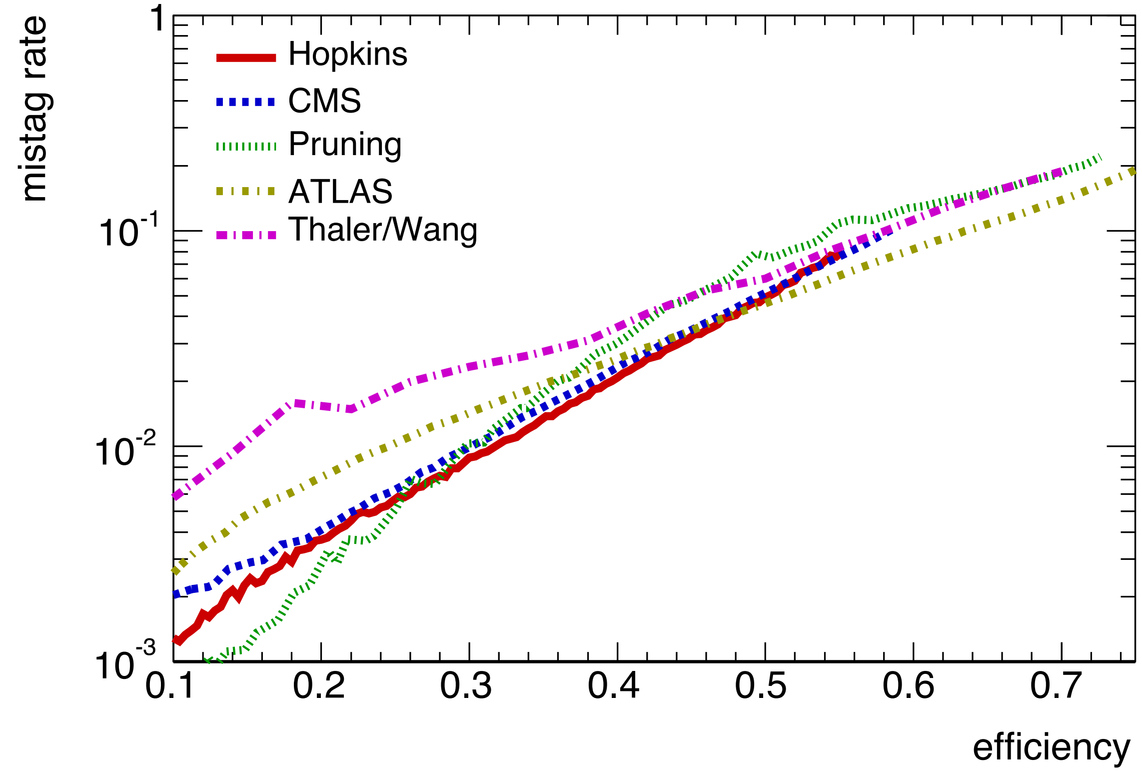
Program Authorized to Offer Degree: Department of Physics

Bonus slides

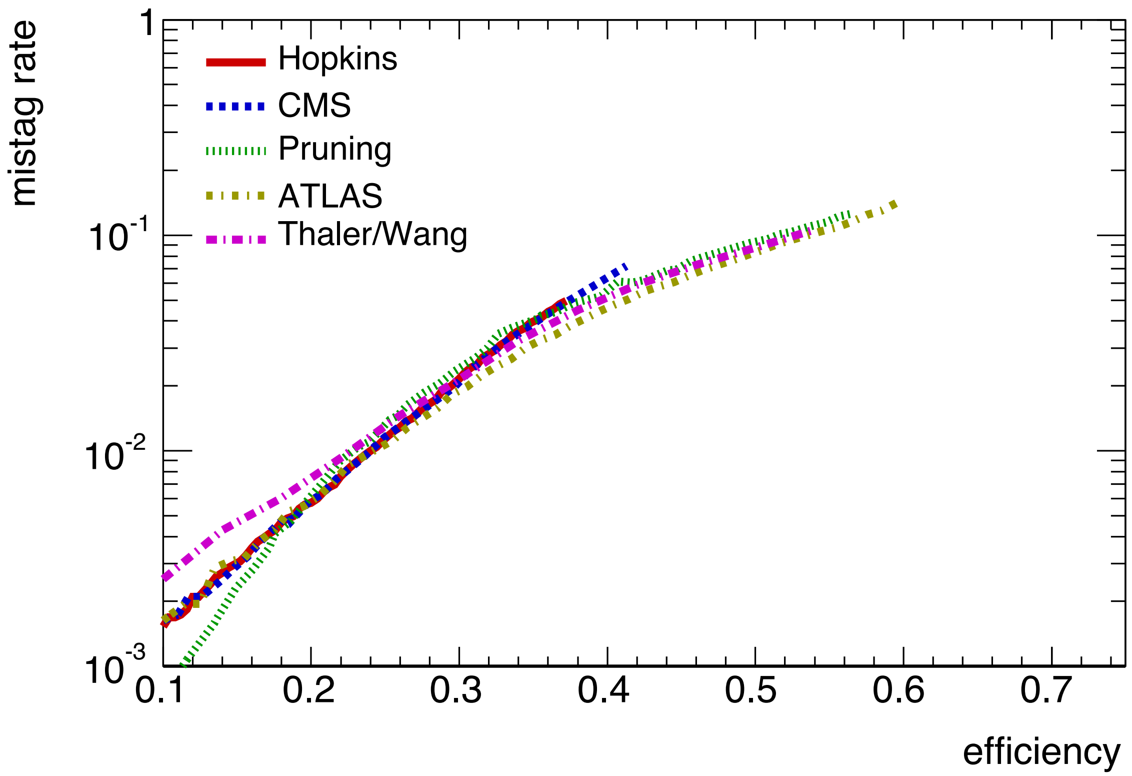




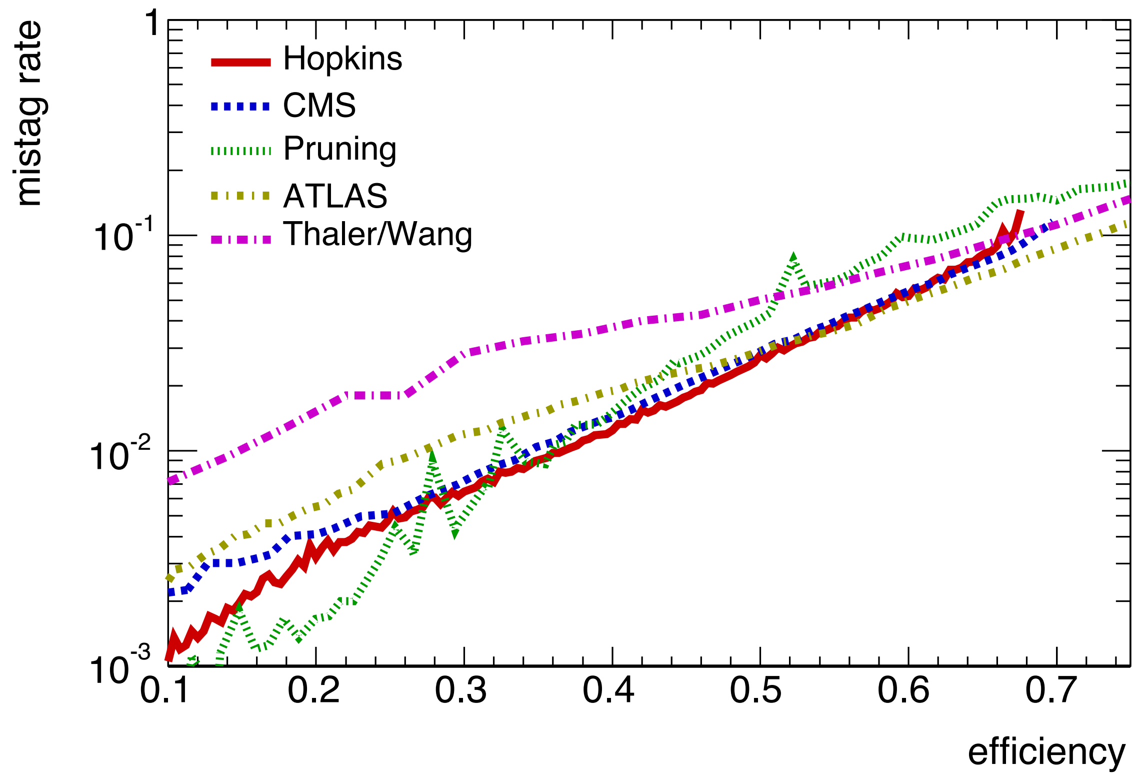
(a) all p_T samples



(b) all p_T samples

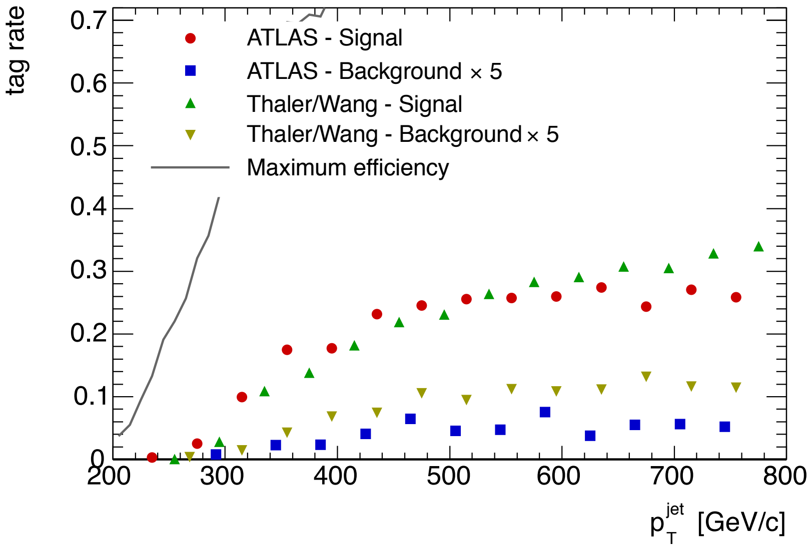


(c) 300–400 GeV

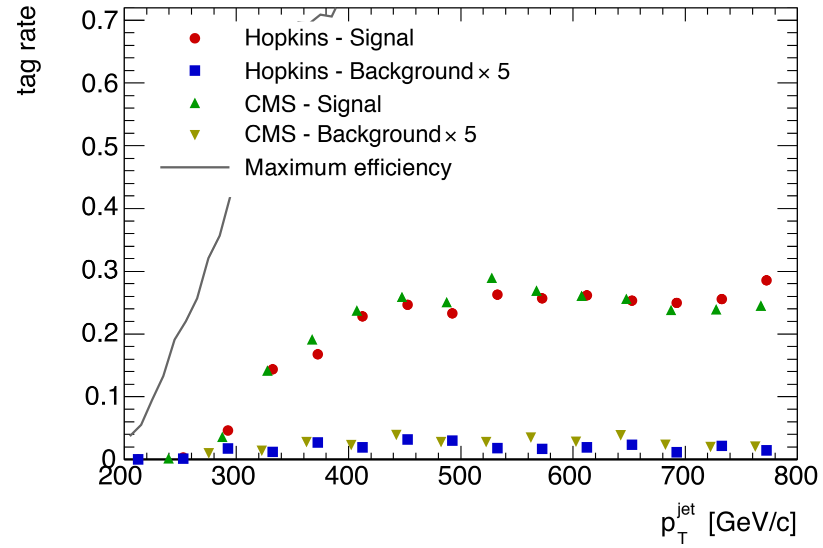


(d) 500–600 GeV

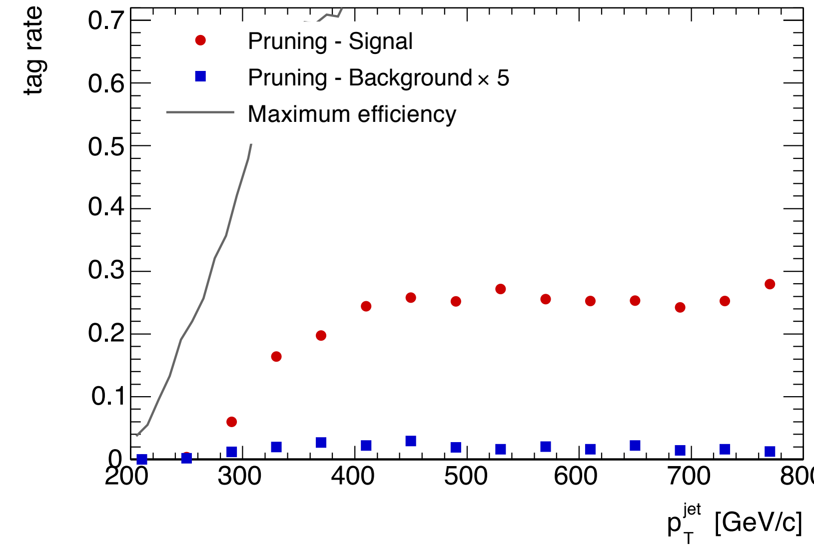
Friday, January 14, 2011



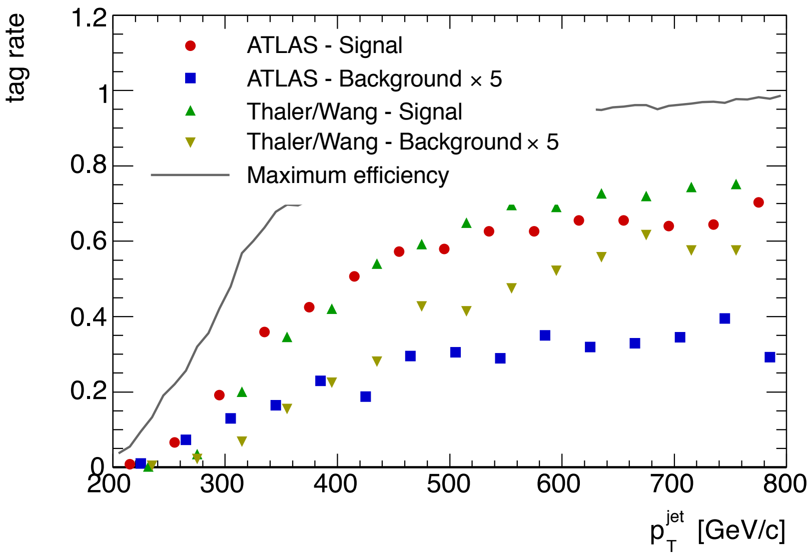
(a) ATLAS, Thaler/Wang, $\epsilon = 20$ %



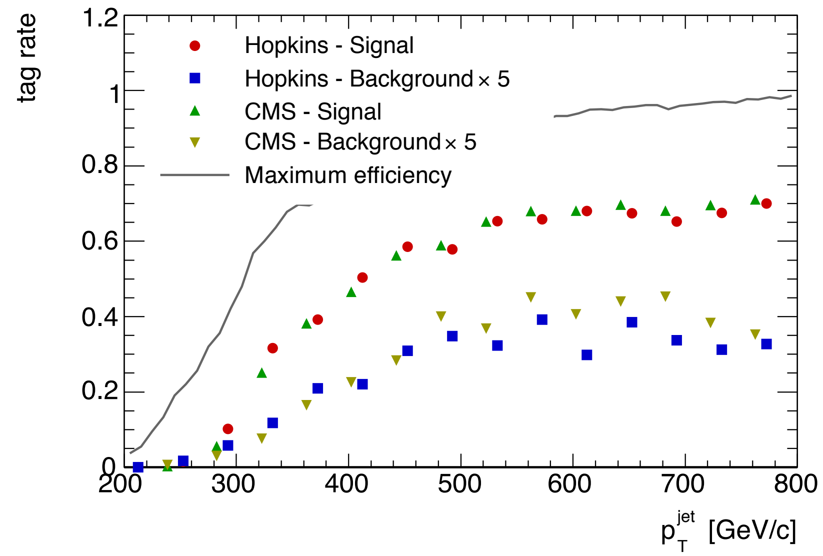
(b) CMS, Hopkins, $\epsilon = 20$ %



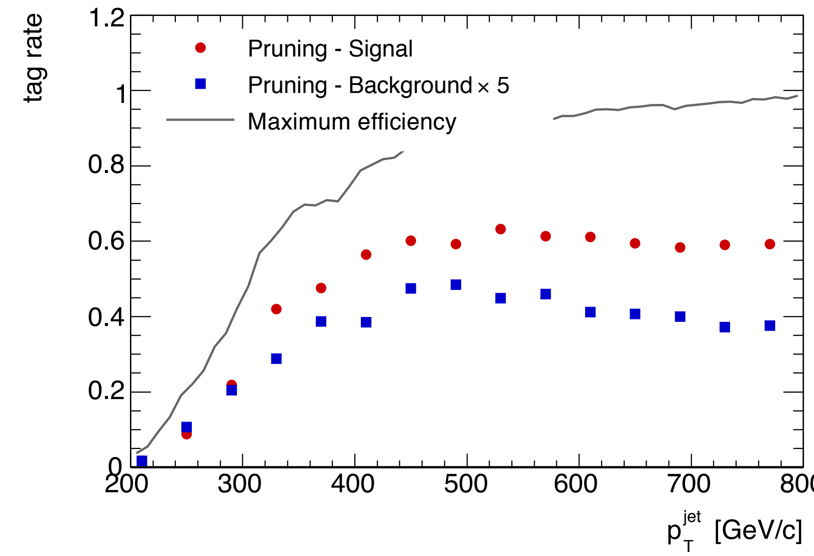
(c) pruning, $\epsilon = 20$ %



(d) ATLAS, Thaler/Wang, $\epsilon = 50$ %

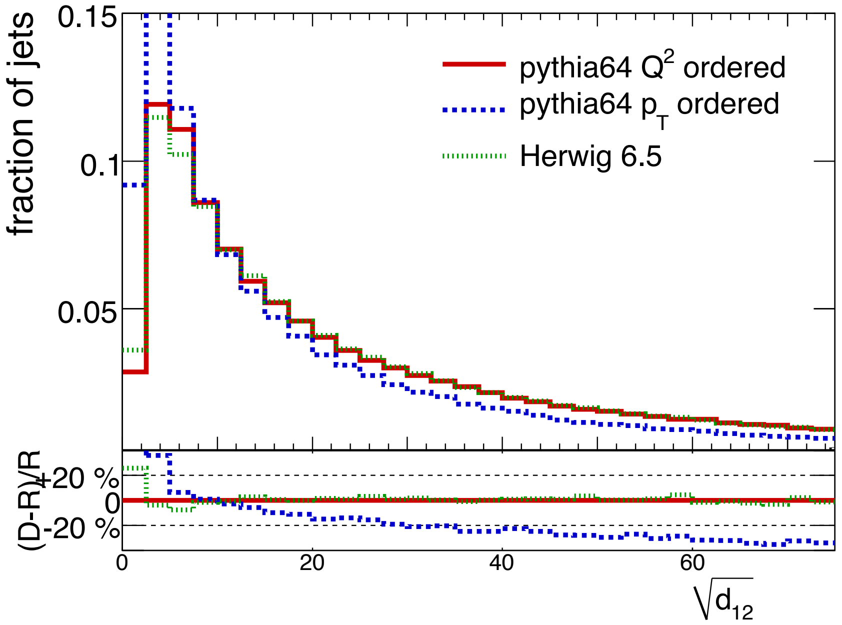


(e) CMS, Hopkins, $\epsilon = 50$ %

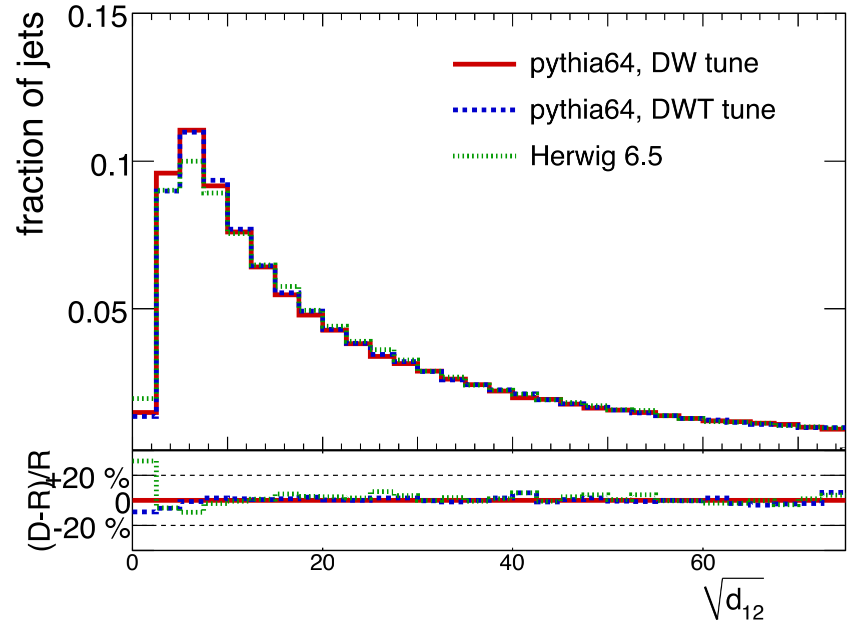


(f) pruning, $\epsilon = 50$ %

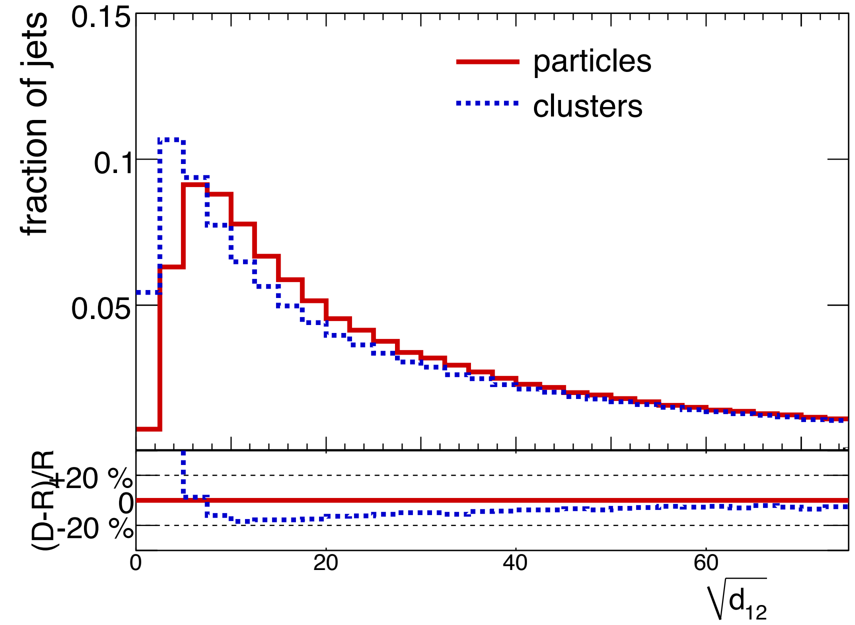
Splitting scales in different MCs



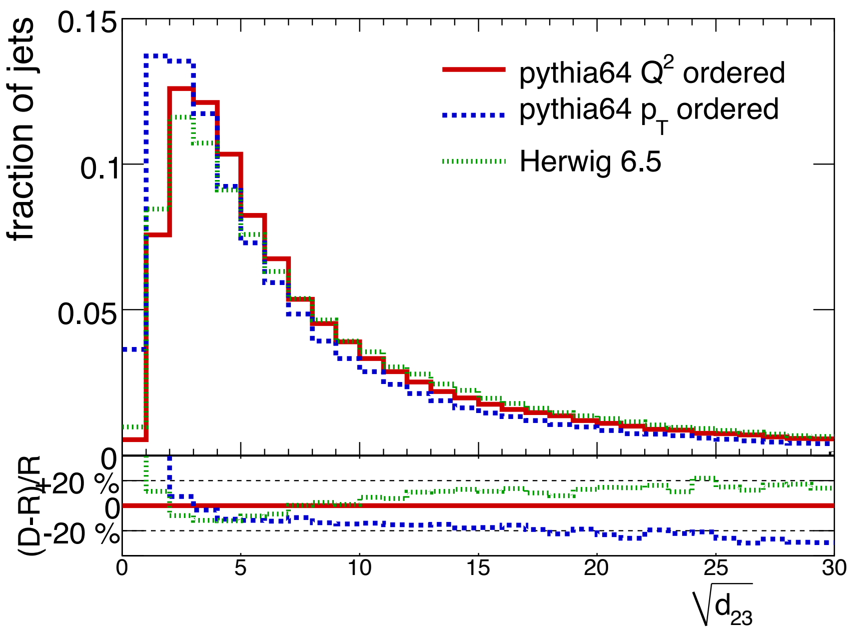
(g) $1 \to 2$ split - PS



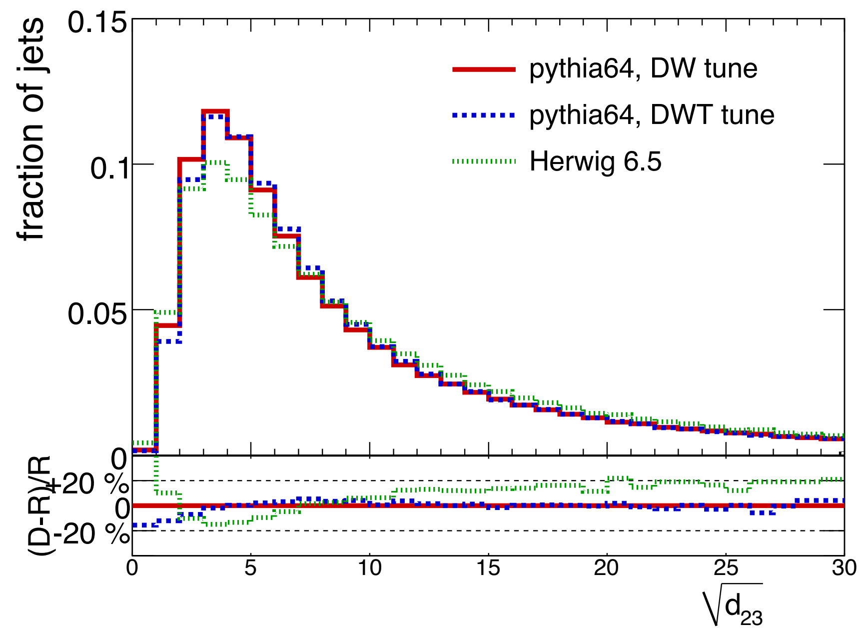
(h) $1 \to 2$ split - UE



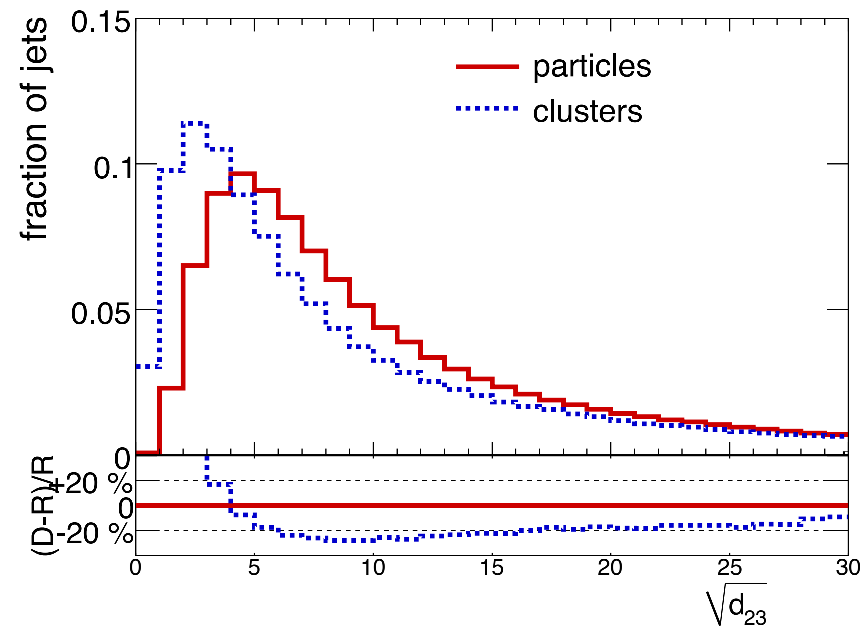
(i) $1 \to 2$ split - detector



(j) $2 \to 3$ split - PS



(k) $2 \to 3$ split - UE



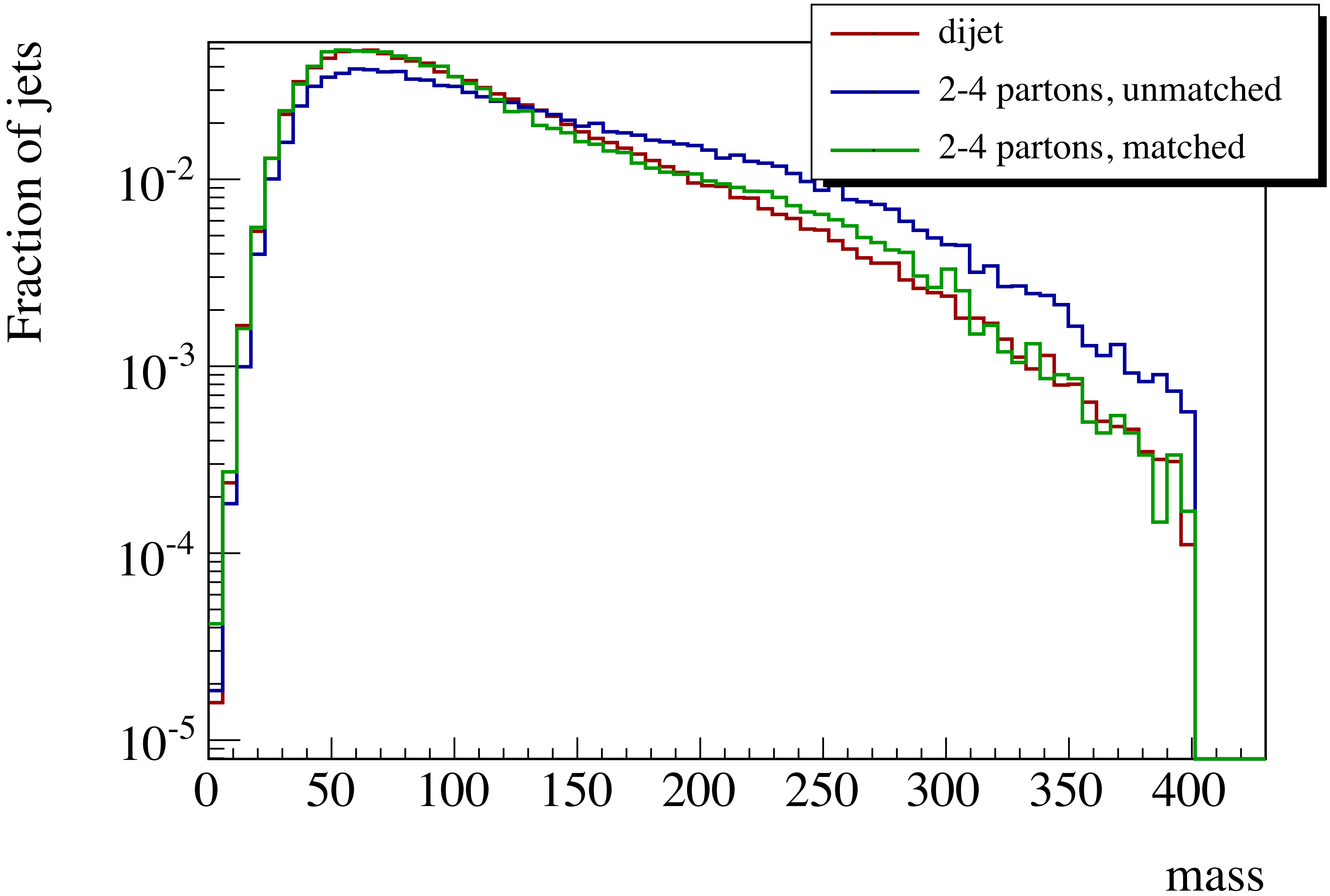
(l) $2 \to 3$ split - detector

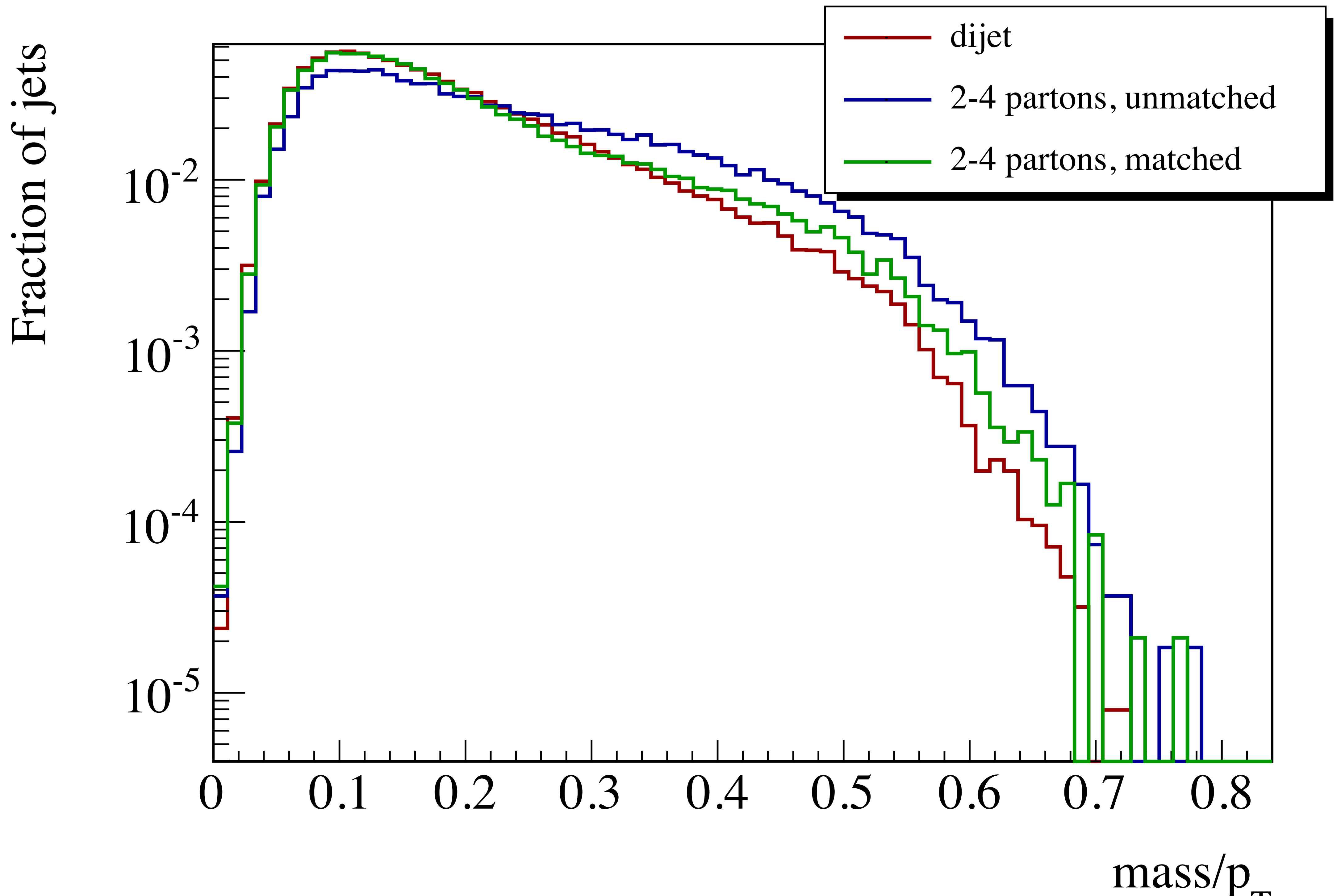
BOOST working points

Tagger	Parameters at 20% working point	Parameters at 50% working point
Hopkins	$\delta_p = 0.1,\ \delta_r = 0.19$ $170 < m_{\rm top} < 195$ GeV, $\cos\theta_h < 0.675$, $75 < m_W < 95$ GeV	$\delta_p = 0.04,\ \delta_r = 0.19$ $160 < m_{\rm top} < 265$ GeV, $\cos\theta_h < 0.95$, $60 < m_W < 120$ GeV
CMS	$170 < m_{\rm jet} < 200$ GeV $m_{\rm min} > 75$ GeV	$164 < m_{\rm jet} < 299$ GeV $m_{\rm min} > 42.5$ GeV
Pruning[a]	$z_{\rm cut} = 0.1,\ D_{\rm cut}/(2m/p_T) = 0.2$ $68 < m_W < 88$ GeV $150 < m_{\rm top} < 190$ GeV	$z_{\rm cut} = 0.05,\ D_{\rm cut}/(2m/p_T) = 0.1$ $28 < m_W < 128$ GeV $120 < m_{\rm top} < 228$ GeV
ATLAS[b]	N/A	N/A
Thaler/Wang	$m_W > 68$ GeV $0.249 < z_{\rm cell} < 0.664$ $183 < m_{\rm jet} < 234$ GeV	$m_W > 59$ GeV $0.0498 < z_{\rm cell} < 0.509$ $162 < m_{\rm jet} < 265$ GeV

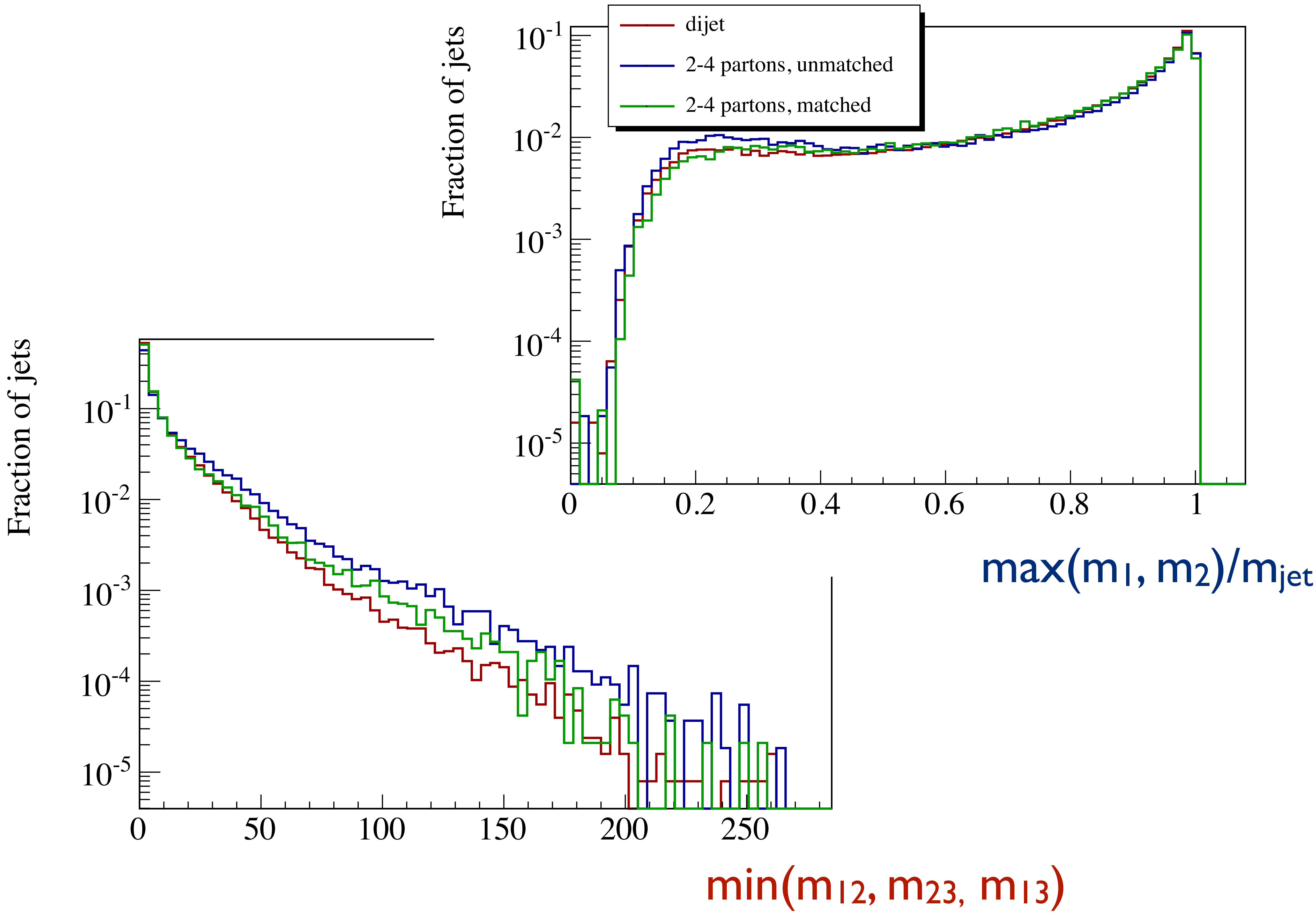
2-4 parton matched vs. dijet

MadGraph+Pythia6





Fraction of jets
10^{-2}
10^{-3}
10^{-4}
10^{-5}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
mass/p_T
dijet
2-4 partons, unmatched
2-4 partons, matched


dijet
2-4 partons, unmatched
2-4 partons, matched
Fraction of jets
Fraction of jets
max(m1, m2)/mjet
min(m12, m23, m13)

Background subtraction may be a challenge!