Recent Developments in Jets and Substructure

Jonathan Walsh, UC Berkeley

Motivations

Jet substructure brings to the table:

- Methods to improve resolution for jetty (and soft QCD) signals
  (q/g tagging, pileup reduction, jet grooming, track-based tools)

- Search techniques in unique kinematic regimes

- Strong test of our understanding of perturbative/nonperturbative QCD

As a theorist, substructure is on the verge of coalescing many tools:

- phenomenology for observable design, BSM models

- MC techniques

- fixed order and resummed calculations, general QCD phenomenology

- experimentally relevant observables

Motivations

Excellent success in Run 1

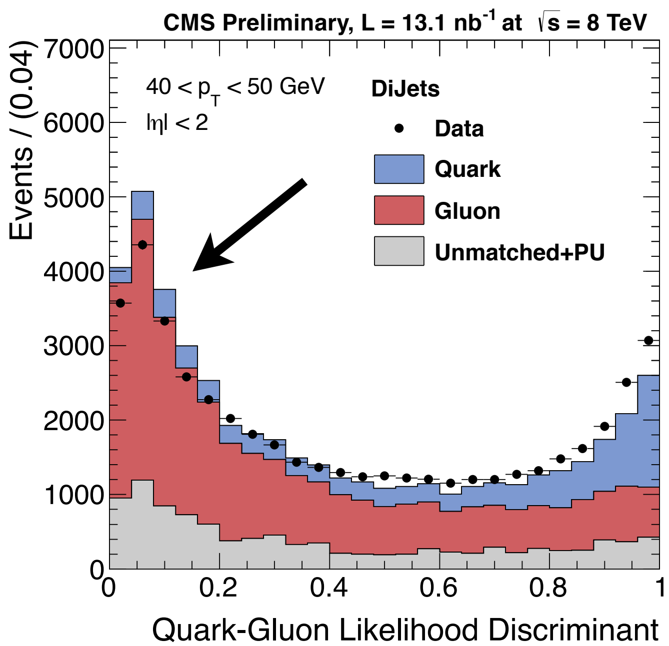
- Substructure started at basically zero in 2008
  - Great set of results: top, W tagging, BSM searches, q/g tagging, pileup suppression, novel observables
  - Amazing turnaround from proposed techniques to study and validation in experiments
- Strong synergy of theoretical and experimental methods

Lots of promise and challenges for Run 2:

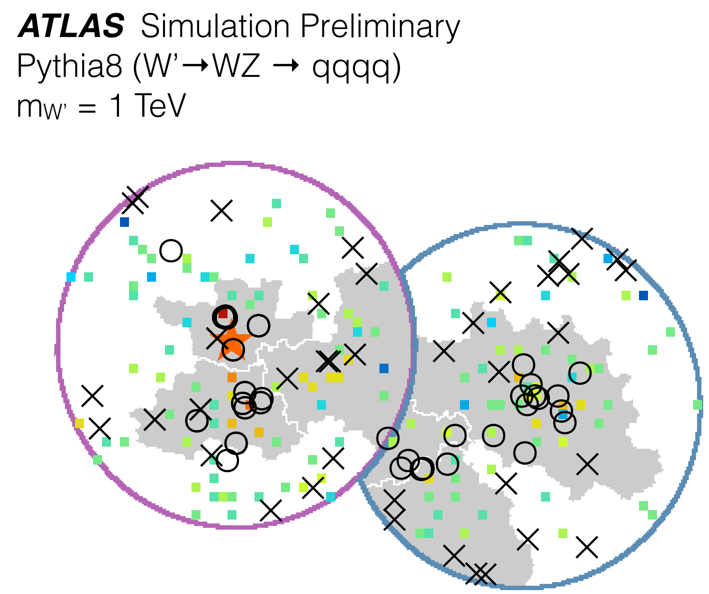
- Boosted regime opens up significantly, higher rates
  - Tests the detector resolution (example: 1 TeV W/Z has an average opening angle of ~0.2)
  - Boosted Higgs studies, models with boosted topologies
- With great luminosity comes great responsibility

Outline

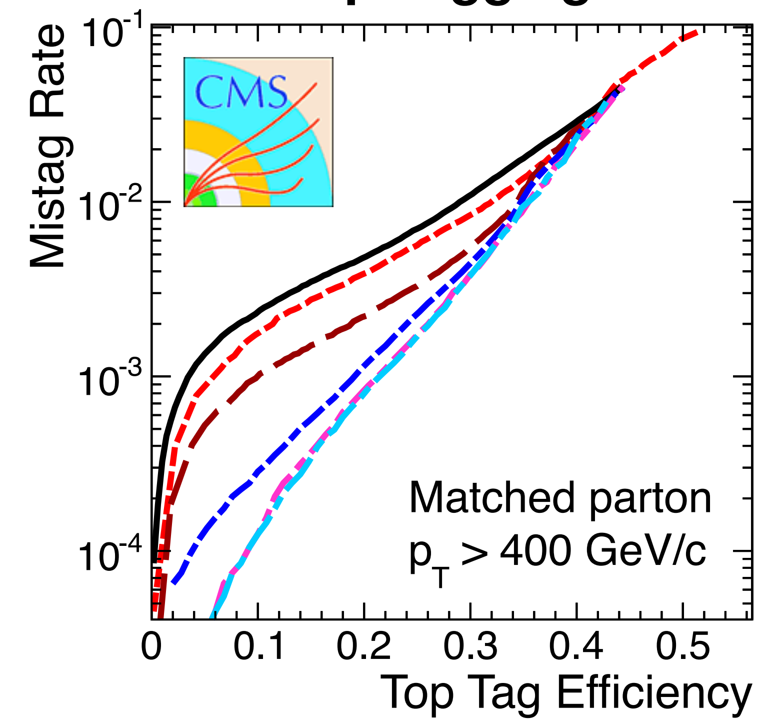
Quark/Gluon Tagging



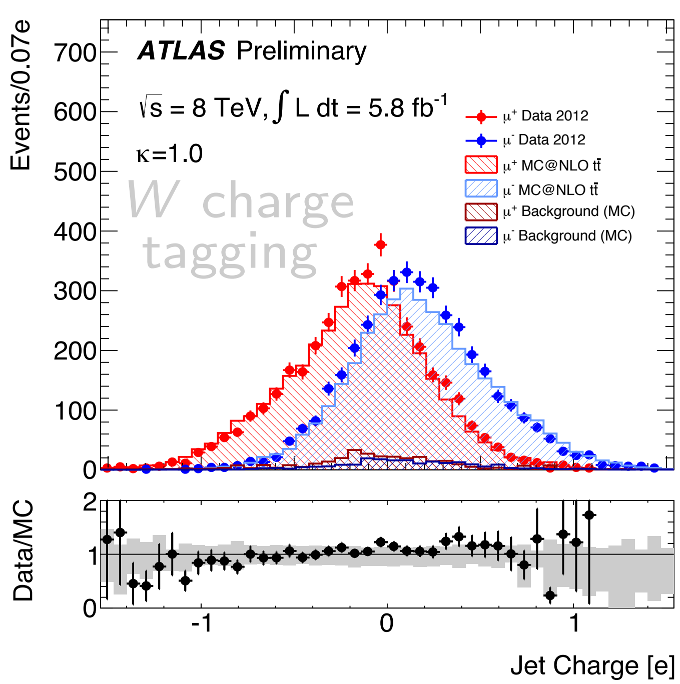
Jet Grooming



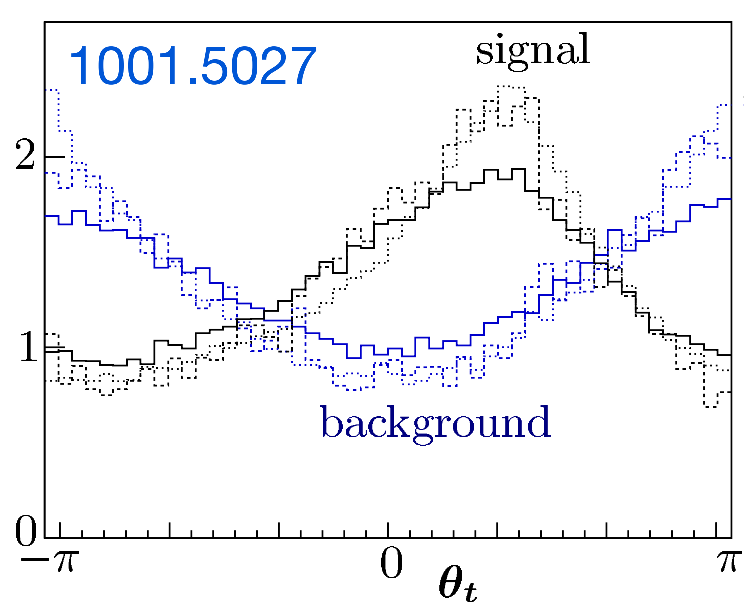
Top Tagging



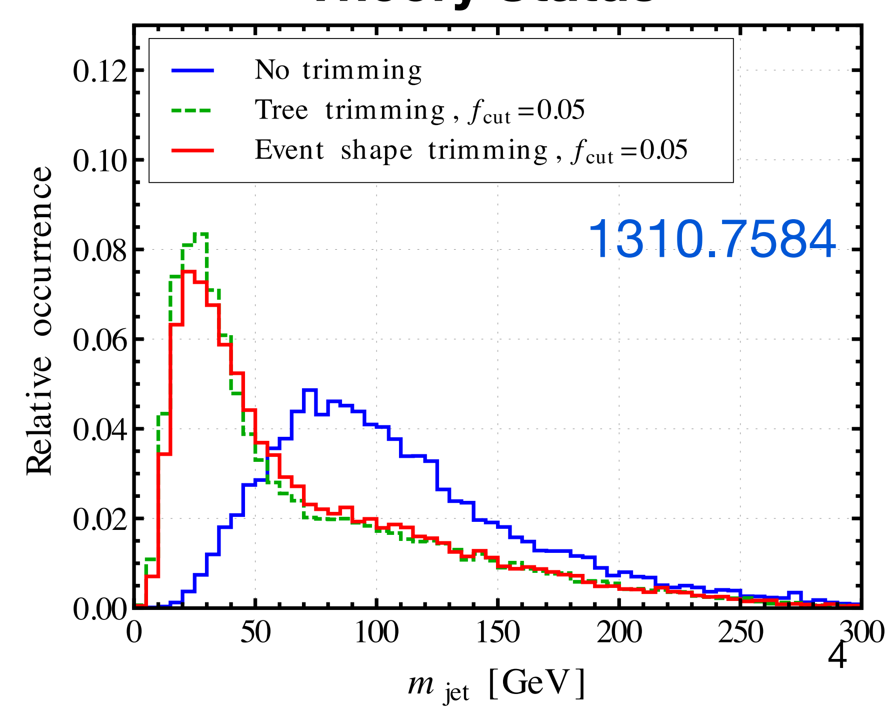
Tracks



Kinematics vs. Color Flow

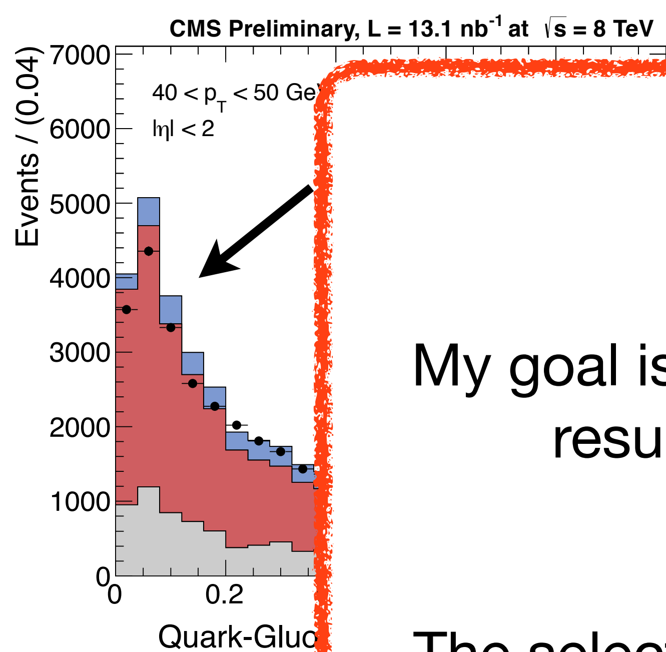


Theory Status



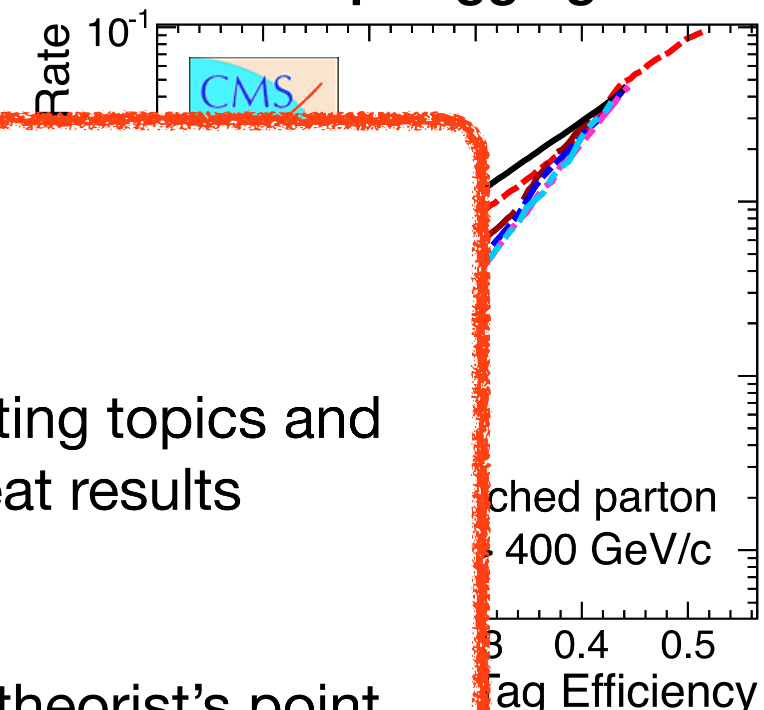
Outline

Quark/Gluon Tagging



Jet Grooming

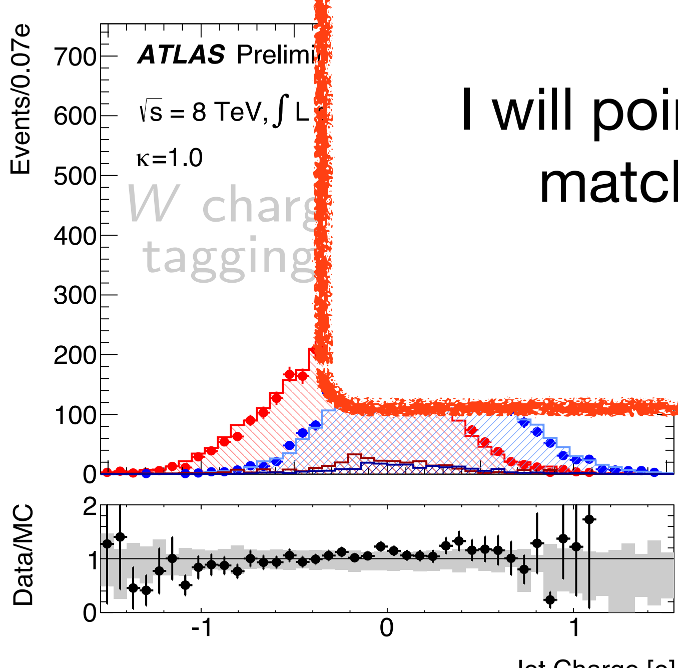
Top Tagging

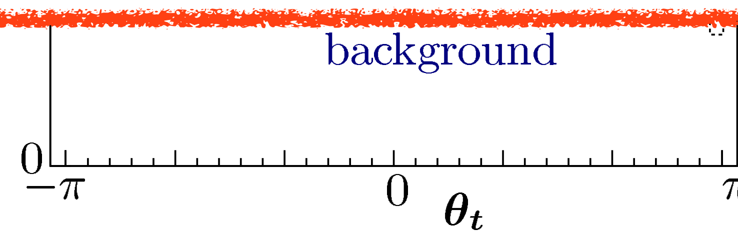


N.B.:
My goal is to give an overview of several interesting topics and results for jet substructure, highlighting great results but also challenges

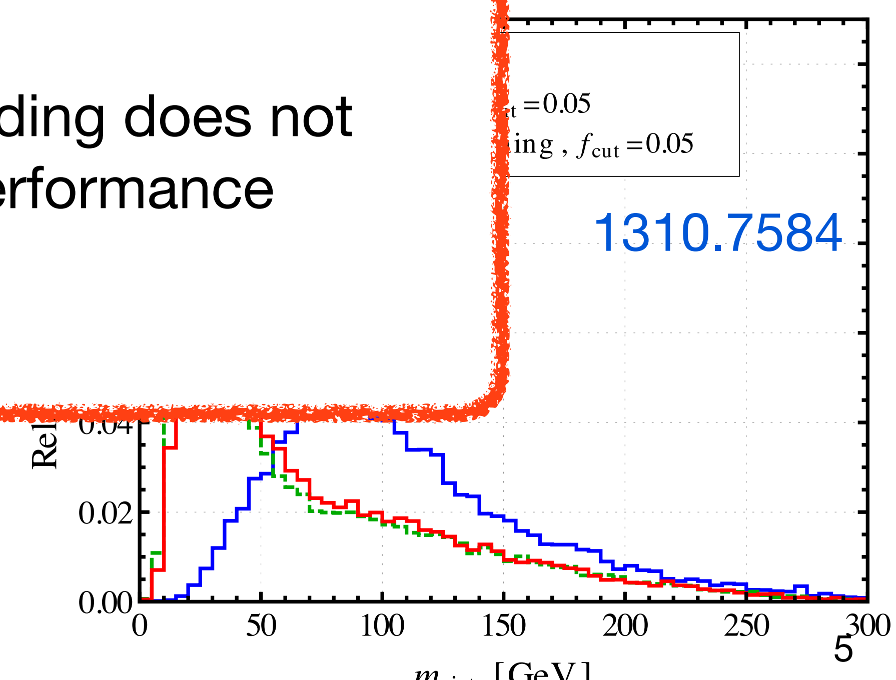
The selection of topics is subjective and from a theorist's point of view, but I hope to give an accurate snapshot of the field

I will point out problems where our understanding does not match data, rather than areas with poor performance



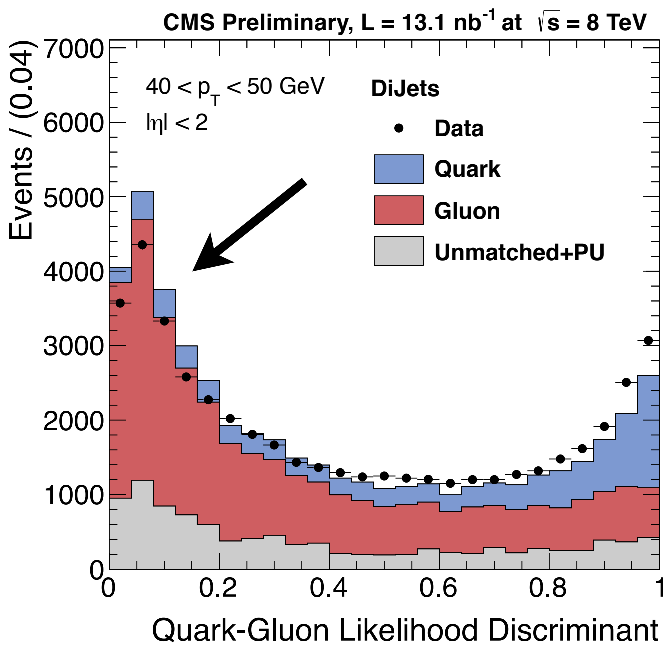


Status

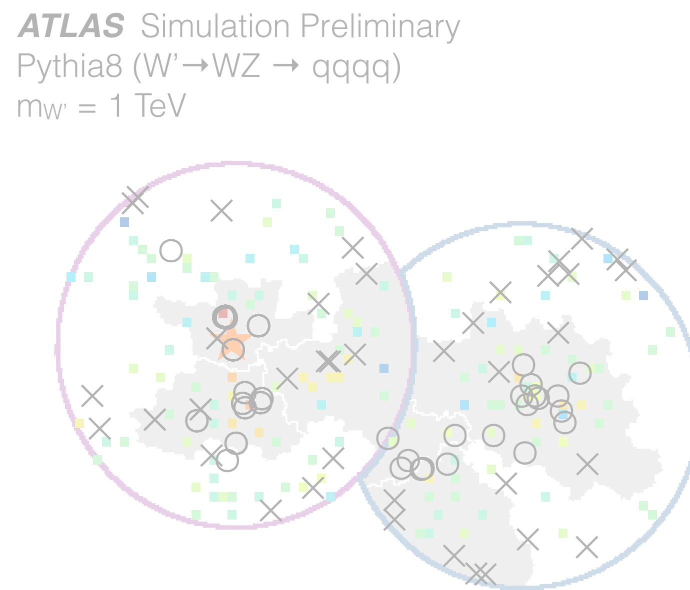


Outline

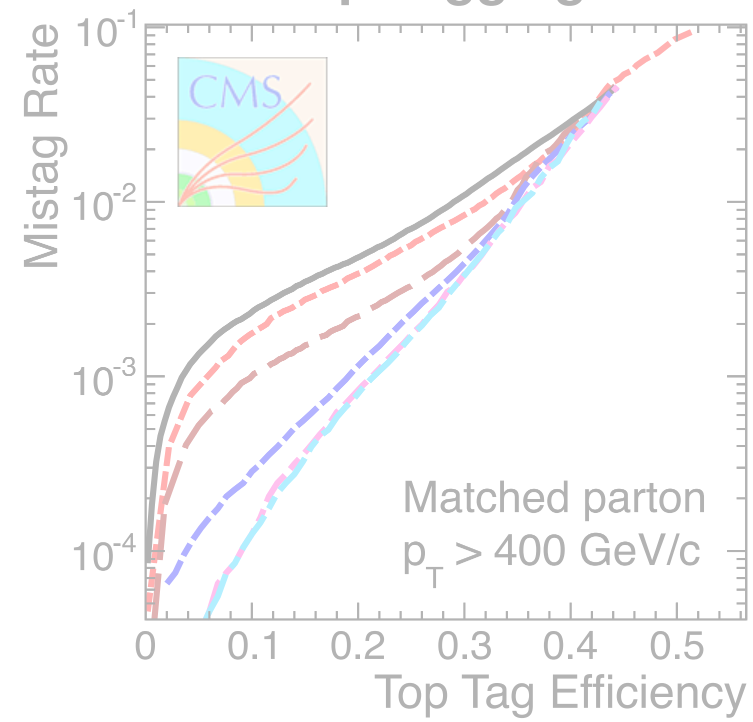
Quark/Gluon Tagging



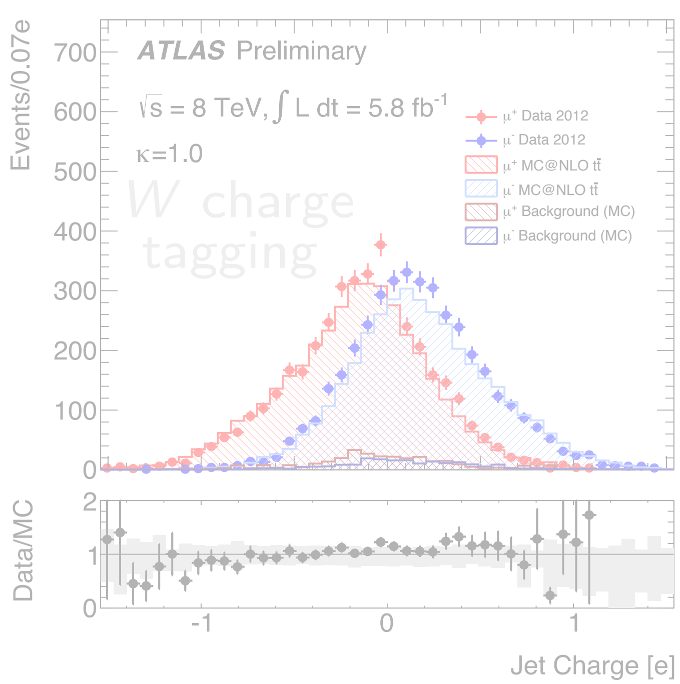
Jet Grooming



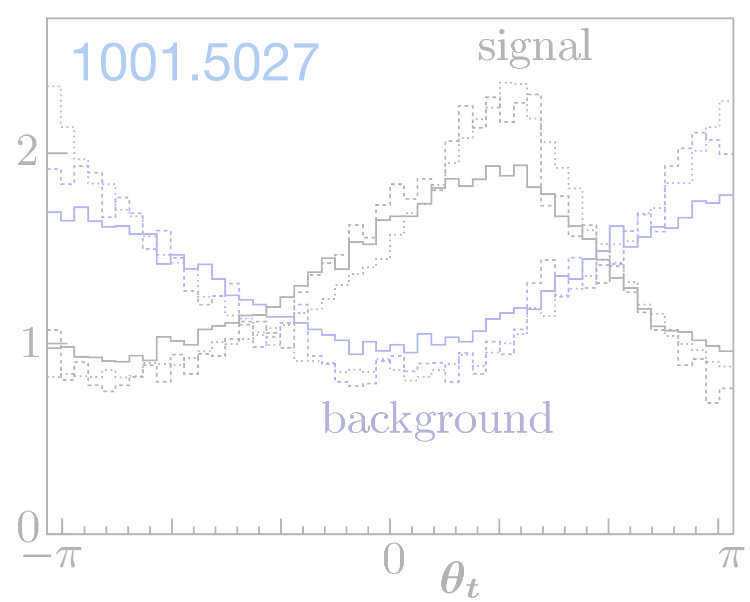
Top Tagging



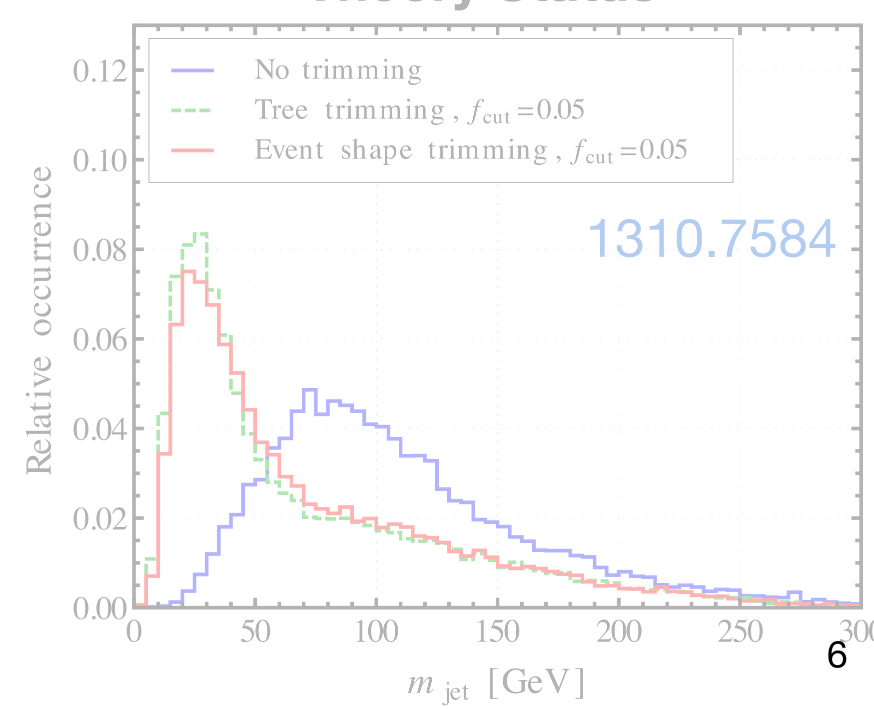
Tracks



Kinematics vs. Color Flow



Theory Status



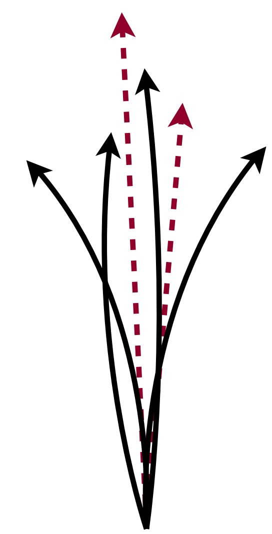
Quarks vs. Gluons: Use

quark/gluon tag is useful discriminant for BSM searches

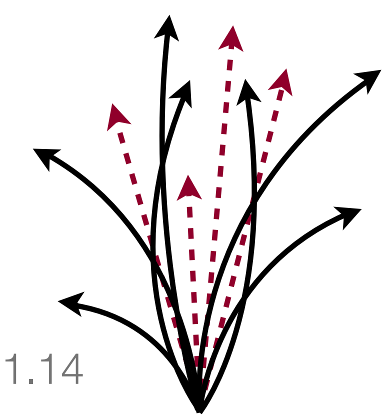
- SUSY is quark enriched over SM
- Higgs signals have strong q/g dominance
    - gg → H → WW vs. WW (quark initiated)
    - VH (q initiated) vs. Z+jets gluon enriched
    - VBF (q jets) vs. ttbar (gluon initiated)
- Important to discriminate by color topology in dijet searches

Robust test of MC/data comparison, interesting QCD problem across many scales

Quark jets:



Gluon jets:



Francesco Pandolfi
Boston Jet Physics Workshop, 21.01.14

Quarks vs. Gluons: Handles

Handles:

- Radiation - gluons shower more
- Color connections - gluons color connected to two sources
- Track based observables (e.g. charge multiplicity)
    - all require good MC modeling of shower/hadronization/color flow
- Kinematics - q/g enrichment in different event topologies, kinematics

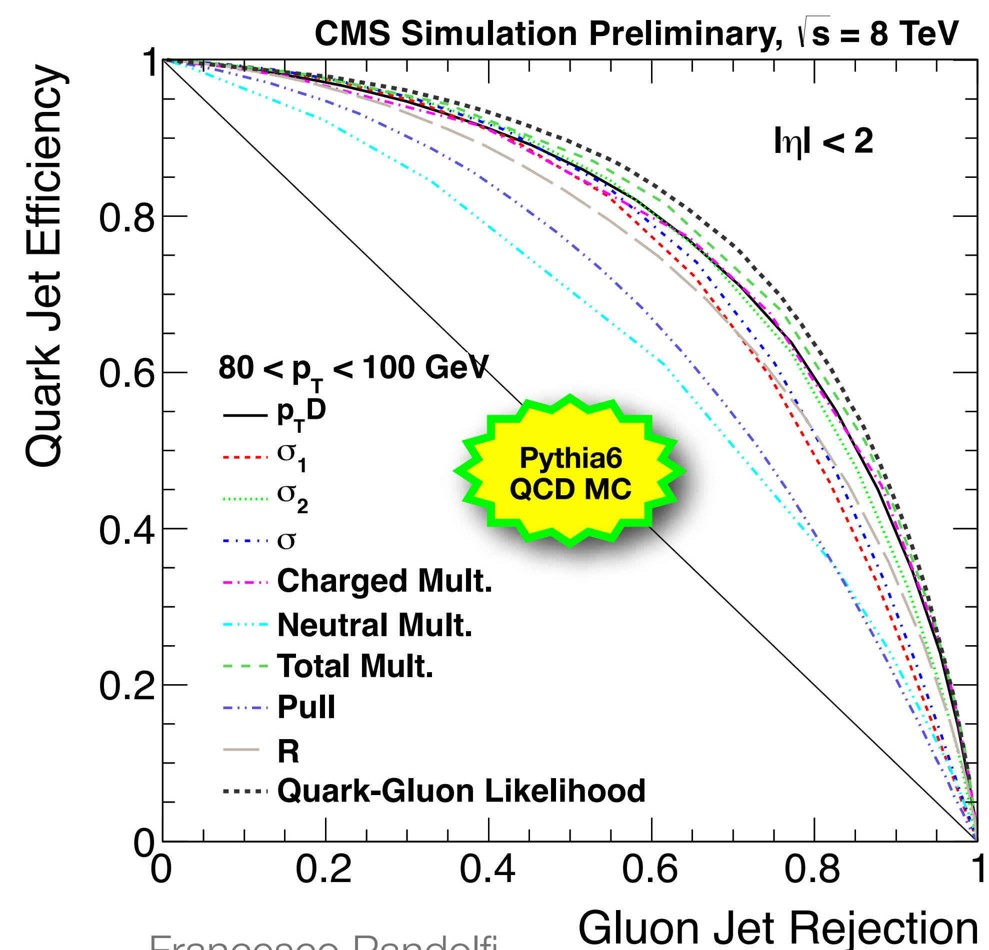
Quarks vs. Gluons: Results

observables used (CMS study)
break down into 3 categories:

- multiplicity (radiation, track based observables)
- width (color connections, radiation)
- energy sharing variables (color connections, kinematics)

kinematic selections may be used to obtain samples of (relatively) high purity q/g samples

challenge: this will bias the observables compared to full data sets

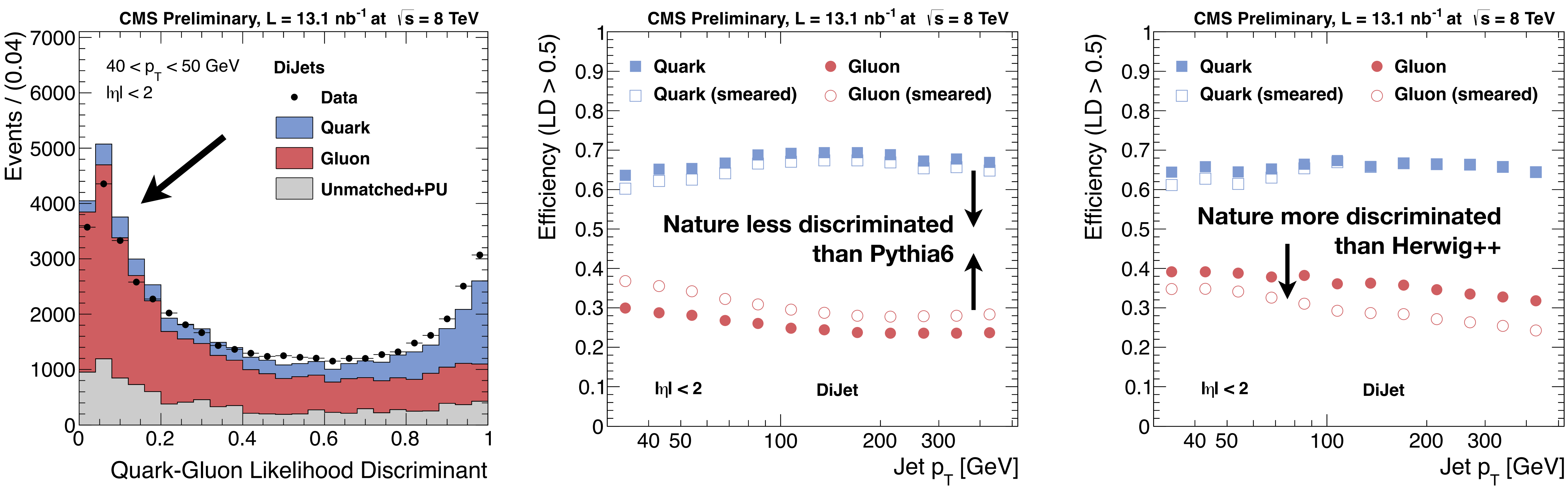


Francesco Pandolfi
Boston Jet Physics Workshop, 21.01.14

some theory papers:

1104.1175, 1106.3076, 1305.0007

Quarks vs. Gluons: Troubling Disagreements



Francesco Pandolfi Boston Jet Physics Workshop, 21.01.14

discrepancy in gluon enriched dijet sample

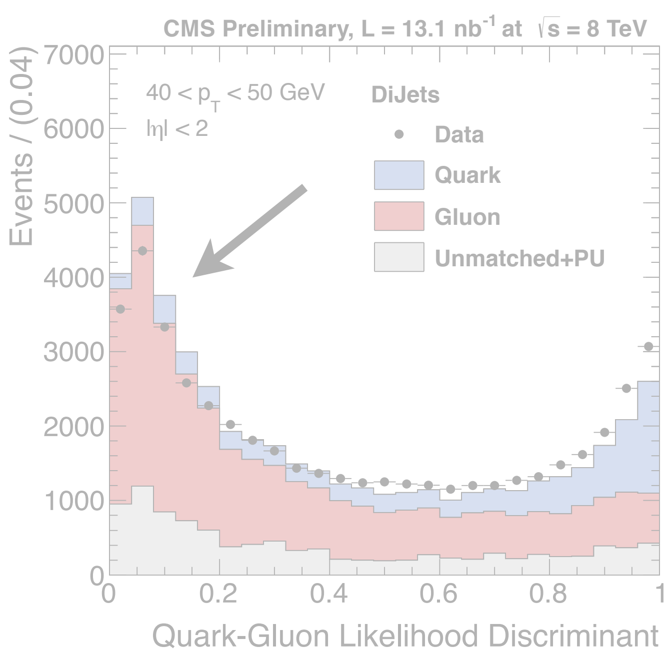
noticeable difference between data/MC, MC/MC

needs innovate approaches from theorists:

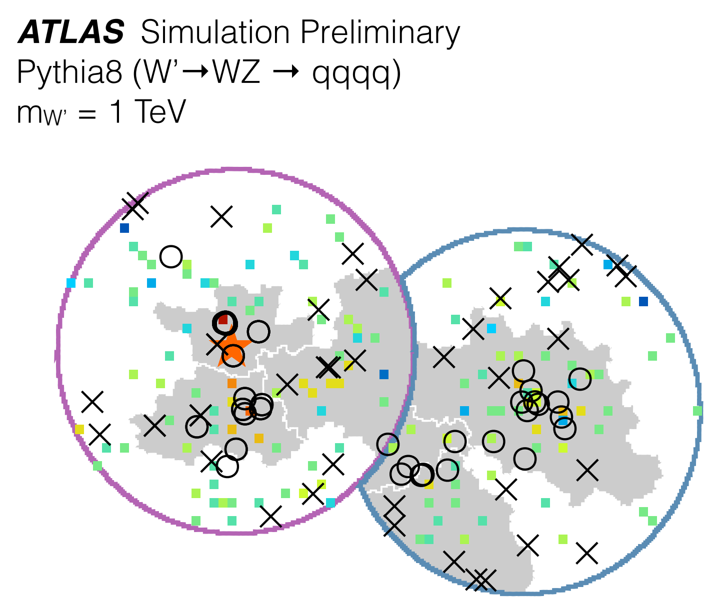
what precisely is driving the disagreement, and do we have any tools to address it (for MCs and also for QCD)? can we do better?

Outline

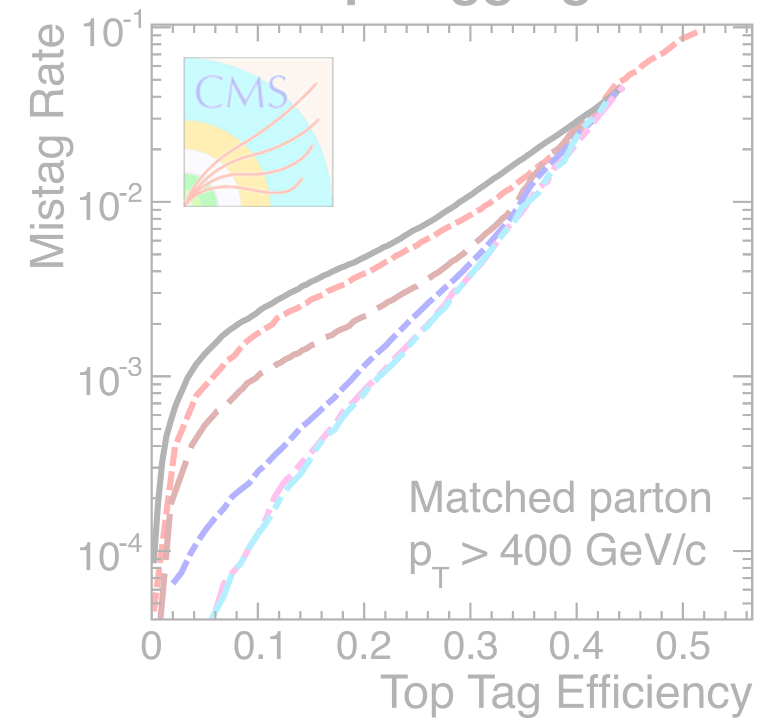
Quark/Gluon Tagging



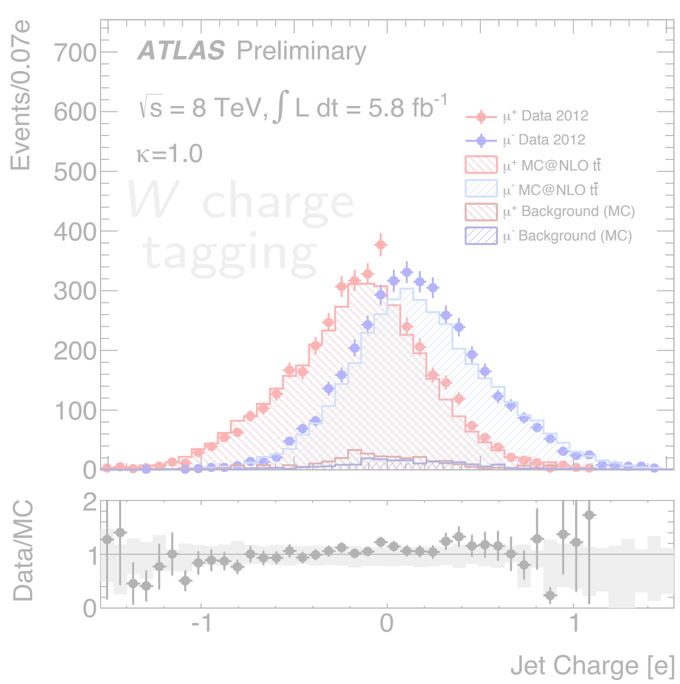
Jet Grooming



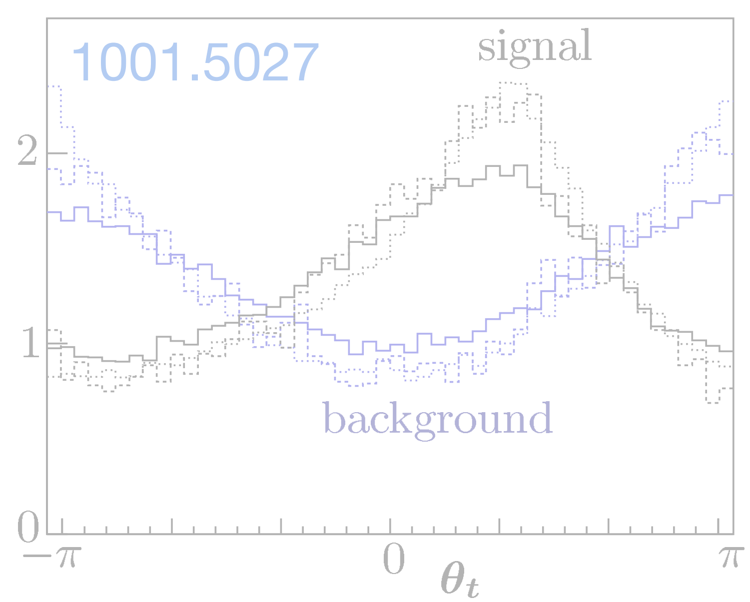
Top Tagging



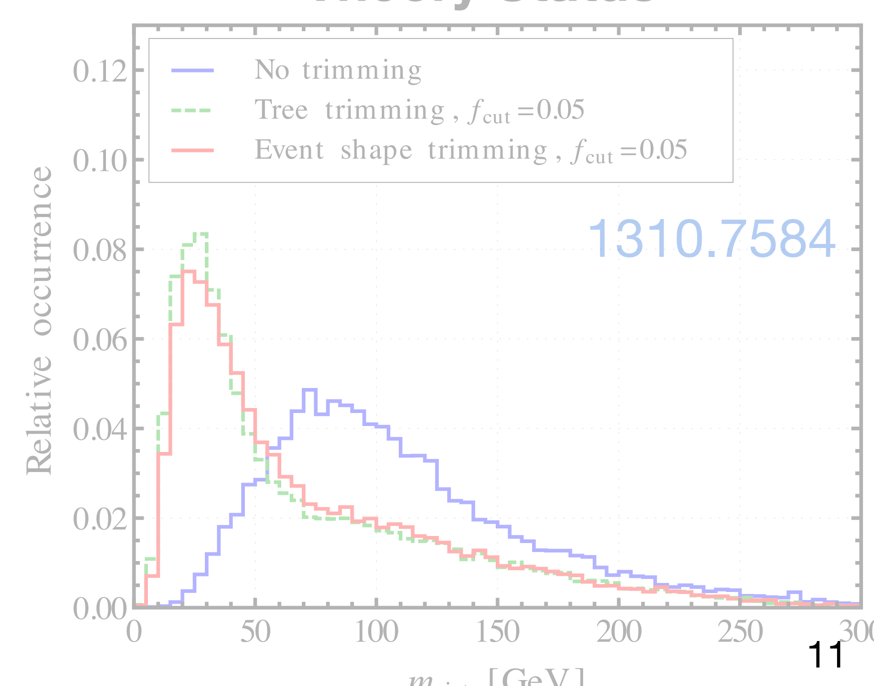
Tracks



Kinematics vs. Color Flow



Theory Status



Jet Grooming

jets are messy, and the environment is messy

- resolution in substructure degraded by soft, uncorrelated radiation
  (pileup, underlying event, jet algorithm effects, soft radiation)

- original grooming tools (mass-drop filter, pruning, trimming)
  designed to reduce sensitivity to this radiation

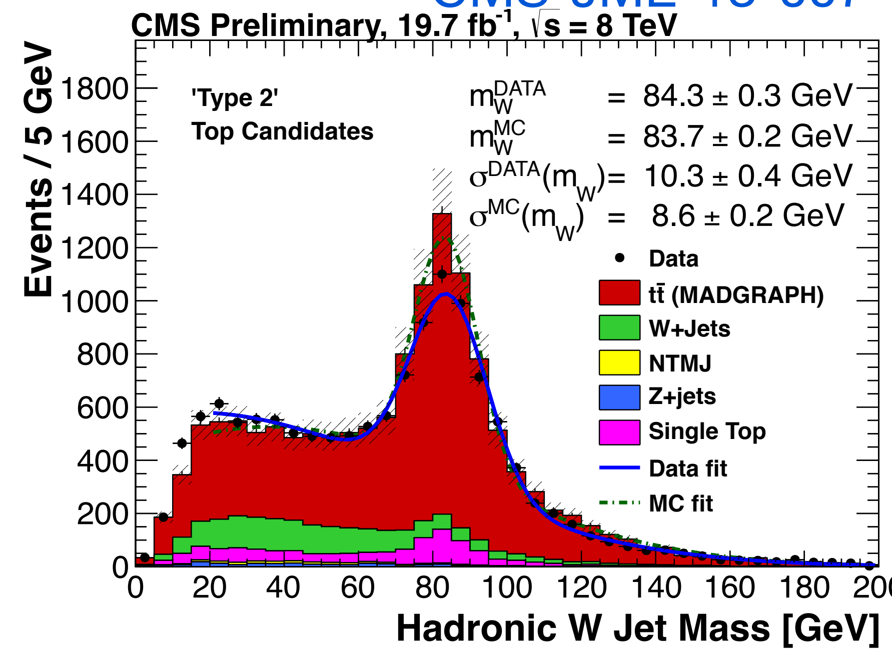
grooming tools used throughout substructure

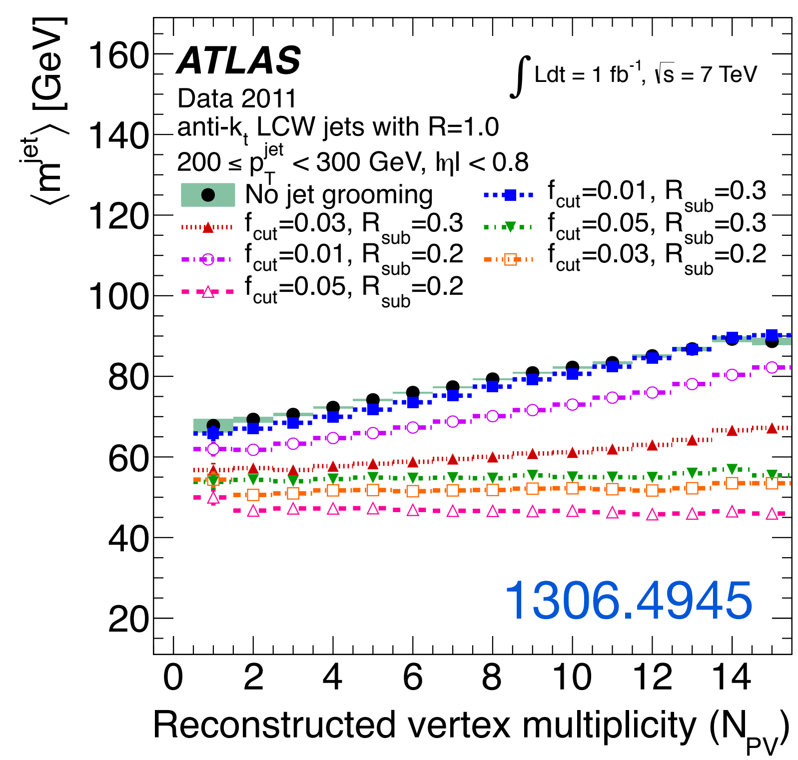
- as part of more complex algorithms

- for pileup reductions

- analytic studies of grooming give refined techniques

CMS-JME-13-007





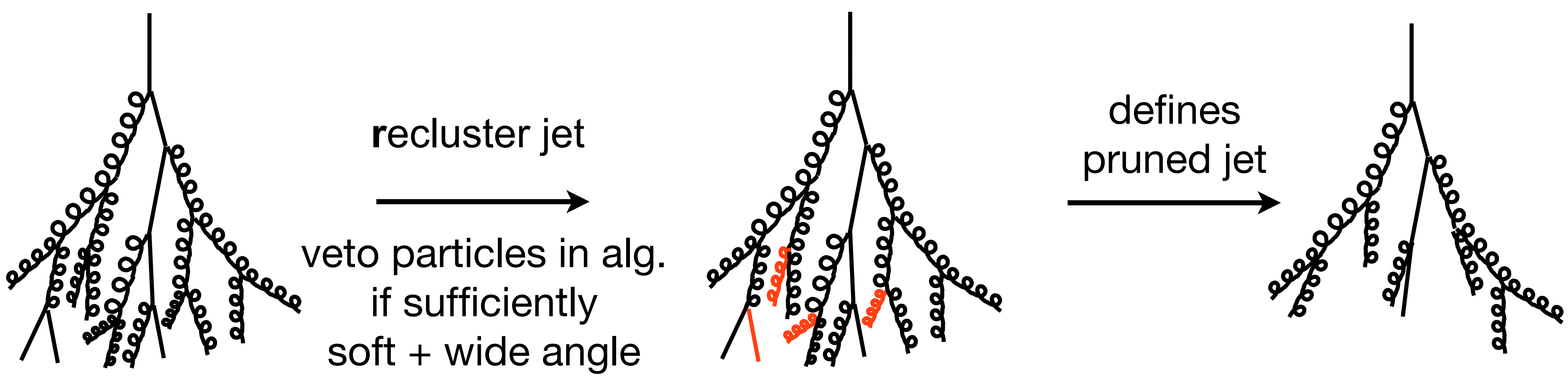
1306.4945

Jet Grooming: Algorithms

filtering

0802.2470

pruning



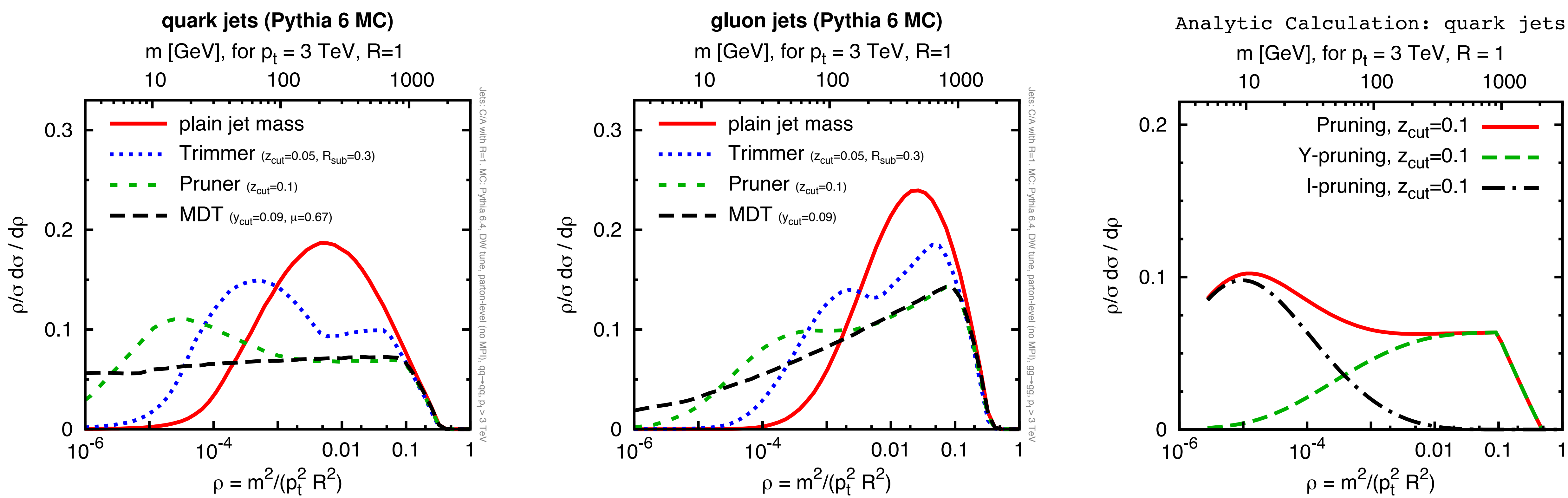
0903.5081
0912.0033

trimming

0912.1342

Analytic Properties of Jet Grooming

1307.0007
1307.0013



nice study of jet grooming's analytic properties

- fixed order, resummed results for jet mass
- nontrivial parameter dependence
- notable differences from ungroomed jets
  (single log vs. double log resummation in some regimes)
- leads to mMDT, a modified mass drop tagger with better analytic properties

Jet Cleansing

1309.4777

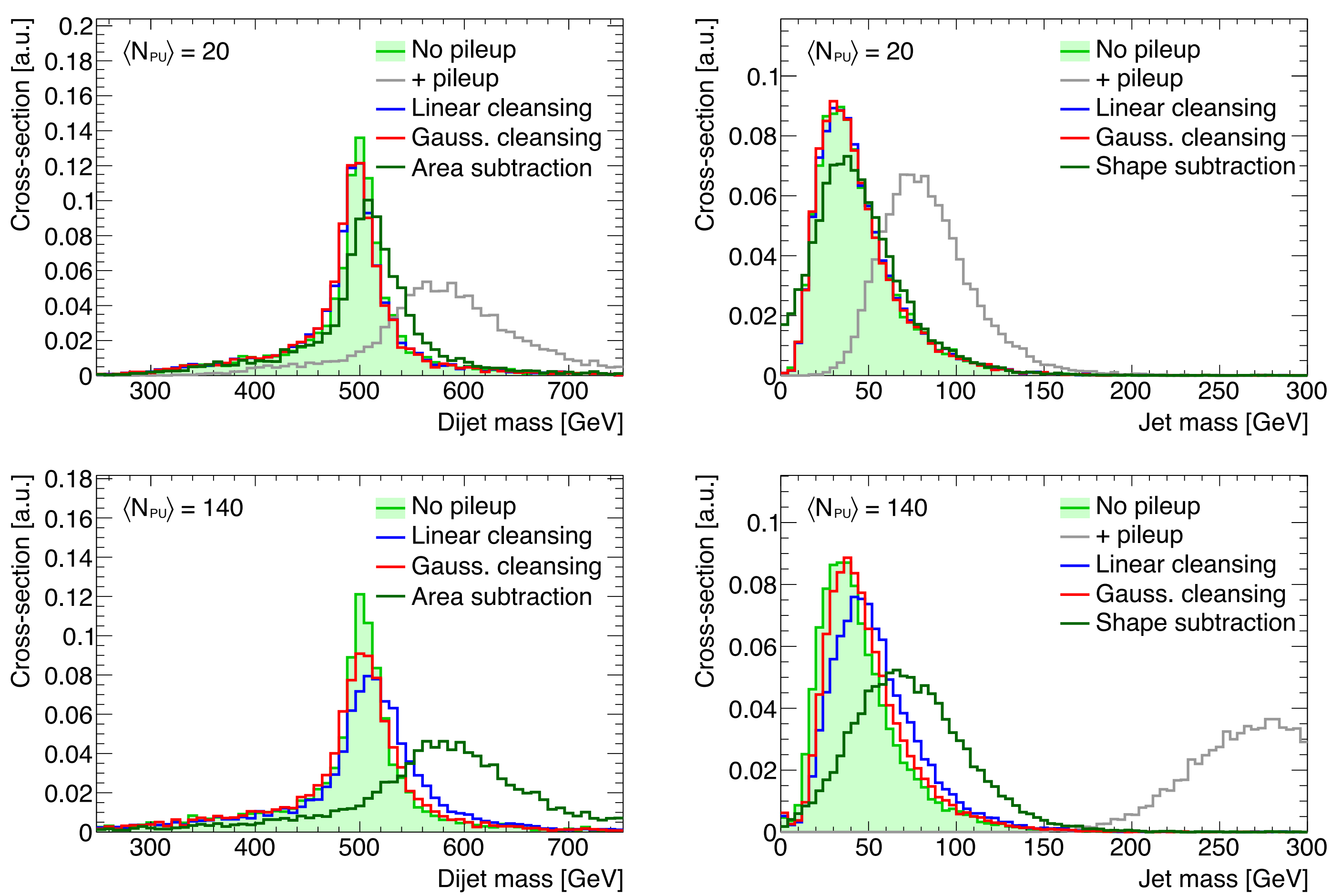
algorithm:

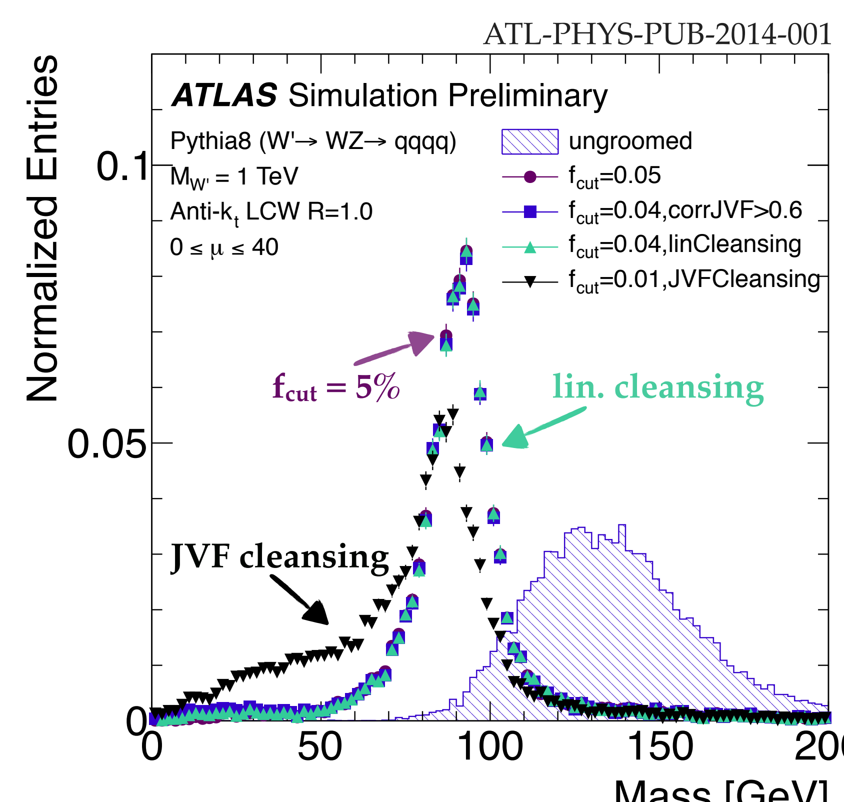
- cluster jet into subjets
- rescale subjet-by-subjet
- reassemble jet

$$\lambda_{\text{sub}} = \frac{\text{charged } p_T \text{ from LV}}{\text{all charged } p_T}$$ JVF cleansing

$$\lambda_{\text{sub}} = \frac{p_T \text{ from LV}}{p_T \text{ from LV} + p_T \text{ from PU}}$$ linear cleansing

$$\lambda_{\text{sub}} = 1 - \frac{p_T^{C,PU}}{\gamma_0 p_T}$$ Gaussian cleansing

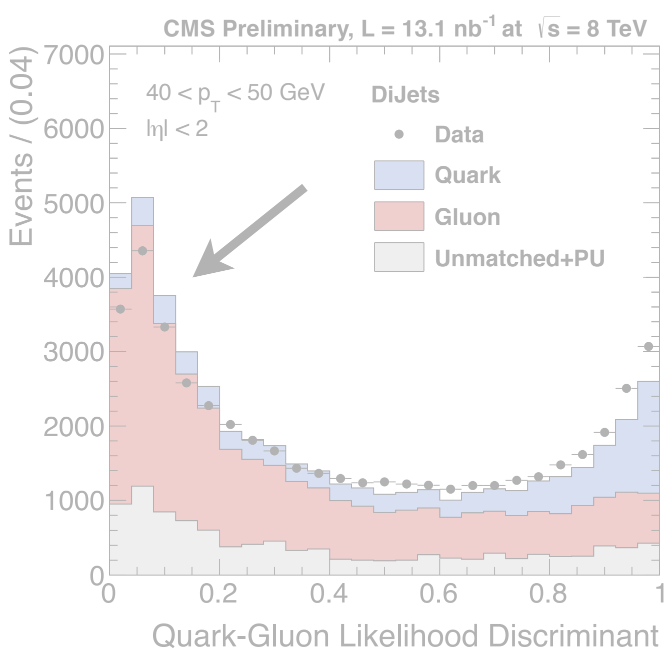




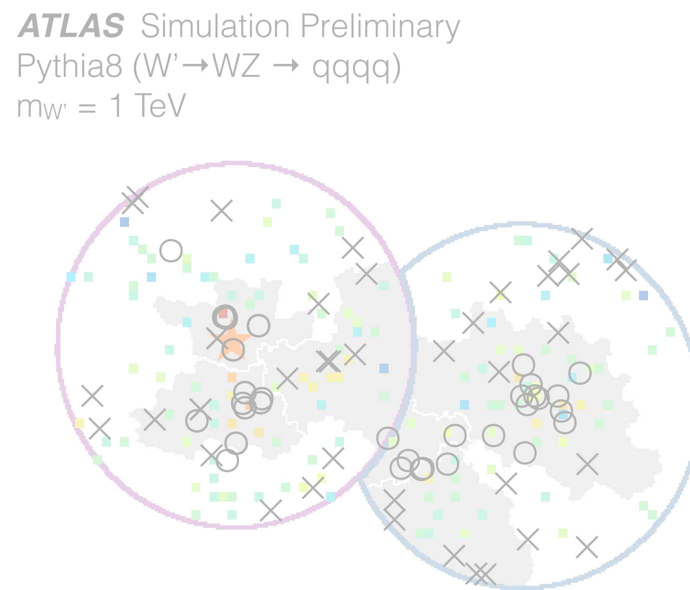
Pascal Nef,
Boston Jets Workshop 2014

Outline

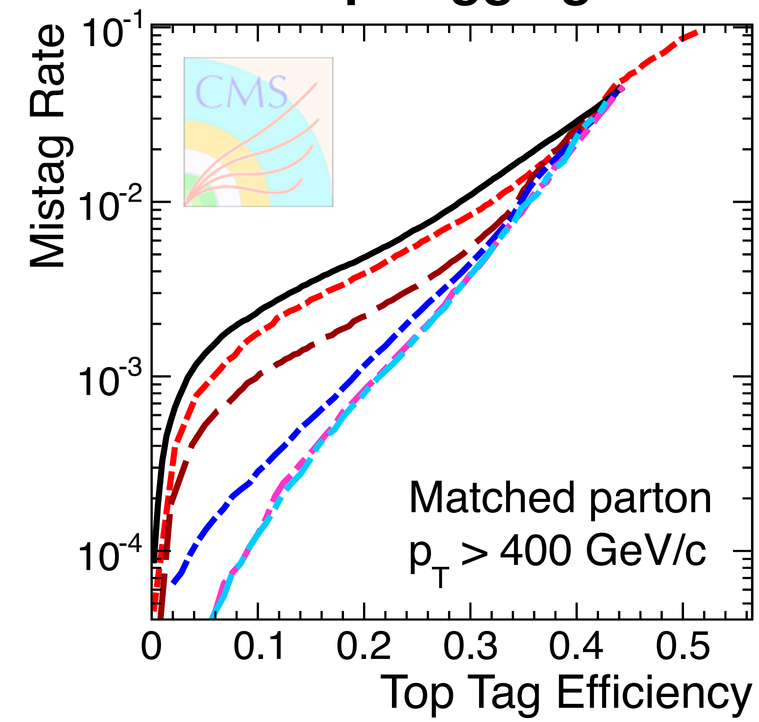
Quark/Gluon Tagging



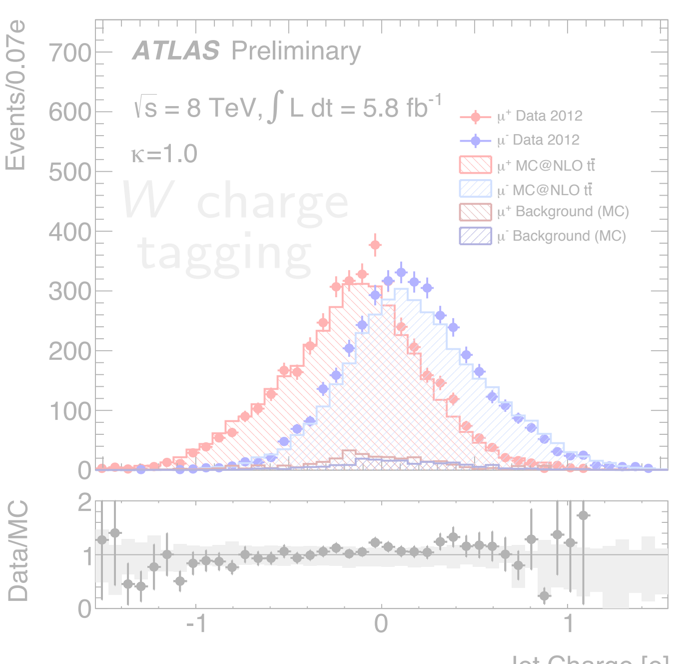
Jet Grooming



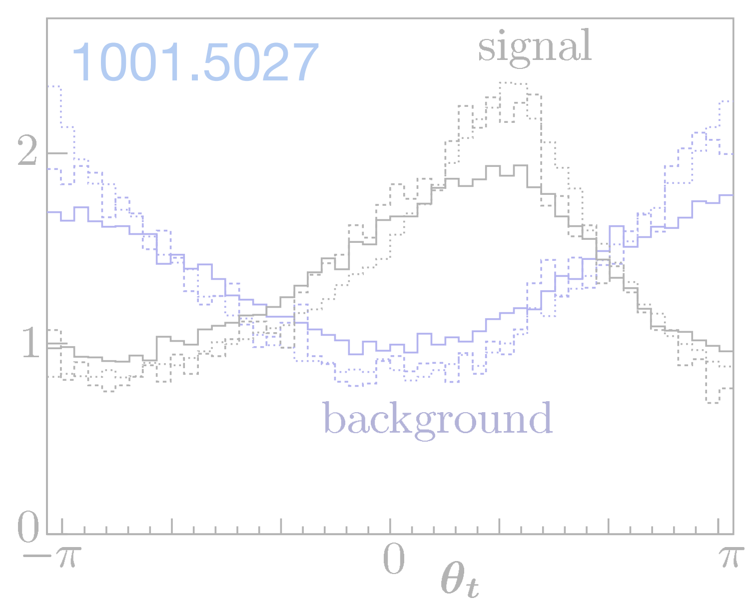
Top Tagging



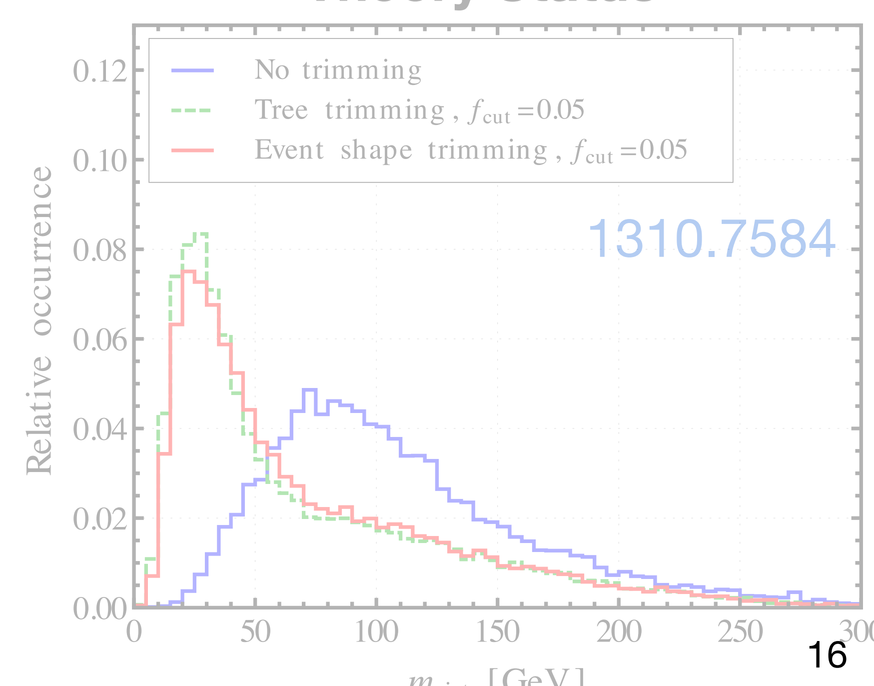
Tracks



Kinematics vs. Color Flow



Theory Status

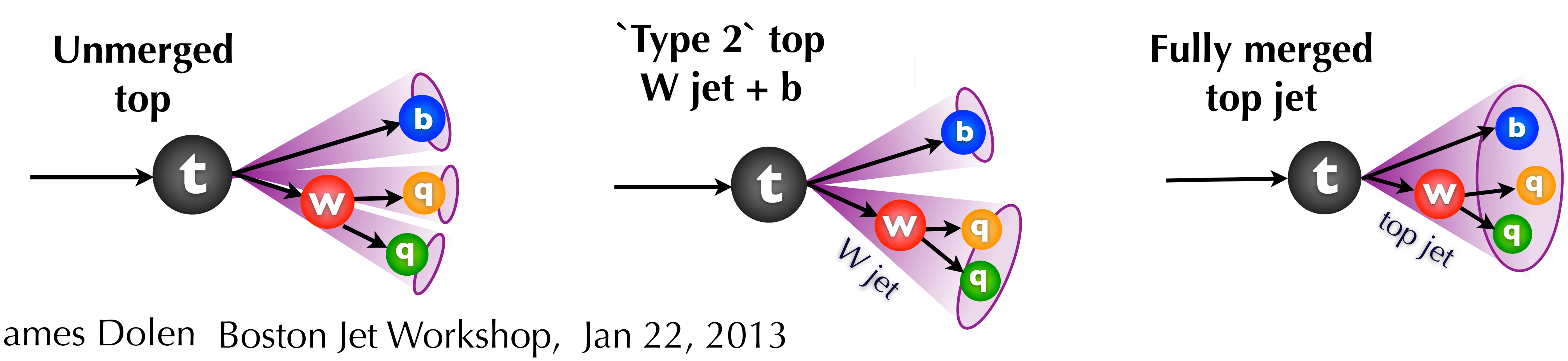


Top Tagging: Overview

boosted tops a signature in different new physics scenarios

many handles for substructure, wide kinematic ranges

- hard kinematics of W, top
- shape variables for color flow, charge
- high ttbar rate to validate

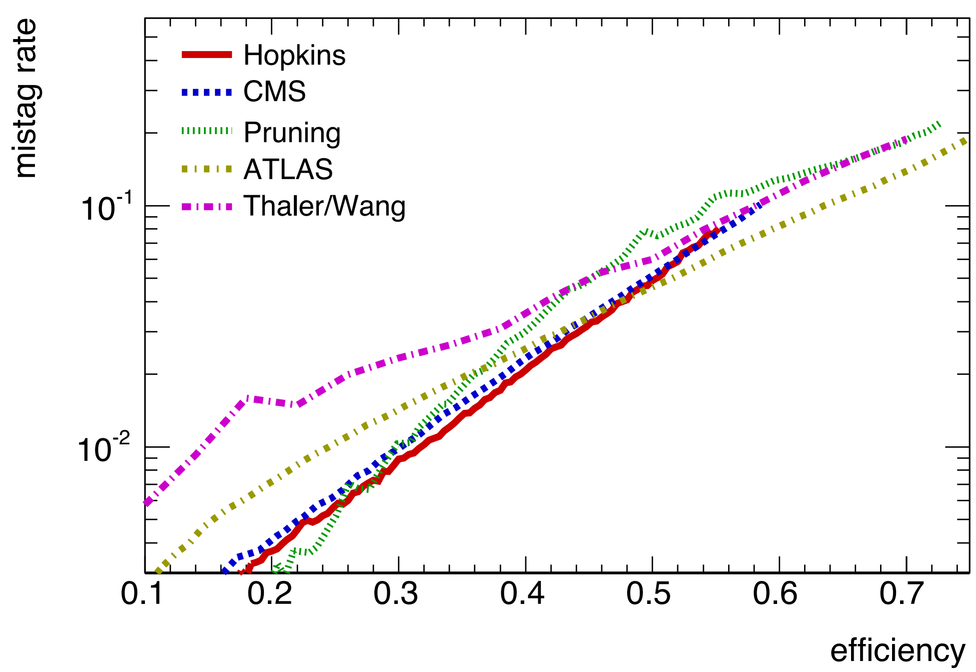


Top Tagging: Tools

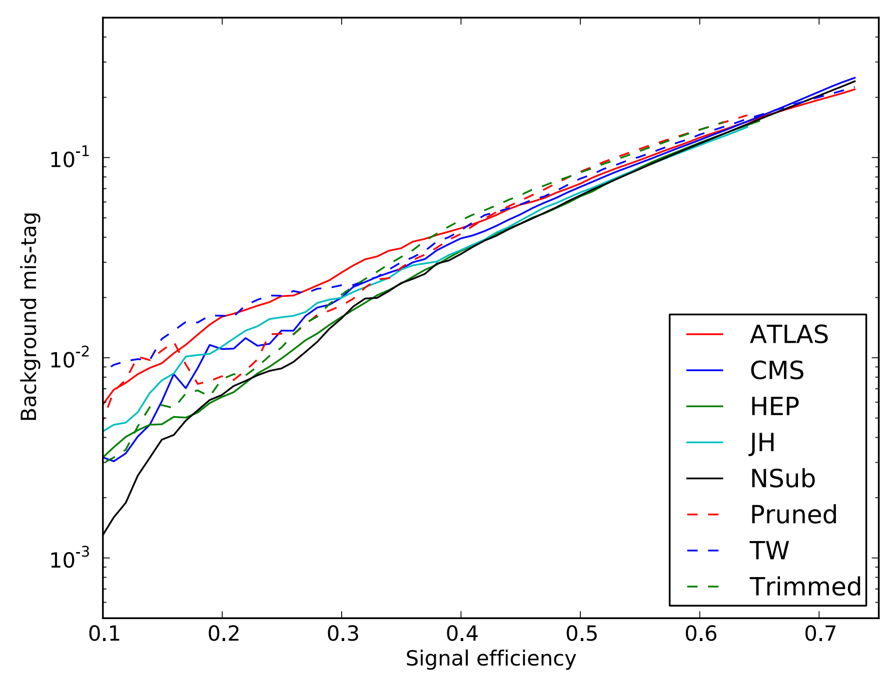
many different types of tools:

ATL-PHYS-PUB-2010-008, CMS-JME-2013-007

- subjet counting, reconstruction tools: JHU, HEP, ATLAS, CMS taggers
  0806.0848, 1006.2833
- shape variables: Y-splitter, N-subjettiness, angular structure function
  ATL-PHYS-CONF-2008-008, 1011.2268, 1108.2701 1201.2688
- jet grooming may be used alone or in concert with shape tools



Boost 2010: 1012.5412

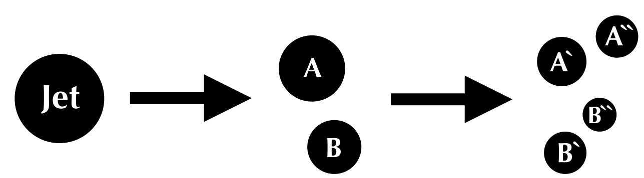


Boost 2011: 1201.0008

Top Tagging: Tools II

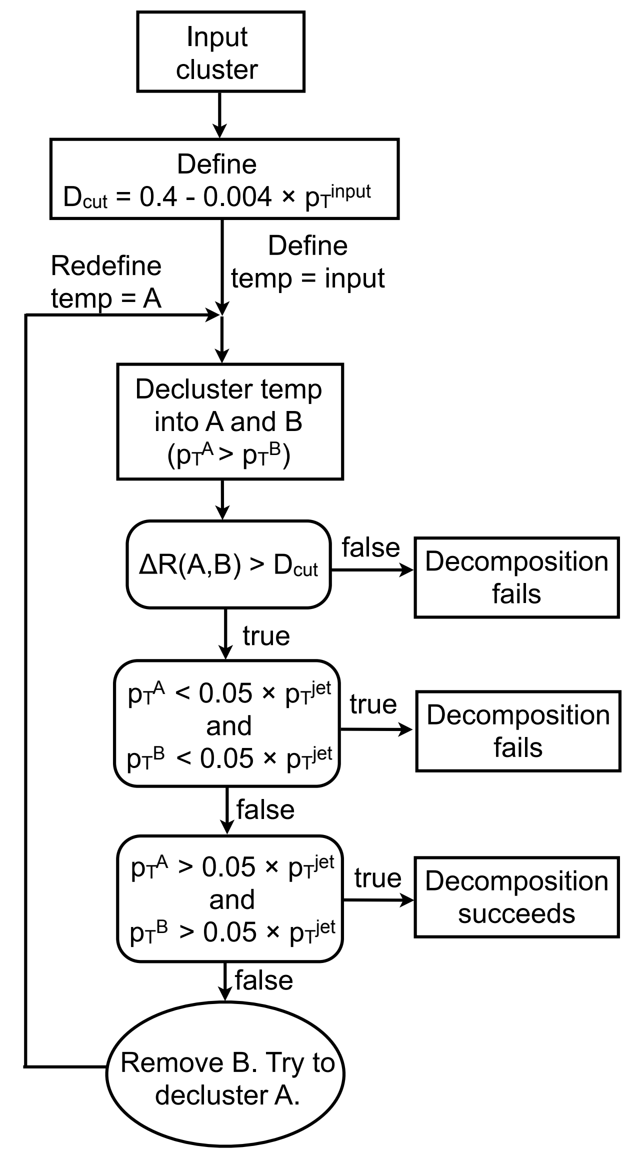
CMS Top Tagger

- Based on JHU Top Tagger (Kaplan et al., Phys.Rev.Lett. 101 (2008) 142001)
- Cluster a jet using a sequential recombination algorithm (CA R=0.8)
- Decluster in two stages in order to find up to 4 subjets



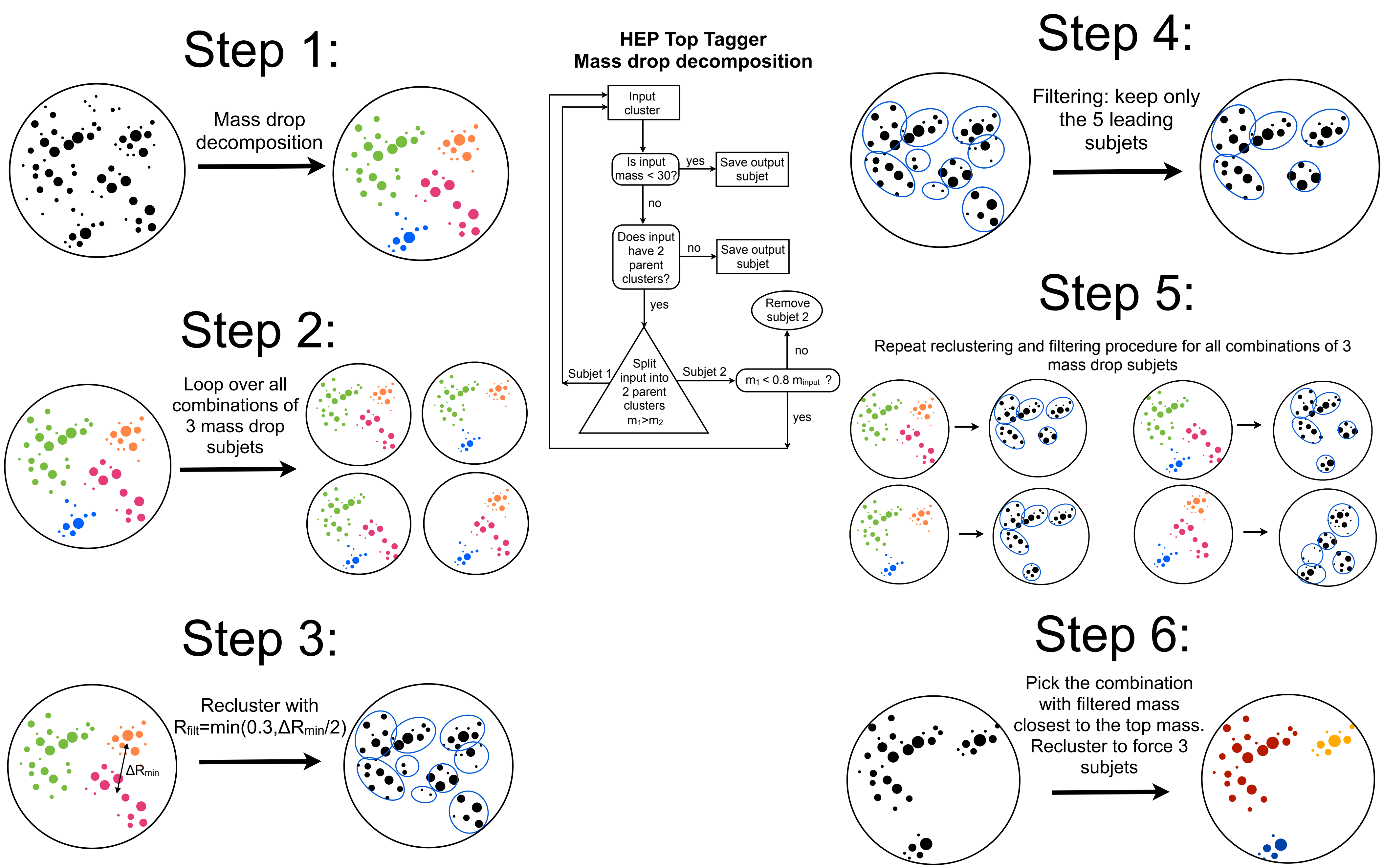
- Subjets must satisfy two requirements
  - Momentum fraction criterion: $p_T^{subjet} > 0.05 \times p_T^{jet}$
  - Adjacency criterion: $\Delta R(C_1, C_2) > 0.4 - 0.0004 \times pT(C)$
- Iterative process - throw out subjets that fail momentum fraction cut and try to decluster again
- Tag top jets with selections on the CA8 jet mass, number of subjets, and minimum pairwise subjet mass ($m_{min} = \min[m_{12}, m_{13}, m_{23}]$)

CMS Top Tagger decomposition



Top Tagging: Tools III

HEP Top Tagger details



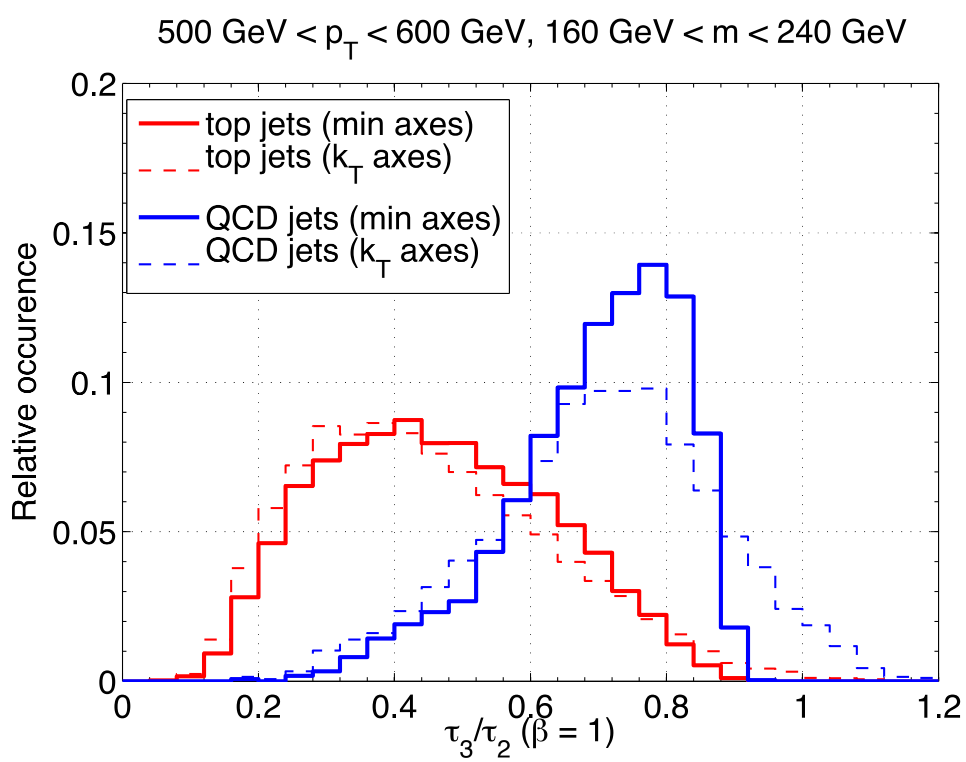
James Dolen Boston Jet Workshop, Jan 22, 2013 12

Top Tagging: Tools IV

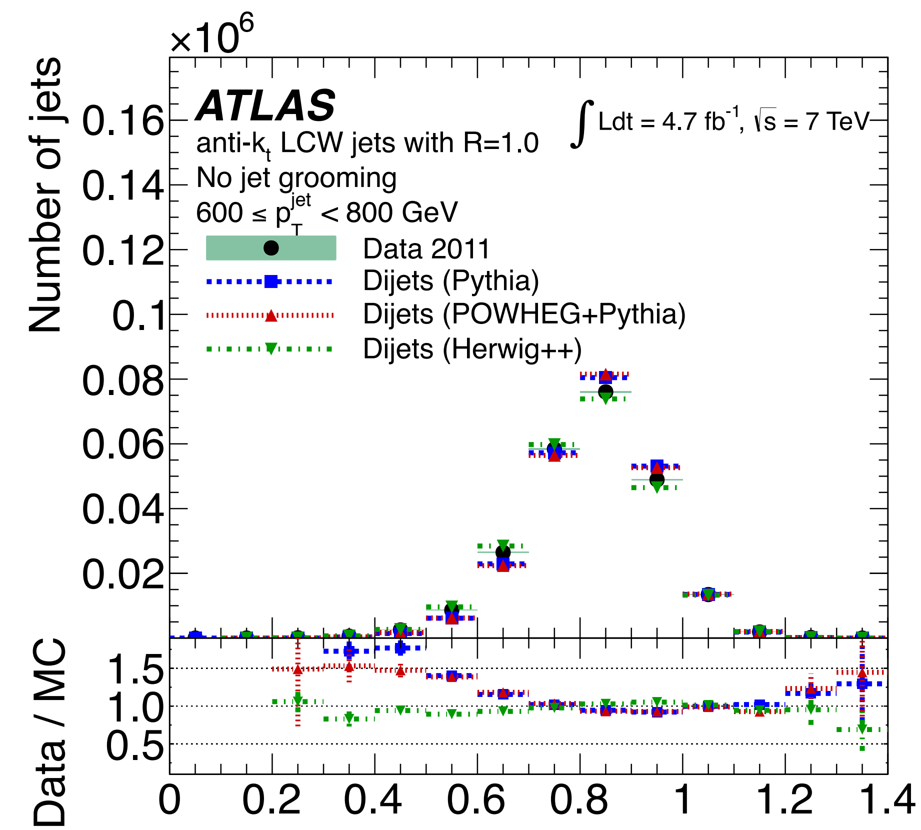
N-subjettiness: $$\tilde{\tau}_N^{(\beta)} = \frac{1}{d_0} \sum_i p_{T,i} \min \left\{ (\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta, \ldots, (\Delta R_{N,i})^\beta \right\}$$

ratio of N-subjettinesses is an effective counter for the number of subjets (3 for tops)

$$\tau_3/\tau_2 \sim \begin{cases} \mathcal{O}(1) & \text{for } < 3 \text{ subjets} \quad [\tau_3 \ll 1\,, \tau_2 \ll 1] \\ \ll 1 & \text{for } 3 \text{ subjets} \quad [\tau_3 \ll 1\,, \tau_2 \sim 1] \\ \mathcal{O}(1) & \text{for } > 3 \text{ subjets} \quad [\tau_3 \sim 1\,, \tau_2 \sim 1] \end{cases}$$

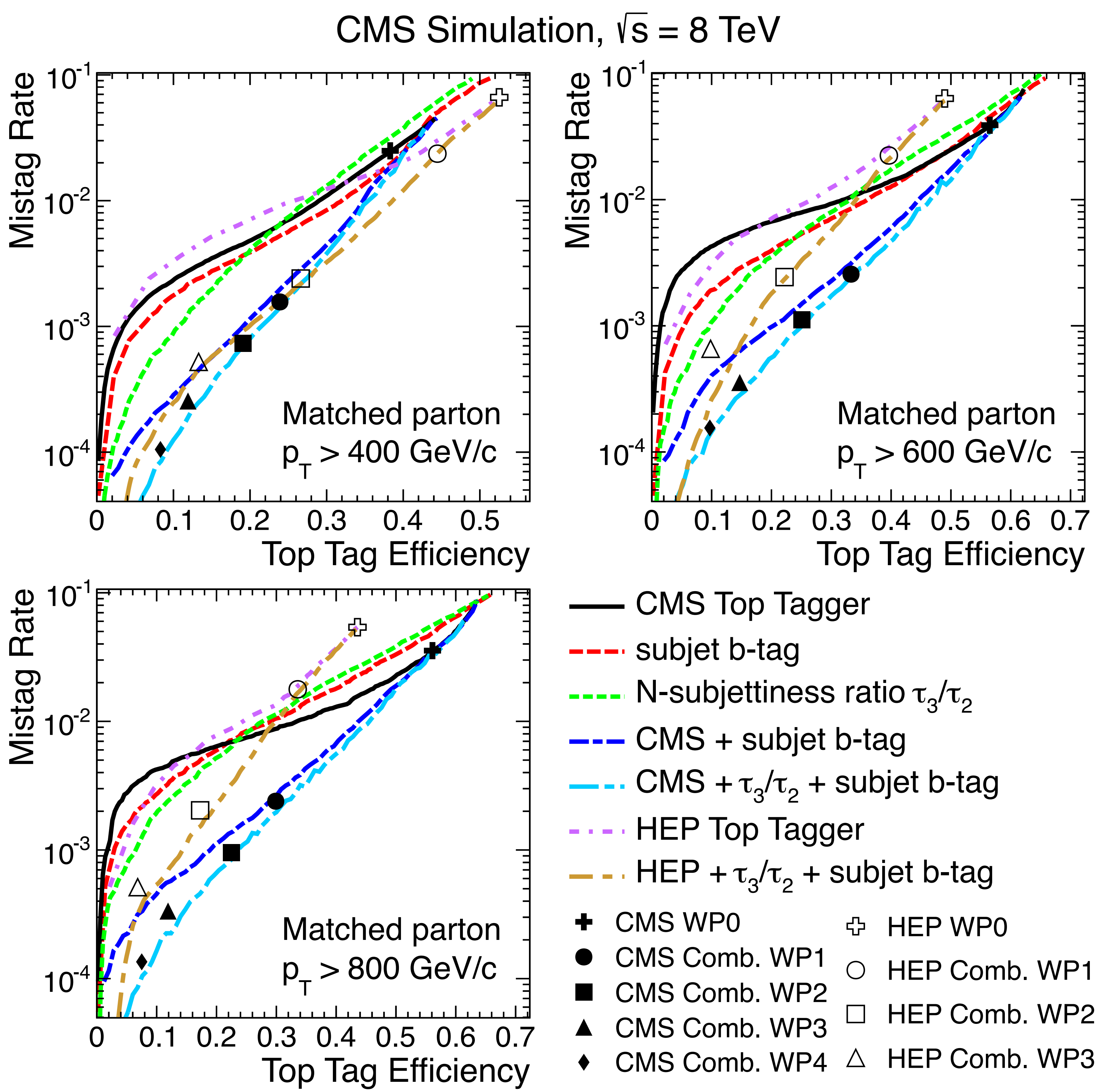


1108.2701



1306.4945

Top Tagging: Results



combination of tools:

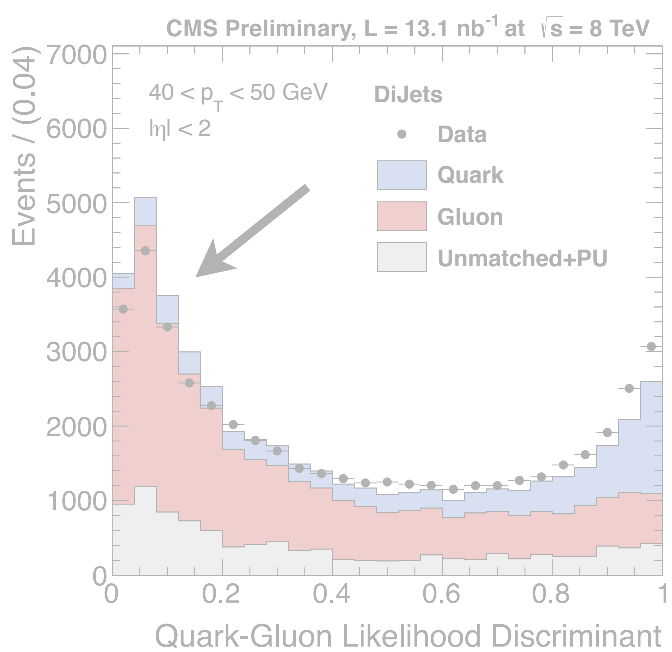
- HEP, CMS, N-subjettiness, subjet b-tagging
- outstanding performance across a wide range of efficiency

remaining improvements possible at low top p_T, where there is a transition to multijet techniques

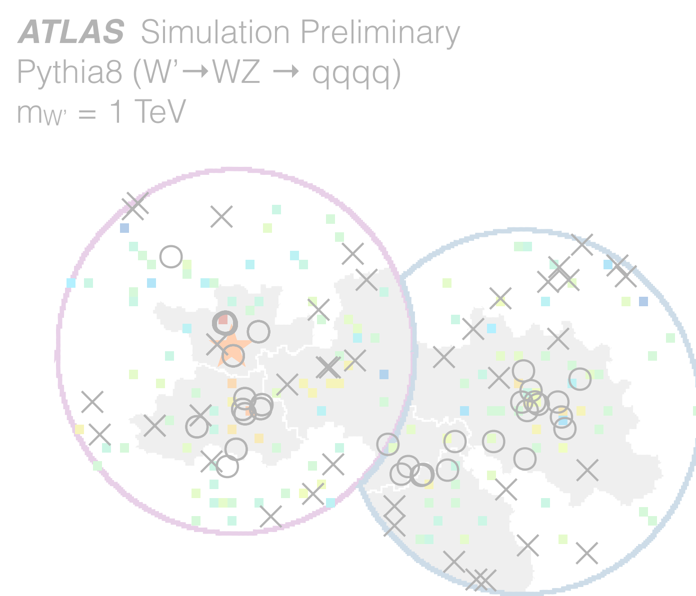
- can effectively tag W jet + b jet tops

Outline

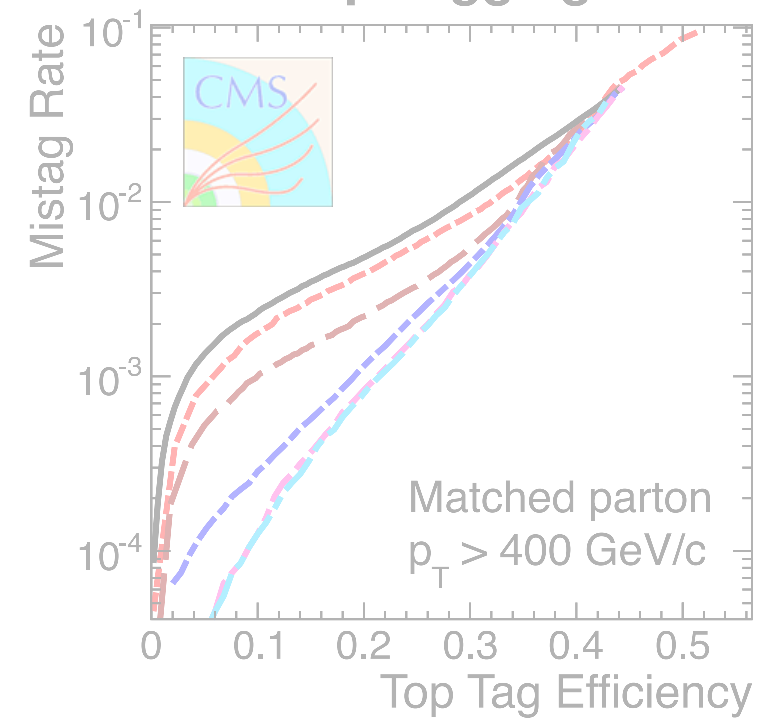
Quark/Gluon Tagging



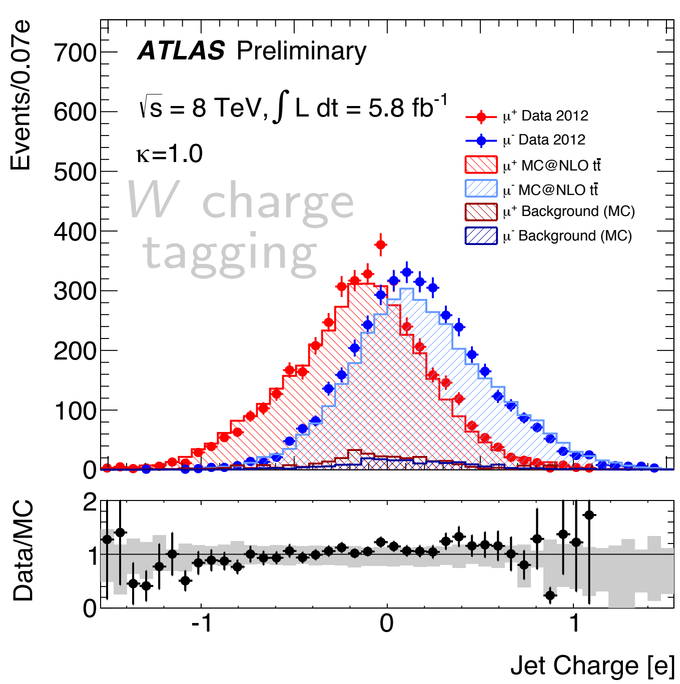
Jet Grooming



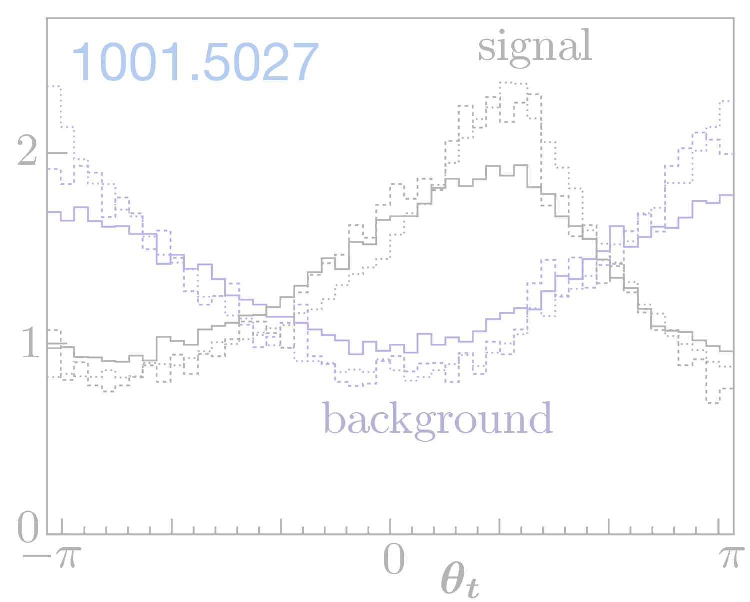
Top Tagging



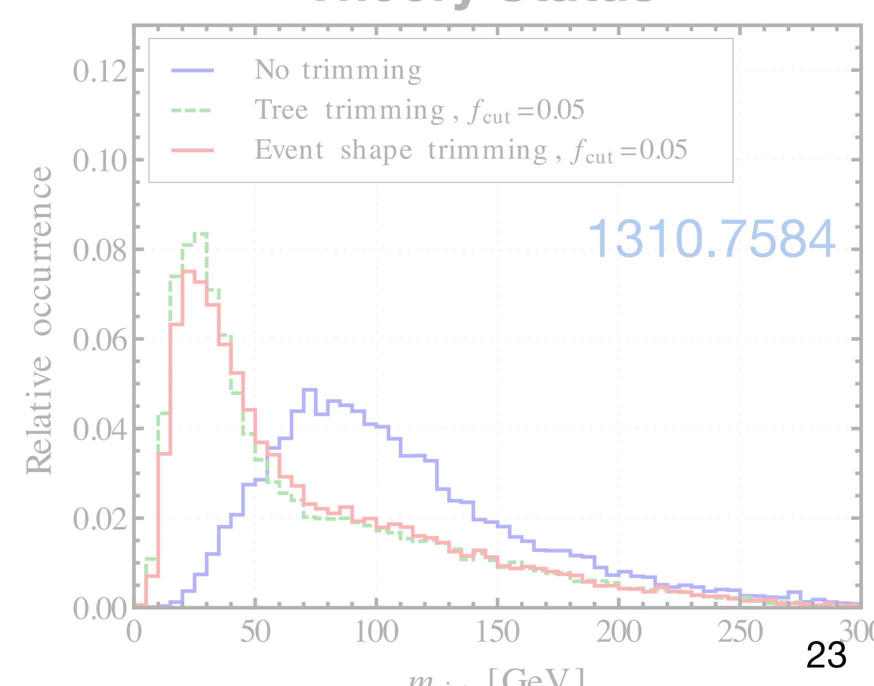
Tracks



Kinematics vs. Color Flow



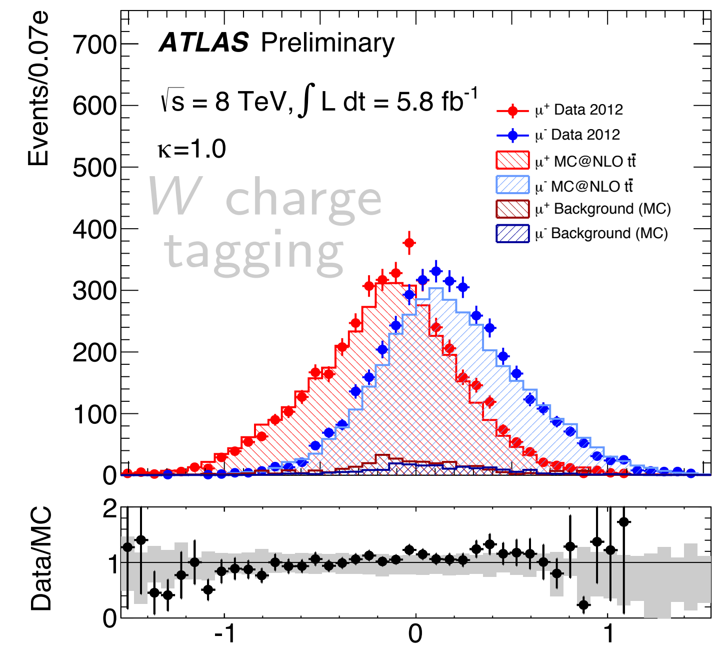
Theory Status



Tracks: Overview

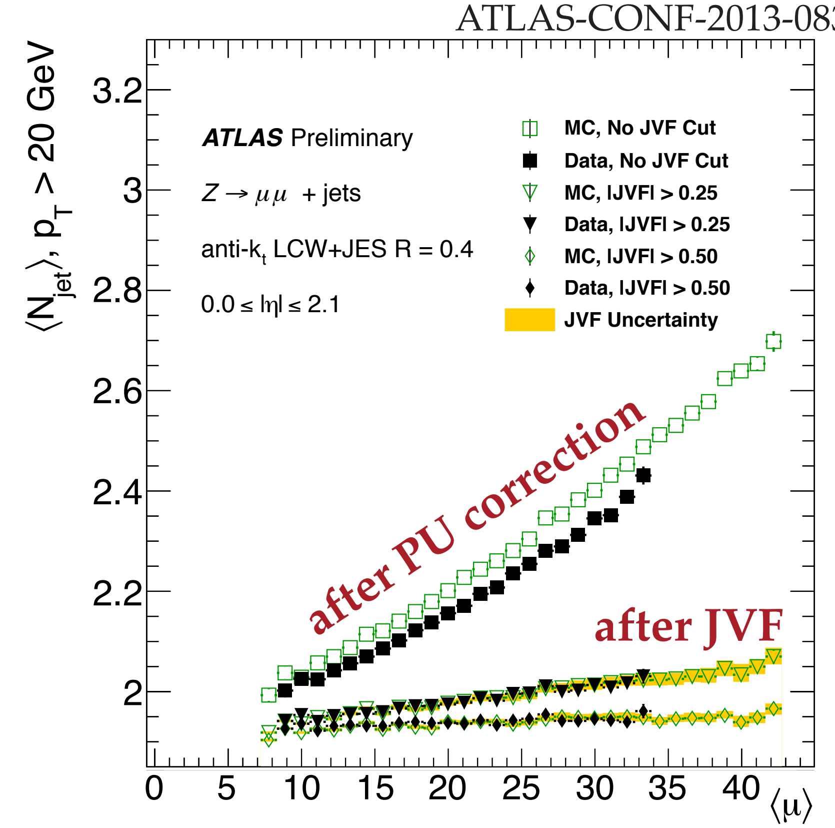
track-based observables useful in several contexts:

- pileup suppression (for ATLAS)
- tests of our understanding of pQCD
  - most observables are collinear IR unsafe (like PDFs), have non-perturbative input
- q/g tagging
  - one of the first discriminants proposed is charge track multiplicity



ATLAS-CONF-2013-086

ATLAS-CONF-2013-083



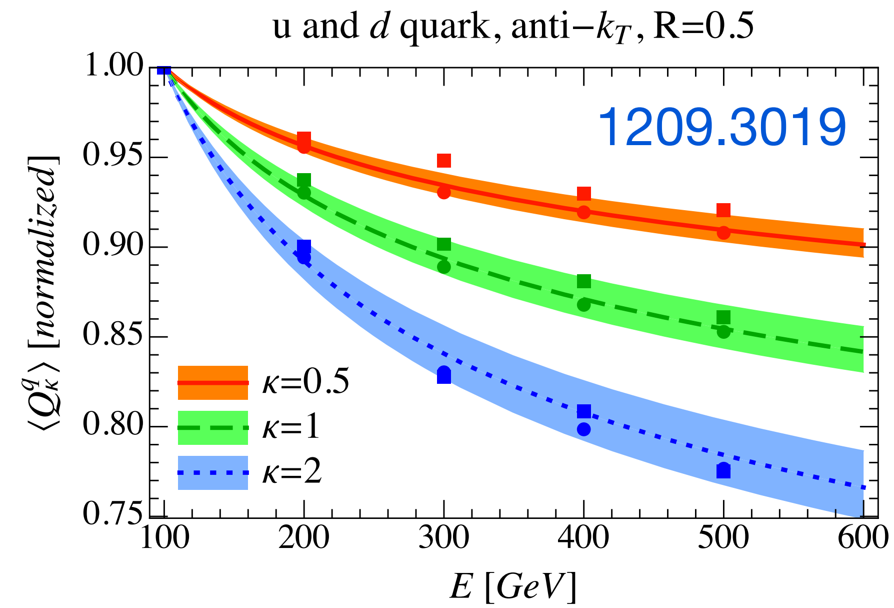
Jet Charge

1209.2421
1209.3019

jet charge: $\mathcal{Q}_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$

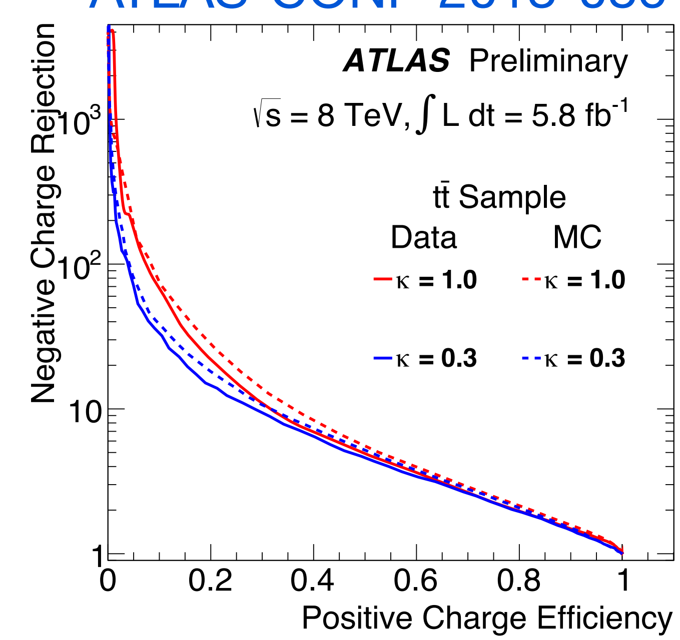
track charge

p_T weighting factor



calculable but IR collinear unsafe (similar to PDFs), depends on fragmentation functions comparisons to MC encouraging

ATLAS-CONF-2013-086



tagging W^+, rejecting W^- in semileptonic ttbar

one example of a jet shape that can be defined on tracks, theoretical framework interesting
see also 1303.6637, 1306.6630

jet charge with boosted objects is actively being studied, likely to add nontrivial dimension to high p_T tagging

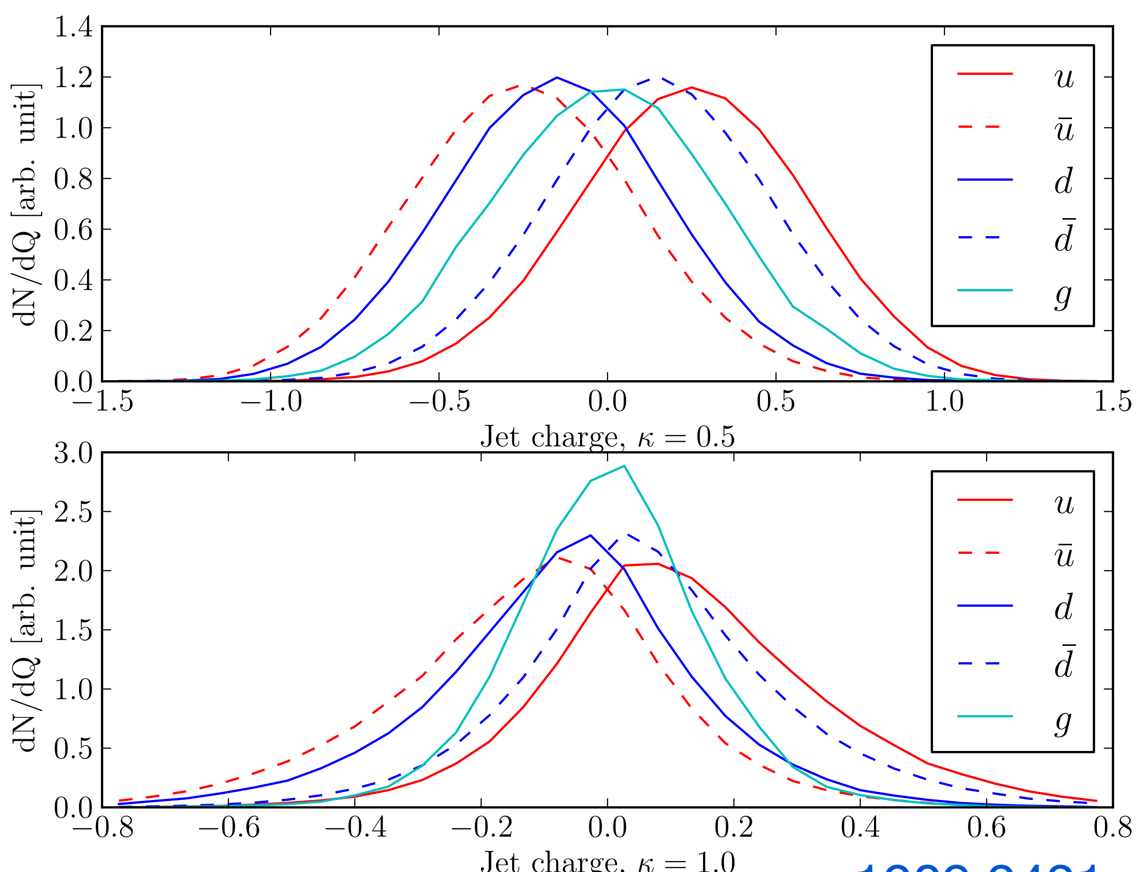
Jet Charge

1209.2421
1209.3019

jet charge: $\mathcal{Q}_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$

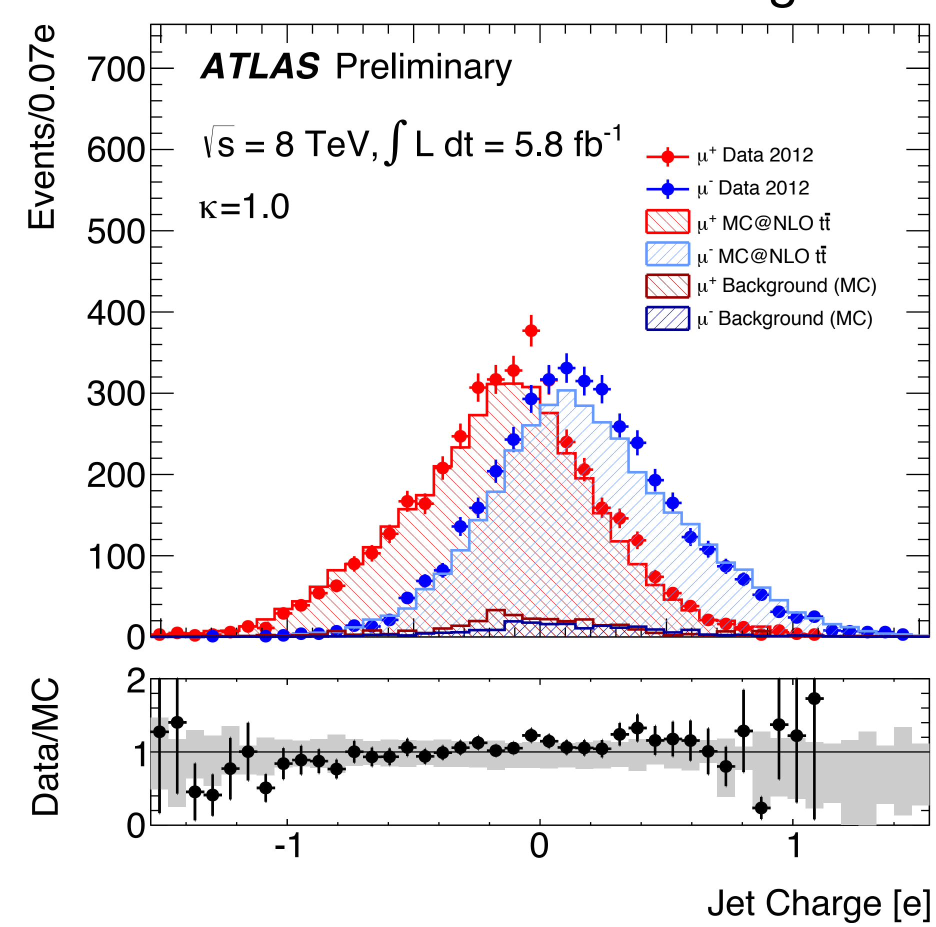
track charge

p_T weighting factor



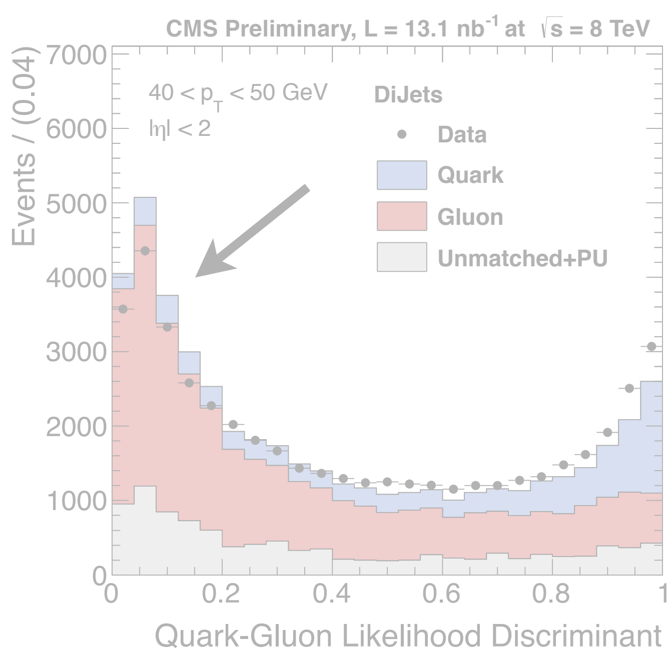
1209.2421

hadronic W+/W- charge

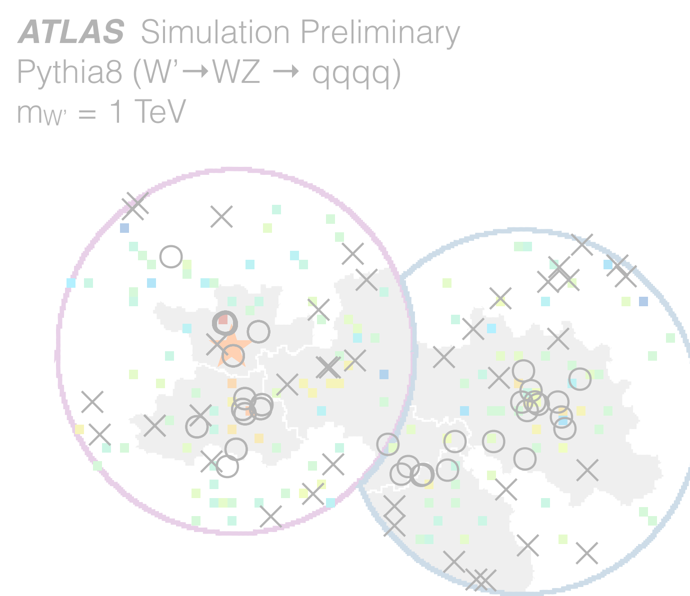


Outline

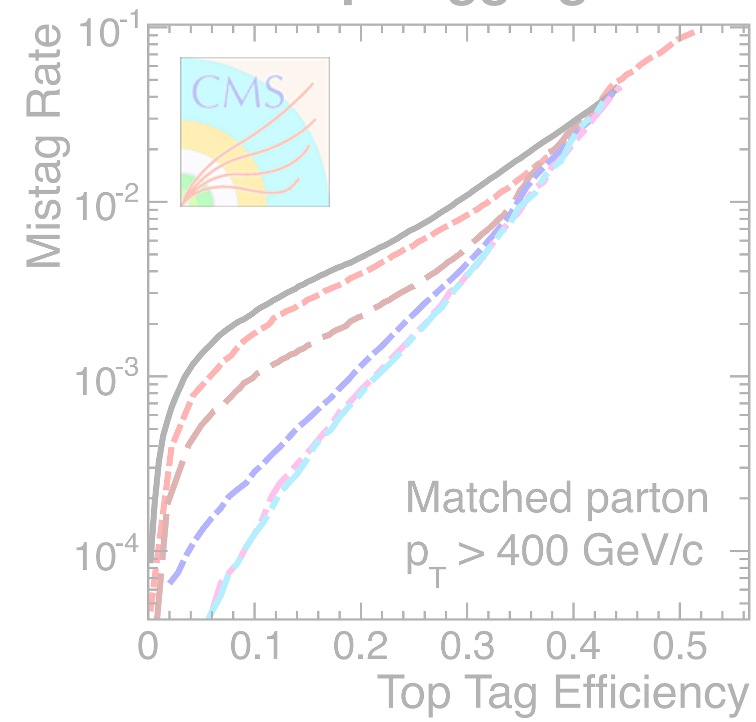
Quark/Gluon Tagging



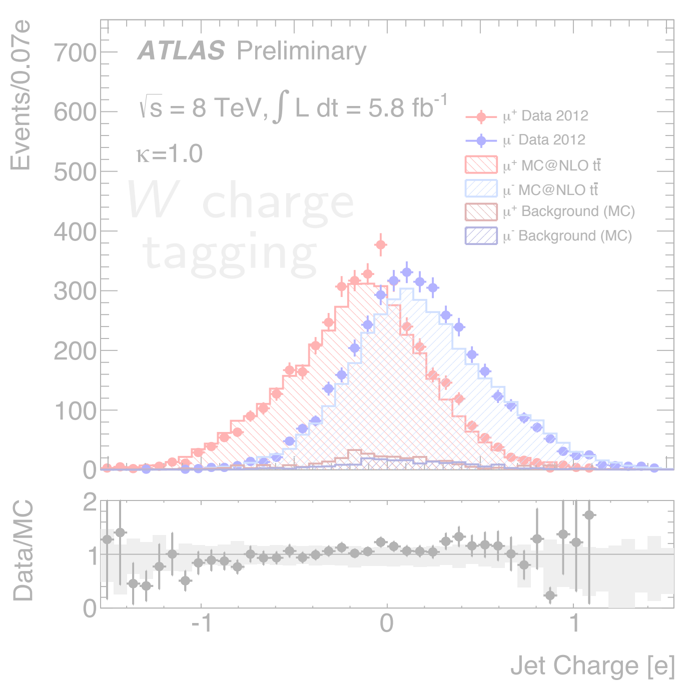
Jet Grooming



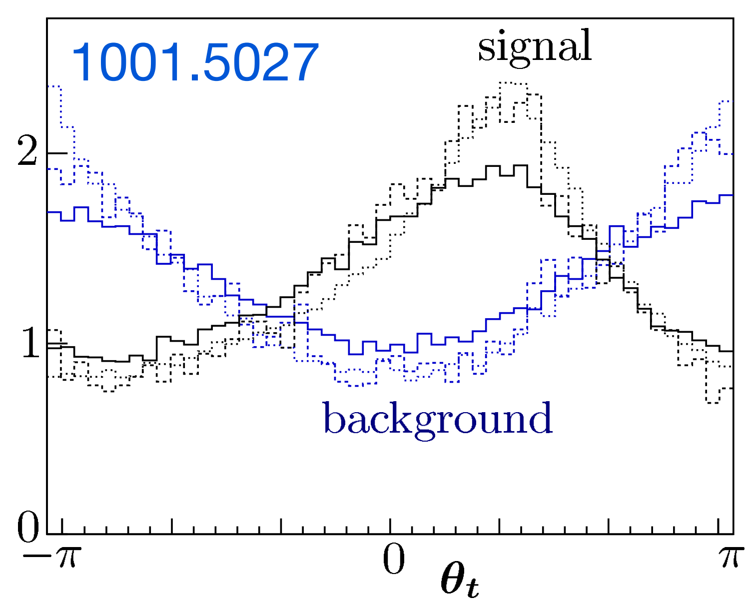
Top Tagging



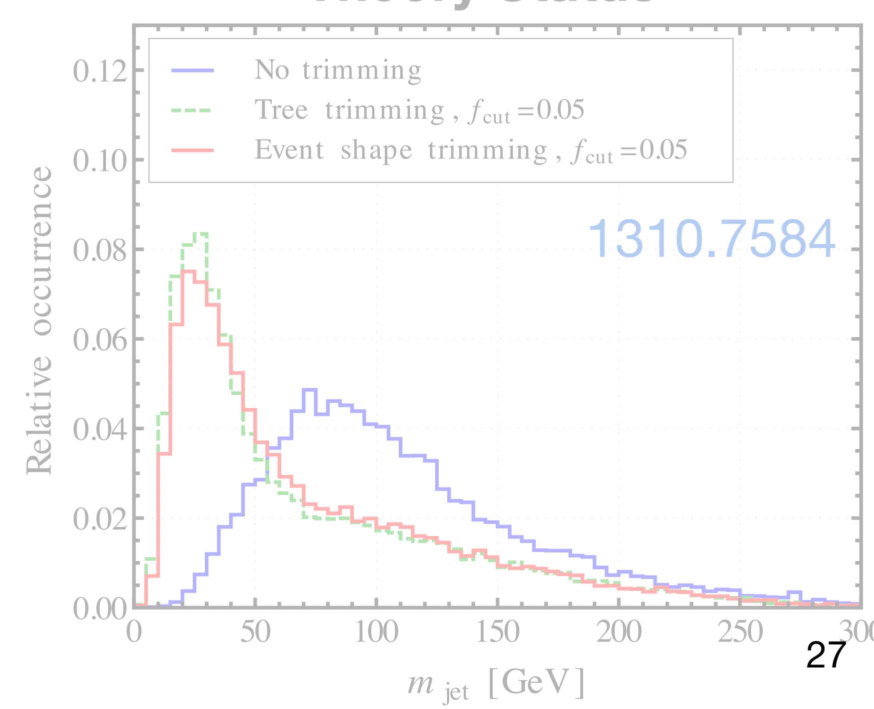
Tracks



Kinematics vs. Color Flow



Theory Status



Kinematic Discriminants

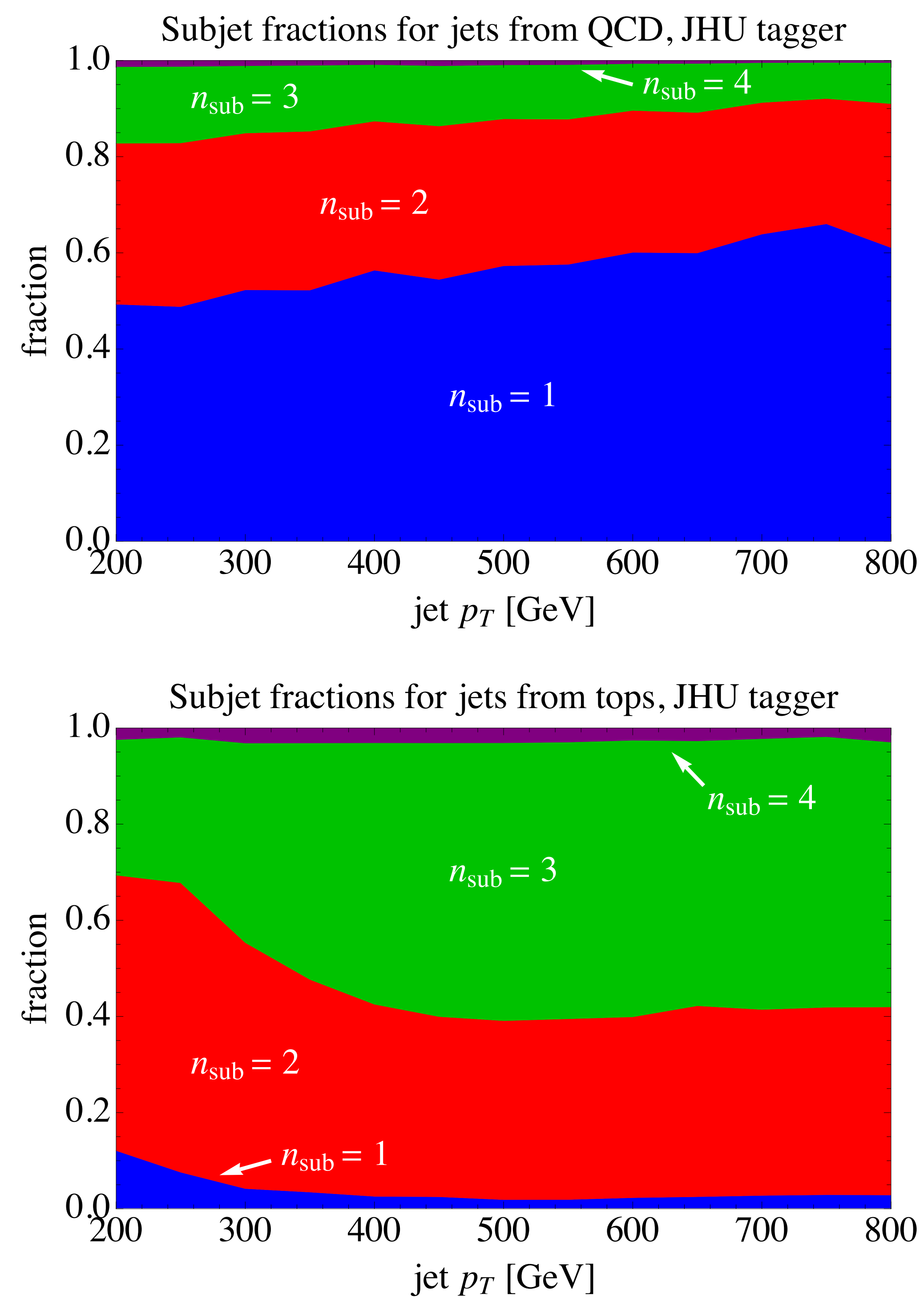
general algorithm flow:

1. groom substructure
2. count subjets
3. make kinematic cuts on subjets (e.g., mass)

N-subjettiness (for top tagging)

$$\tau_3/\tau_2 \sim \begin{cases} \mathcal{O}(1) \quad \text{for} \;\; < 3 \text{ subjets} & [\tau_3 \ll 1\,, \tau_2 \ll 1] \\ \ll 1 \quad \text{for 3 subjets} & [\tau_3 \ll 1\,, \tau_2 \sim 1] \\ \mathcal{O}(1) \quad \text{for} \;\; > 3 \text{ subjets} & [\tau_3 \sim 1\,, \tau_2 \sim 1] \end{cases}$$

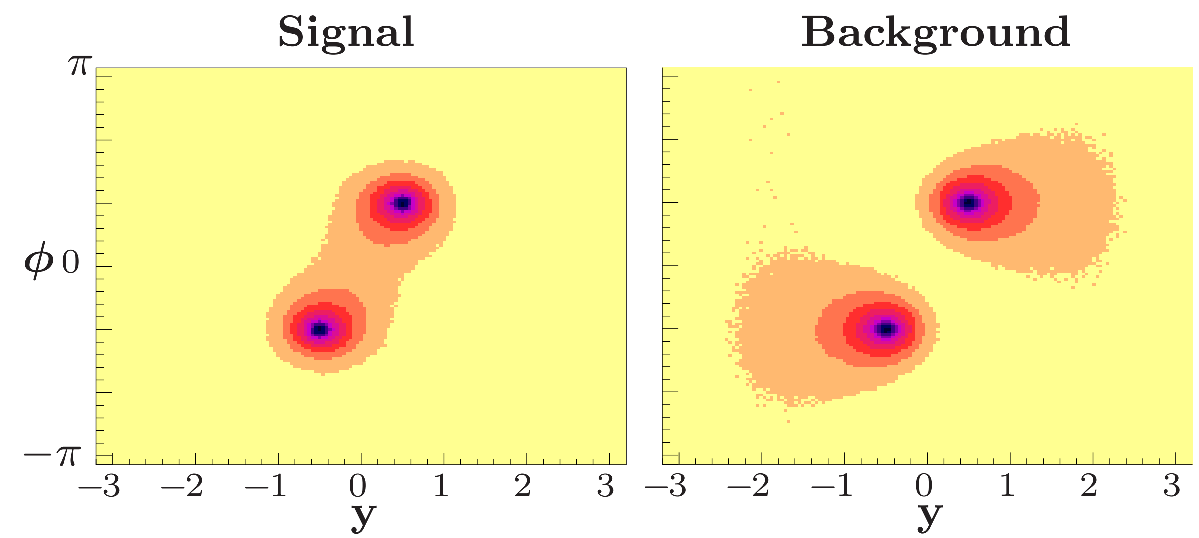
an effective counter for the number of subjets

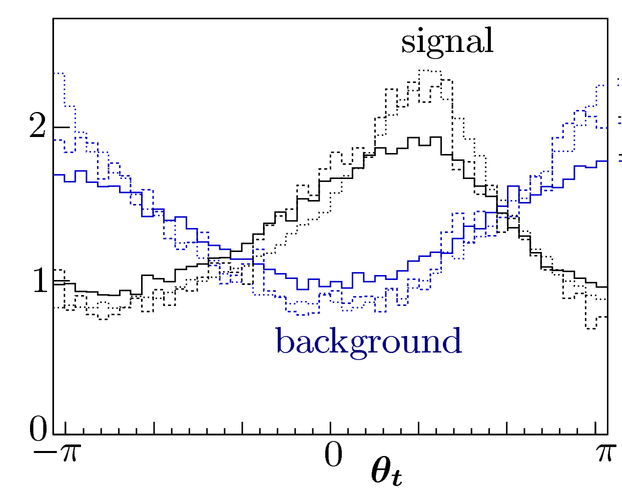


Color Flow

color flow variables cut on the pattern of soft radiation in the jet

- can try to distinguish color singlet/vector/octet jets
- also useful for spin information
- challenging due to high variance, variables never give strong separation
- can better theory understanding of what subjet axes to choose help?
- how do different algorithms overlap?

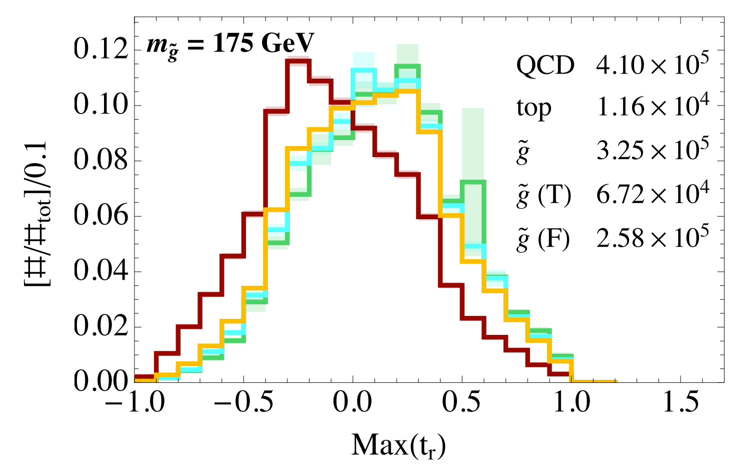




1001.5027

$$\vec{t} = \sum_{i\in\text{jet}} \frac{p_T^i\,|r_i|}{p_T^{\text{jet}}}\,\vec{r}_i$$

pull (angle)



1210.5523

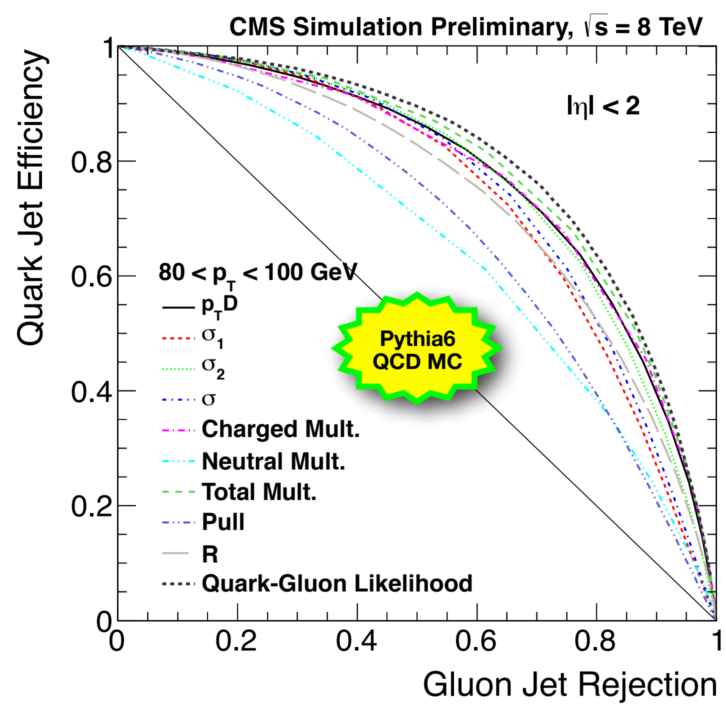
$$t_{\text{r}} = \frac{1}{N}\sum_{j=1}^{N} \hat{t}_j\cdot\hat{n}_j$$

radial pull

Reducing the Space of Observables

space of observables is unlimited, and the number of proposed observables is trending that way

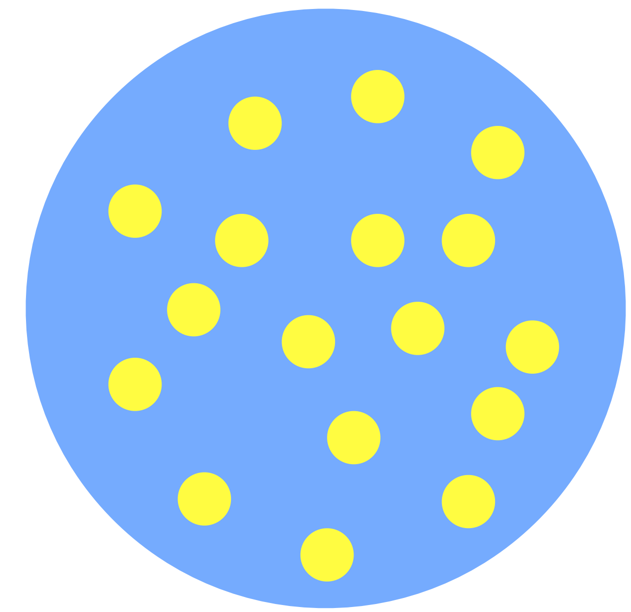
but in MVA studies, typically only 2 or 3 variables have an impact on ROC

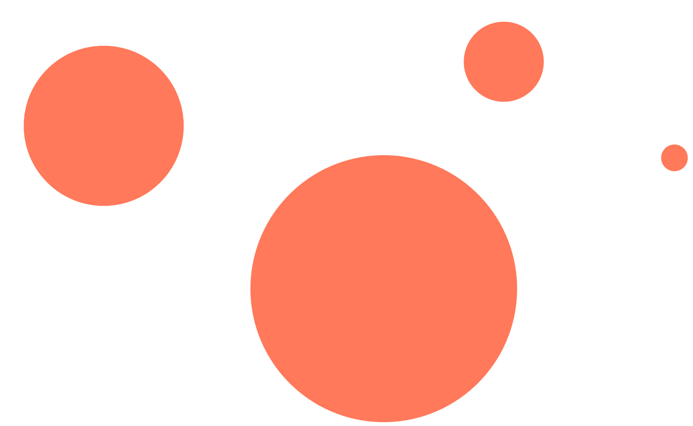


- huge projection from theory space
- validating new observables requires lots of work for experimentalists

when new observables are proposed, will they have any additional impact? which observables are most orthogonal, and why?
[crucial in design/implementation efficiency]
needs good theory ideas; active area of research

theory space
for observables





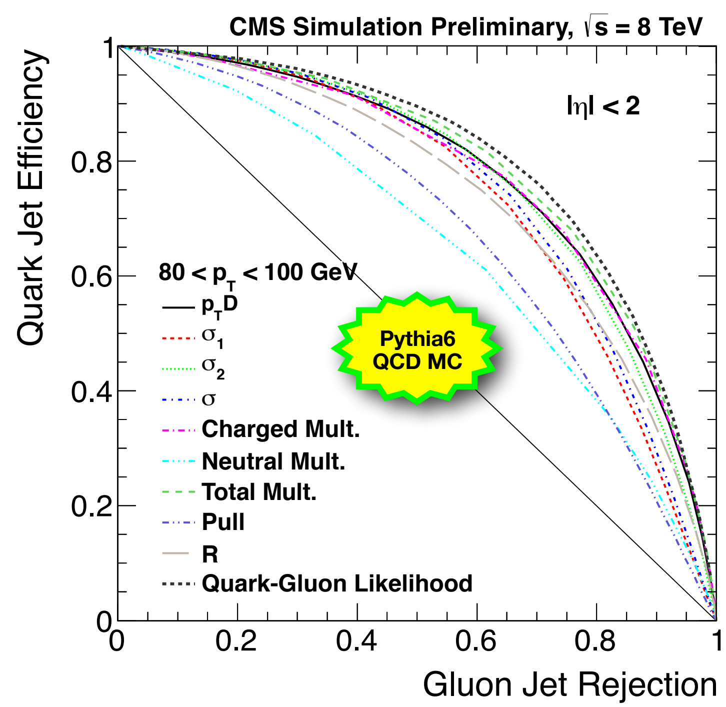
useful discriminants

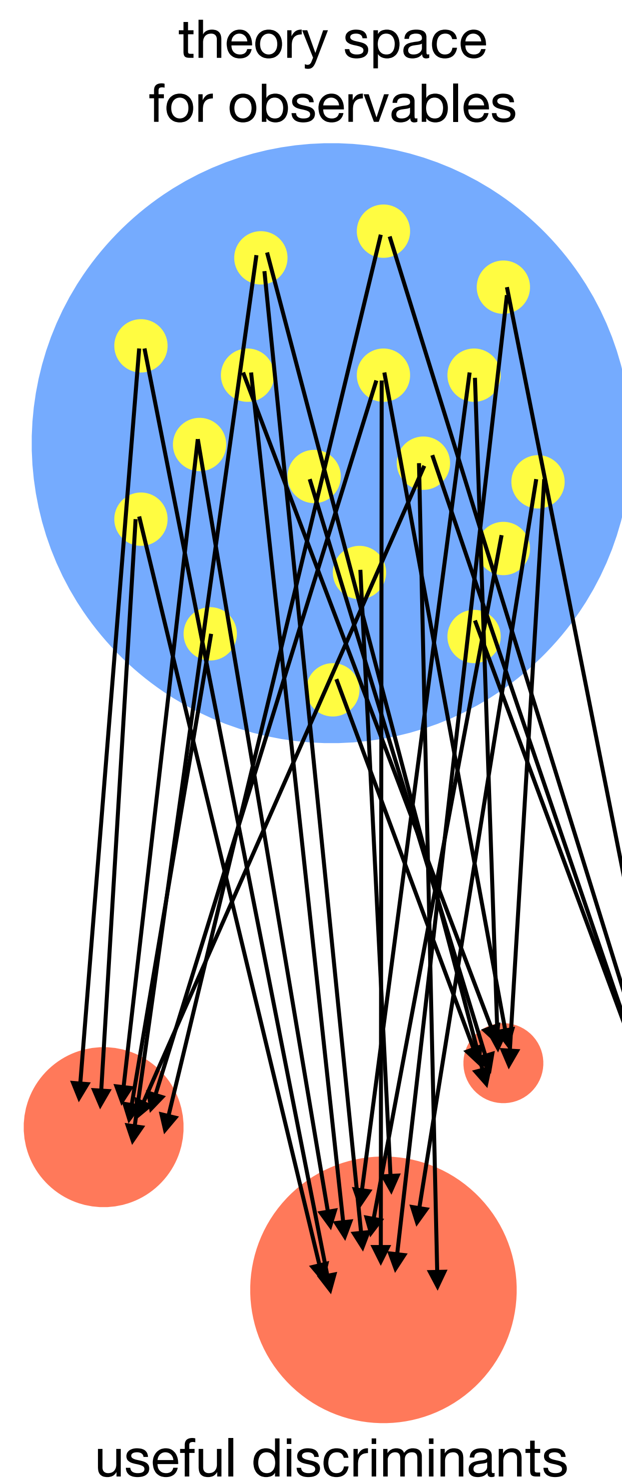
Reducing the Space of Observables

space of observables is unlimited, and the number of proposed observables is trending that way

but in MVA studies, typically only 2 or 3 variables have an impact on ROC

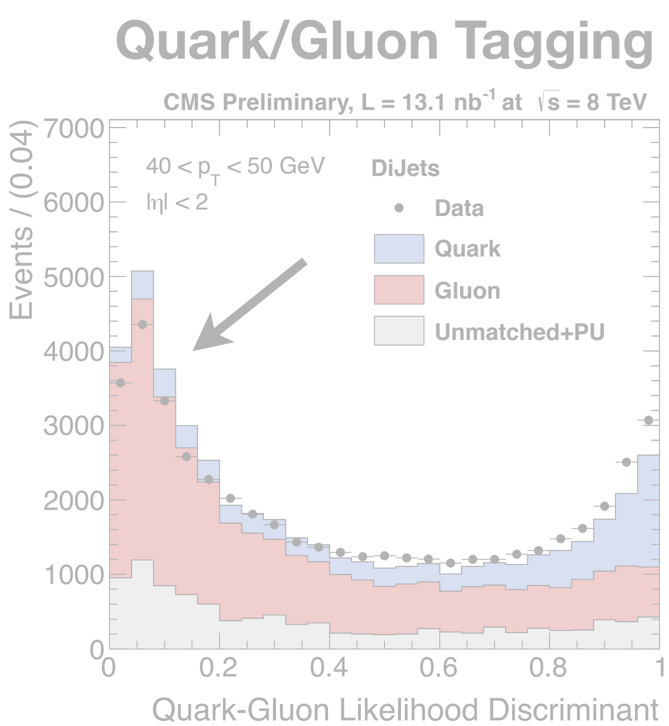
- huge projection from theory space
- validating new observables requires lots of work for experimentalists

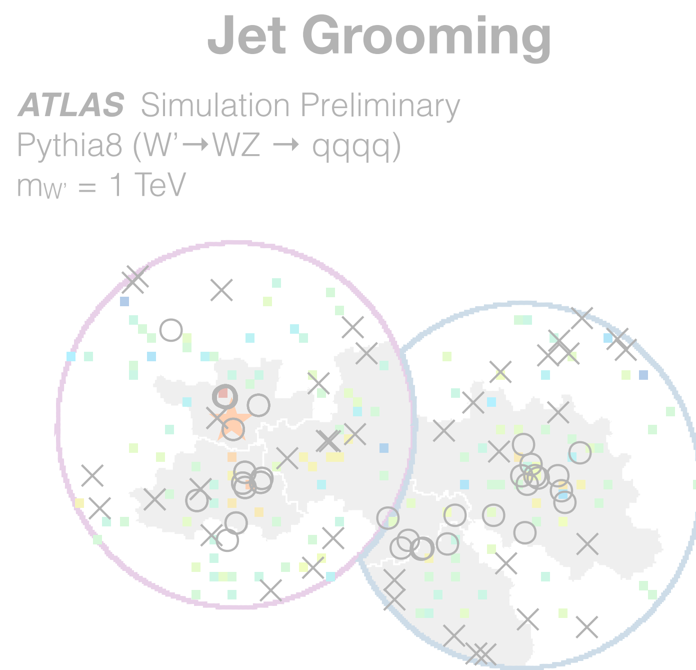


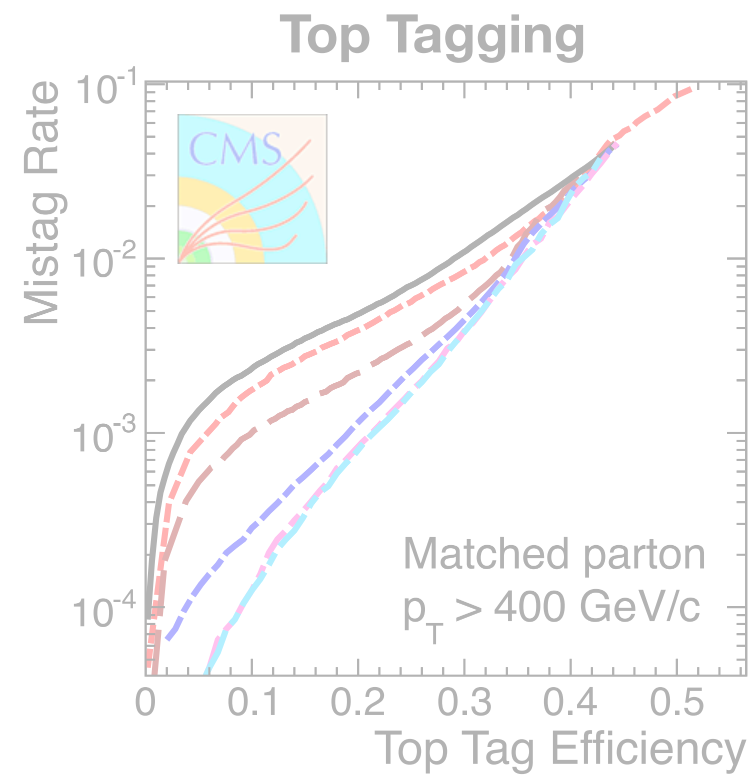


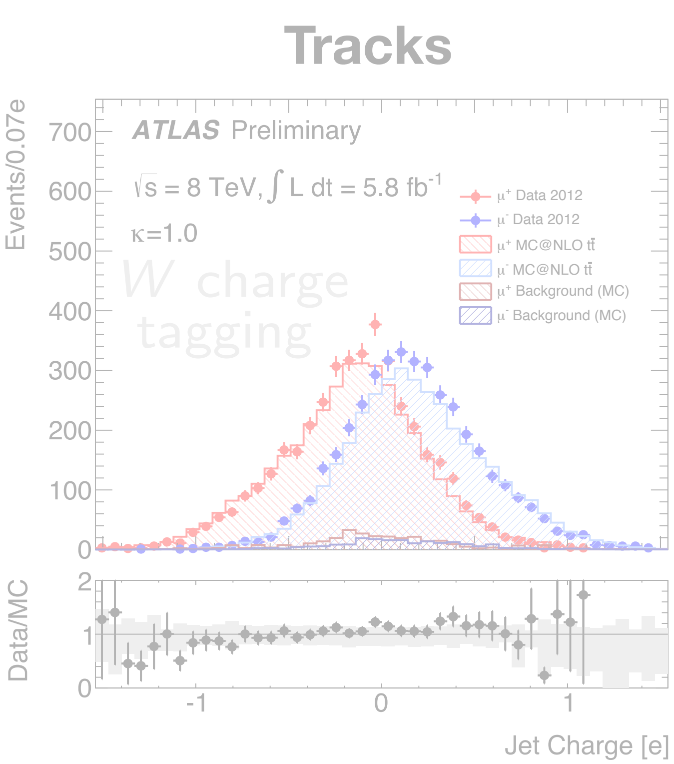
when new observables are proposed, will they have any additional impact? which observables are most orthogonal, and why?
[crucial in design/implementation efficiency]
needs good theory ideas; active area of research

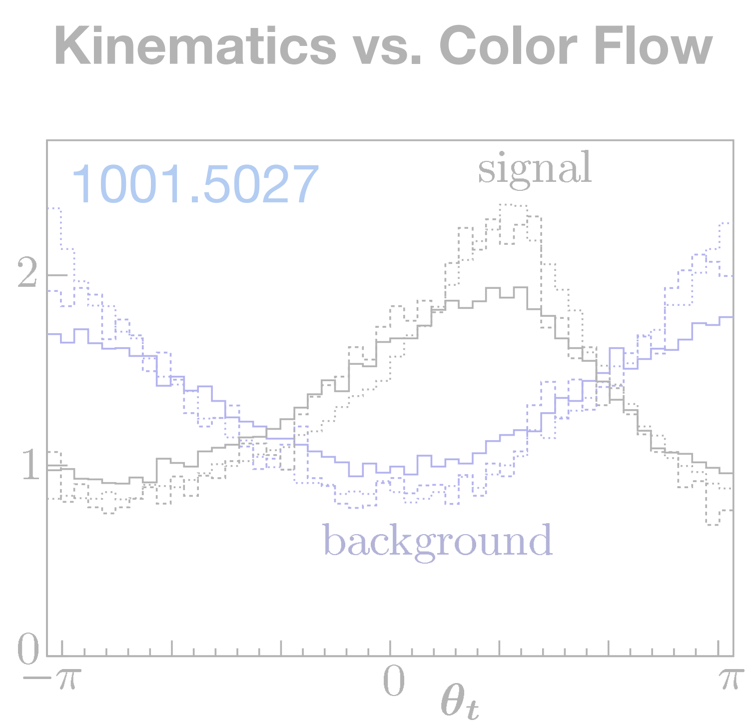
Outline

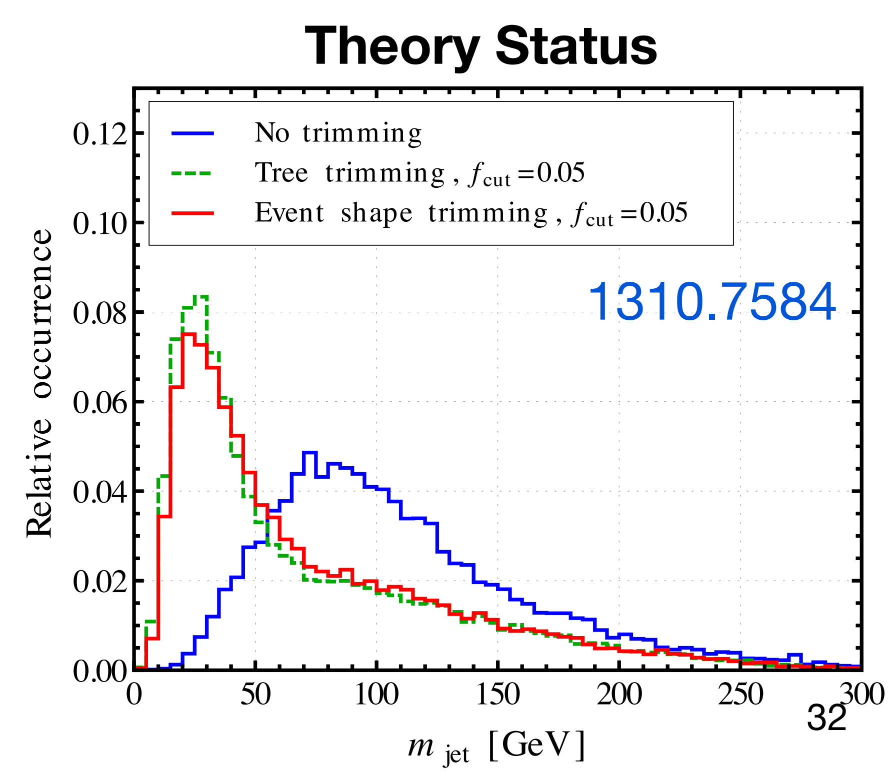

Quark/Gluon Tagging
CMS Preliminary, L = 13.1 nb-1 at √s = 8 TeV
40 < pT < 50 GeV
|η| < 2
DiJets
Data
Quark
Gluon
Unmatched+PU
Events / (0.04)
Quark-Gluon Likelihood Discriminant


Jet Grooming
ATLAS Simulation Preliminary
Pythia8 (W'→WZ → qqqq)
mW' = 1 TeV


Top Tagging
Mistag Rate
CMS
Matched parton
pT > 400 GeV/c
Top Tag Efficiency

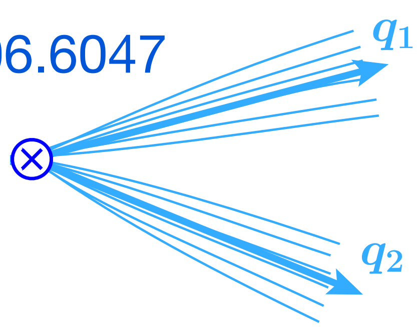

Tracks
ATLAS Preliminary
√s = 8 TeV, ∫ L dt = 5.8 fb-1
κ=1.0
W charge tagging
Events/0.07e
Data/MC
Jet Charge [e]


Kinematics vs. Color Flow
1001.5027
signal
background
θt


Theory Status
No trimming
Tree trimming, fcut = 0.05
Event shape trimming, fcut = 0.05
1310.7584
Relative occurrence
m jet [GeV]

Theory Status

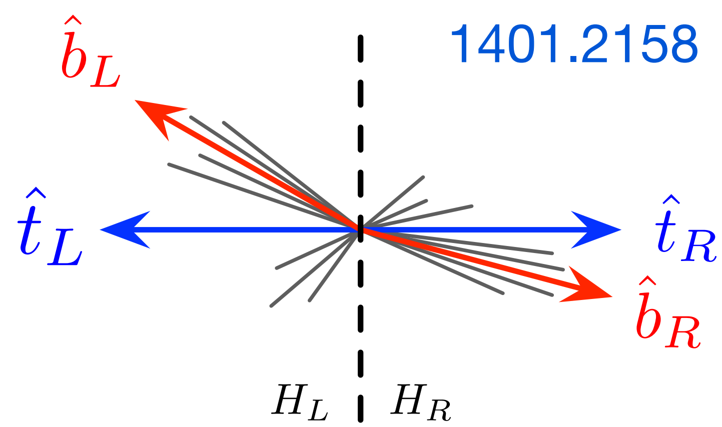
1106.6047



can we calculate at least the leading properties of shape observables, e.g. N-subjettiness?

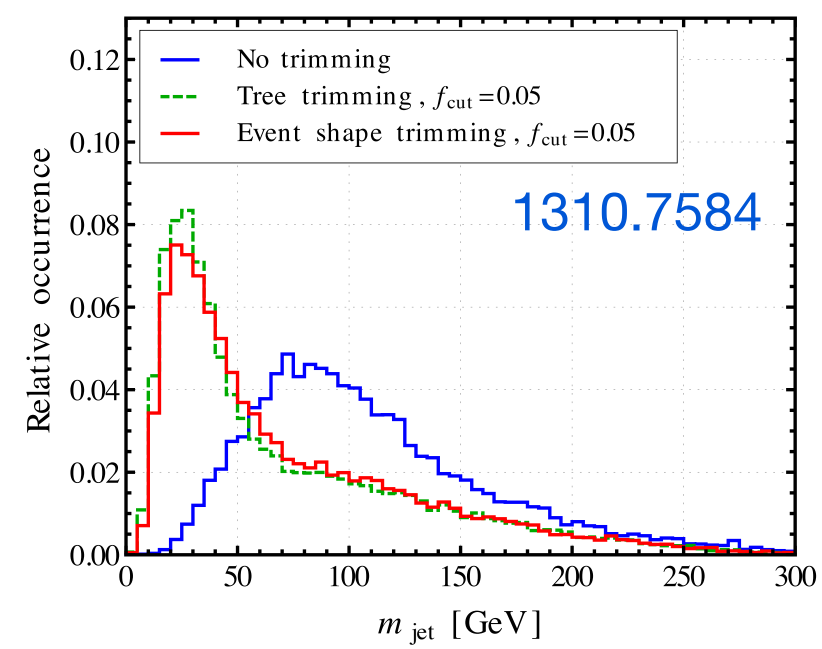
here we know the physics, but calculations are hard

1401.2158



are there choices for subjet axes that will give more calculable/robust observables?

here we know the physics, but meaningful conclusions for phenomenology have not been made



can we understand jet grooming from a more analytic and less algorithmic point of view? same for jet algorithms?

here the physics is being clarified (major steps taken)

Theory Status

observables

many interesting observables defined

when do/don't they overlap?

calculations

lessons to be learned from basic calculations

what can we calculate, and how to we inform observable design?

MC generators

experiments are good at identifying when MC is not reliable

how do we solve these problems? improved MC accuracy? better approaches (e.g. q/g)?

lots of impressive observables, calculations
(great heavy ion results not discussed here)

Conclusions

Trends:

- Observables being refined with improved theoretical understanding
- Experiments maximizing potential of tools; great idea-to-validation pipeline

Needs:

- Better understanding of why different techniques do/don't overlap
  - Will reduce the theory space of substructure, clarify understanding, help develop high-caliber methods
- More theory studies of substructure variables/methods
  - Simple results can lead to improved tools
  - Quark/gluon tagging is a prime candidate for progress