Dipolarity: Top-Tagging with Color Flow

Martin Jankowiak

with Anson Hook and Jay Wacker arXiv:hep-ph/1102.1012

SLAC March 9th, 2011

Outline

- Jet substructure
- Sequential jet clustering algorithms
- Some jet substructure techniques
- The HEPTopTagger
- Color flow and pull
- Dipolarity
- Results
- Summary & Outlook

• the excellent resolution of the ATLAS & CMS detectors means that we can "peer inside" jets and measure how energy is distributed within jets

• the excellent resolution of the ATLAS & CMS detectors means that we can "peer inside" jets and measure how energy is distributed within jets

What is this good for?

• the excellent resolution of the ATLAS & CMS detectors means that we can "peer inside" jets and measure how energy is distributed within jets

> What is this good for? • as a probe of QCD

• the excellent resolution of the ATLAS & CMS detectors means that we can "peer inside" jets and measure how energy is distributed within jets

What is this good for?

- as a probe of QCD
- event discrimination

Jet substructure as a probe of QCD

• make jet substructure measurements in real data and compare to perturbative QCD calculations

Jet substructure as a probe of QCD

- make jet substructure measurements in real data and compare to perturbative QCD calculations
- use to tune Monte Carlo event generators

Jet substructure for event discrimination

• the LHC inverse problem:

how do we connect what we measure (jets) to the hard scattering ?

• use the characteristic energy distribution of signal jets (e.g. top jets) to discriminate against background jets (e.g. QCD jets initiated by light partons)

• especially relevant for boosted objects

Sequential jet clustering algorithms 2.1 Iteration is a series of the series o
1.1 Iteration is a series of the series

two jets (with the associated problem of "overlapping" cones).

- one approach is to take a some the procedure. Combine four-vectors of the particle is the procedure. Combine four-vectors of particles to yield a list of jet four-vectors

a let use to place to place the cones. The cones widely used in the cones of particles to yield a list of jet • using an iterative procedure, combine four-vectors
- $\left\{\n \begin{array}{ccc}\n \text{simple properties under longitudinal boosts (the \\
 \text{matrix of the initial basis})\n \end{array}\n\right\}$ P longitudinal momentum participudinal momentum participus est. And direction has been direction has • procedure is formulated in coordinates with **rapidity y and the azimuthal angle** ϕ and ϕ cone algorithms are "iterative cones" (IC). In such a seed particle in such algorithms, a set of particle in such a set some particle in such a set of \sim

$$
y \equiv \frac{1}{2} \ln \frac{E + p_z}{E - p_z}
$$

 \mathbb{R}^2

2p

tj)

all j such that

O use the euclidean distance ΔR_{ij} in the $y-\phi$ [33] $r_{\rm max}$ are invariant under longitudinal boosts, whereas differences in pseudorapidity are invariant only in for the euclidean distance ΔR_{ij} in the $y-\varphi$ plane **•** use the euclidean distance ΔR_{ij} in the $y-\phi$ plane where \mathbf{w} are respectively the rapidity and azimuth of particle i. The direction of the direction

ti , (10b) , (
The contract of the contract o

$$
\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2
$$

proposal. The Sterman-Weinberg \$ parameter is less-directly mirrored in hadron-collider

cone algorithms. Rather, most physics analyses will use a cone algorithm to obtain jets

, (10a)

² < R²

2.1.2 Overlapping cones: the progressive removal approach

and pseudorapidity \mathbb{R}^n = \mathbb{R}^n \mathbb{R}^n \mathbb{R}^n and \mathbb{R}^n \mathbb{R}^n and $\mathbb{R$

Sequential jet clustering algorithms

Cambridge-Aachen Algorithm

- 1. find the smallest of the ΔR_{ij}
- 2. combine i and j and return to step 1
- 3. continue until all $\Delta R_{ij} > R$
- 4. the remaining four-vectors define a list of jets

Boosted particles

• at the LHC many of the particles considered 'heavy' at previous colliders will be produced with transverse momenta far exceeding their rest masses (W^\pm,Z^0,t,h)

• in many Beyond the Standard Model scenarios boosted particles appear in the decay of heavy resonances $(e.g. \phi \rightarrow t\bar{t})$

Boosted higgs search

• for $p_T \geq m_H$ the decay products of the higgs will typically be close together and reconstructed as a single jet ∼
∼− m_H

- about 5% of the cross-section for VH has $p_T > 200$ GeV
- backgrounds (V+jets, VV, top pairs) fall faster with p_T than the signal
- can pay to go to the boosted regime if substructure techniques can reduce backgrounds

Jon Butterworth, Adam Davison, Mathieu Rubin, Gavin Salam arXiv/hep-ph:0802.2470

• to capture all of the higgs decay products in a single jet, we need to use "fat" jets

• to accurately reconstruct the mass of the higgs, we want to "clean up" our jet to get rid of contamination from the underlying event and pile-up

• use jet substructure techniques to identify the heavy particle neighborhood of the jet

1. Break a fat $(R=1.5)$ C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

- 1. Break a fat $(R=1.5)$ C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$
- 2. If there is a significant mass drop $m_{j_1} < \mu m_j$ then exit the loop
- 3. Otherwise redefine j_1 as j and go back to step $\mathsf I$

Here $\mu = 0.67$

1. Break a fat $(R=1.5)$ C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

- 2. If there is a significant mass drop $m_{j_1} < \mu m_j$ then exit the loop
- 3. Otherwise redefine j_1 as j and go back to step $\mathsf I$

Here $\mu=0.67$

4. Finally, recluster with $R_{\text{filt}} = \min(0.3, \frac{R_{b\bar{b}}}{2})$ and use the three hardest subjets to calculate the filtered Higgs mass $R_{b\bar{b}}$ $\frac{b_{b\bar{b}}}{2}\Big)$

 $h \circ (P \cap D \cap \emptyset)$ show undown the jet radius) until the jet radius p radius) until the jet splits in the jet spl the "BDRS" procedure in diagrams

 $J_{\rm eff}$ definition of $B_{\rm eff}$ fb β fb β

 $C_{\rm eff}$, $C_{\rm eff}$,

 K , proposed by the 1.0, year of 1.0, year of

subjets, so as to filter away UE contamination while retaining hard perturbative radiation from the Higgs decay products.

objects (particles) i and j, recombines the closest pair,

updates the set of distances and repeats the procedure

until all objects are separated by a R $>$ R $>$

The HEPTopTagger hadronic top decay seems well-suited as a discriminant

color configurations. Thus the dipolarity of the W in a

5. Finally, require that the total *p^T* of the three sub-

We introduce dipolarity cuts into the HEPTopTagger

by modifying step 4 above. For a top candidate that has

passed one of the three pairs of mass cuts we calculate

conditions is satisfied in step 4, we choose the smaller

dipolarity. We find that this procedure performs better

than calculating the dipolarity of the pair of subjets that

make cuts on the filtered mass of the reconstructed top,

*m*filt, which is not done in the original HEPTopTagger,

where the cuts have been chosen so as to avoid any ex-

plicit mass scales. We introduce cuts on *m*filt for two

reasons. The first is to improve background rejection.

The second and main reason is that we are interested in

determining whether dipolarity cuts are sufficiently or-

 $\overline{}$

In addition to introducing dipolarity cuts, we also

 \mathbb{R}^n

reconstructs *m^W* most accurately.

HEP = Heidelberg/Eugene/Paris

Tilman Plehn, Gavin Salam, Michael Spannowsky, Michihisa Takeuchi, Dirk Zerwas arXiv/hep-ph:1006.2833 arXiv/hep-ph:0910.5472 Herton Commanded Michil

- essentially a generalization of the BDRS procedure to identify the three-pronged hard substructure of a top jet rate dipolarity into the Hermann is de- $\frac{1}{2}$ signed that effectively at the state boost, with $\frac{1}{2}$ e dipolarity of the mass cuts: e.g. for a top candidate that satisfies ii) and ii') the W is identified as *j*1+*j*2. If more than one of the pairs of mass
- Why do we care about jets? • designed for intermediate boost $200 \,\text{GeV}$ $200 \text{ GeV} \lesssim p_T \lesssim 800 \text{ GeV}$

The HEPTopTagger construction of the *W±*. The HEPTopTagger algorithm is defined as follows. The

1. Using the Cambridge/Aachen algorithm cluster the event into fat $R = 1.5$ jets.

implemented by the HEPTopTagger results in accurate re-

2. Break each fat jet into hard subjets using the following mass-drop criterion. Undo the last stage of clustering to yield two subjets *j*¹ and *j*² (with $m_{j_1} > m_{j_2}$, keeping both j_1 and j_2 if $m_{j_1} < 0.8m_j$

1 The HEPTOPTAGER does not make use of \mathcal{L} the HEPTOPTAGER does not make use of \mathcal{L} and otherwise dropping j_2 . Repeat this procedure recursively, stopping when the m_{j_i} drop below 30 GeV.

2 Consider in turn all possible triplets of hard sub improvement in background regulation. Since dipolarity cuts and regulate the since dipolarity cuts are the since dipolarity cuts and the since dipolarity cuts are the since dipolarity cuts and the since dipolarity cuts are jets. First, filter each triplet with a resolution $R_{\text{filter}} = \min(0.3, \Delta R_{ij}/2)$. Next, using the five 3. Consider in turn all possible triplets of hard subhardest constituent subjets of the filtered triplet calculate the jet mass m_{filt} . Finally, choose the triplet whose m_{filt} lies closest to m_t .

 $\Rightarrow m_t$

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1 , j_2 , and j_3 ordered in descending p_T . Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

triplet whose *m*filt lies closest to *mt*.

i) 0.2
$$
\leq
$$
 arctan $m_{13} \leq 1.3$
\n*i'*) $R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$
\n*ii)* $R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$

$$
iii') \frac{m_{23}}{m_{123}} \ge 0.35
$$

Here $R_{\text{min}} = 85\% \times m_W/m_t$ and $R_{\text{max}} = 115\% \times$ m_W/m_t .

5. Finally, require that the total *p^T* of the three subjets defined in step 4 be greater than 200 GeV.

0.04

dipolarity

0.02

0

250 300 350 400 450 500 550 600 650 700 750

FIG. 4: Dipolarity distributions for *W±*s reconstructed by

the HEPTopTagger and passing default mass cuts with *m*filt ∈

[150 GeV, 210 GeV]. Thick solid lines indicate central values,

whereas thin dashed lines correspond to values at 10% and

90%. Here and throughout the *p^T* is that of the fat *R* = 1*.*5

jet. For all *p^T* the central value of the dipolarity for the back-

ground is *O*(50% −100%) larger than for the signal. This fig-

ure uses the HERWIG event samples; the PYTHIA event samples

yield similar distributions. The dot-dash line roughly indi-

cates where dipolarity cuts are made at the S=20% working

Note that the *jⁱ* selected in step 4 contain only the hard

substructure of the fat jet. Some amount of soft radiation

has been thrown out by filtering and the mass drop crite-

"

"

i)
$$
0.2 \le \arctan m_{13} \le 1.3
$$

\n*i'*) $R_{\min} \le \frac{m_{23}}{m_{123}} \le R_{\max}$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \le 1 - \frac{m_{23}^2}{m_{123}^2} \le R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$

*m*²

Here *R*min = 85% × *m^W /m^t* and *R*max = 115% ×

12

"

[≤] ¹ [−] *^m*²

23

[≤] *^R*²

 $\overline{}$

1 +

*m*²

*m*²

*m*²

12

"

the HEPTopTagger and passing default mass cuts with *m*filt ∈

[150 GeV, 210 GeV]. Thick solid lines indicate central values,

whereas thin dashed lines correspond to values at 10% and

90%. Here and throughout the *p^T* is that of the fat *R* = 1*.*5

jet. For all *p^T* the central value of the dipolarity for the back-

ground is *O*(50% −100%) larger than for the signal. This fig-

ure uses the HERWIG event samples; the PYTHIA event samples

yield similar distributions. The dot-dash line roughly indi-

cates where dipolarity cuts are made at the S=20% working

13

12

$$
iii') \frac{m_{23}}{m_{123}} \ge 0.35
$$

iii) *R*²

 $\overline{}$

1 +

mass closest to *mt*.

i)
$$
0.2 \le \arctan m_{13} \le 1.3
$$

\n*i'*) $R_{\min} \le \frac{m_{23}}{m_{123}} \le R_{\max}$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \le 1 - \frac{m_{23}^2}{m_{123}^2} \le R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$
\n*iii'*) $\frac{m_{23}}{m_{123}} \ge 0.35$

*m*²

*m*²

*m*²

12

"

the HEPTopTagger and passing default mass cuts with *m*filt ∈

[150 GeV, 210 GeV]. Thick solid lines indicate central values,

whereas thin dashed lines correspond to values at 10% and

90%. Here and throughout the *p^T* is that of the fat *R* = 1*.*5

jet. For all *p^T* the central value of the dipolarity for the back-

ground is *O*(50% −100%) larger than for the signal. This fig-

ure uses the HERWIG event samples; the PYTHIA event samples

yield similar distributions. The dot-dash line roughly indi-

cates where dipolarity cuts are made at the S=20% working

13

12

than five filtered constituents use all of them). Finally, select the set of three-subjet pairings with a jet

*m*²

Here *R*min = 85% × *m^W /m^t* and *R*max = 115% ×

12

"

[≤] ¹ [−] *^m*²

23

[≤] *^R*²

 $\overline{}$

1 +

1 +

*m*¹²³

iii) *R*²

 $\overline{}$

mass closest to *mt*.

The HEPTopTagger

Legoplot for a top jet with hard substructure as identified by the HEPTopTagger

• a top jet has more structure than is encoded by kinematic constraints: $(p_1+p_2+p_3)^2=m_t^2$ $(p_1 + p_2)^2 = m_W^2$

• the effective operator for the decay has a particular color configuration $t \to b q \bar{q}$

• in particular the W boson is a color singlet and the color indices of q and \bar{q} are contracted

• a top jet has more structure than is encoded by kinematic constraints: $(p_1+p_2+p_3)^2=m_t^2$ $(p_1 + p_2)^2 = m_W^2$

• the effective operator for the decay has a particular color configuration $t \to b q \bar{q}$

• in particular the W boson is a color singlet and the color indices of q and \bar{q} are contracted

Question: can we use color information to improve background rejection in top tagging algorithms?

- in a QCD event radiation is controlled by i) the kinematics of the hard partons and by ii) how color indices are contracted together (color flow)
- how does a color singlet radiate?
- apart from some color algebra, $QED \sim QCD$; so let's first ask this question in the context of QED

The Chudakov Effect: \blacksquare Consider emission of soft photon in pair \blacksquare

Soft bremsstrahlung from e^+e^- pairs is suppressed Consider emission of soft photon at angle θ from electron in pair with opening from e^+e^- pairs is suppressed. emsstrahlung from e^+e^- pairs is suppre

Heuristic explanation: Transverse momentum of photon is ktion is ktion is ktion is ktion is ktion is ktion in balance at e Transverse momentum of photon is ktic and energy in photon is ktic at even at even at each \sim suc e

 \mathbb{R}

opening angle of pair, which is angular ordering.

- there is an energy imbalance at the vertex $\Delta E \sim k_T^2 / zp \sim zp\theta^2$ at the vertex $\Delta E \sim$
- time available for emission is $\Delta t \sim 1/\Delta E$
- in which time the pair separates T imbolon is Δt \sim 1/ ΔL
Dair separates Δb . Δt pair separates $\Delta\theta \sim$ $\Delta b \sim \theta_{e^+e^-} \Delta t.$
- for emission photon must resolve this distance: For non-negligible probability of the probability of the state of the state of the state this transverse this transverse this transverse that

Introduction to QCD at CollidersLecture III: Shower Monte Carlo – p.23/32 $\Delta b > \lambda/\theta \sim (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-} (zp\theta^2)^{-1} > (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-} > \theta$

Introduction to QCD at CollidersLecture III: Shower Monte Carlo – p.23/32

The Chudakov Effect: \blacksquare Consider emission of soft photon in pair \blacksquare

Soft bremsstrahlung from e^+e^- pairs is suppressed Consider emission of soft photon at angle θ from electron in pair with opening from e^+e^- pairs is suppressed. emsstrahlung from e^+e^- pairs is suppre

Heuristic explanation: Transverse momentum of photon is ktion is ktion is ktion is ktion is ktion in balance at each energy in balance at e Transverse momentum of photon is ktic and energy in photon is ktic at even at even at each \sim suc e

 \mathbb{R}

opening angle of pair, which is angular ordering.

- there is an energy imbalance at the vertex $\Delta E \sim k_T^2 / zp \sim zp\theta^2$ at the vertex $\Delta E \sim$
- time available for emission is $\Delta t \sim 1/\Delta E$
- in which time the pair separates T imbolon is Δt \sim 1/ ΔL
Dair separates Δb . Δt pair separates $\Delta\theta \sim$ $\Delta b \sim \theta_{e^+e^-} \Delta t.$ For non-negligible probability of the probability of the state of the state of the state this transverse this transverse this transverse that
	- for emission photon must resolve this distance:

$$
\Delta b > \lambda/\theta \sim (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-} (zp\theta^2)^{-1} > (zp\theta)^{-1} \Rightarrow \boxed{\theta_{e^+e^-} > \theta}
$$

angular ordering

Introduction to QCD at CollidersLecture III: Shower Monte Carlo – p.23/32

$$
\bigwedge \bigwedge \bigwedge \bigwedge \bigwedge \bigwedge \theta_{e^+e^-}
$$

- this so-called angular ordering property of soft emission is common to all gauge theories
- soft emissions that are not angular-ordered are suppressed by deconstructive interference azimuthal integration, we find \mathcal{L}
- the radiation from a pair of partons i and j in a color singlet configuration is mostly limited to two cones centered around i and j $\frac{1}{2}$ $\overline{ }$ is m $\frac{1}{2}$, compared the contribution of the contribution of the contribution of the contribution of the contribution of
Section of the contribution of
 ed to two cones

Color flow and pull jet in polar coordinates. The signal (connected to ¯b jet) is on $\sqrt{ }$ alors the right. the same, but we have found that the most effective computation of \mathbf{h} bination is a pull, which we call the call pull, which we can consider the call \mathbb{R}^n

| • this observation led to the introduction of the jet | observable "pull" $T = 0$ this one \mathbf{u} from approximations used in participations used in participations \mathbf{u} ers [7, 8]. In the color dipoles are allarge energy scales as social the scales associated with the hard interaction and the hard interaction and inter to the lower energy scale associated with confinement. pythia 8 [15] for the parton shower, hadronization and rvation lad to the tion. To begin, we isolate the effect of the color connec- $\mathbf{t}_{\mathbf{r}}$ fixing the parton momentum. We compare events with \mathbf{p} in the final state (with \mathbf{p} \mathbf{p} \mathbf{p} \mathbf{p} which the quarks are color-connected to each other (sig-connected to each other (sig-connected to each other (sig-connected to each other (sig-connected) α $\overline{\text{ク}}$

θt

θt

 \mathcal{F}_1 and the pull vector of the by-

y.

represent factors of 2 increase in radiation.

FIG. 2: Accumulated p^T after showering a particular partonic phase space point 3 million times. Left has the ^b and ¯^b color-connected to each other (signal) and right has the ^b and α color-connected to the beams (background). Contours (background). Contours (background). Contours (co

In order to extract the color connections, they must

These emissions transpire in the rest frame of the dipole. When boosting back to the lab frame, the lab frame, the radiation appears dominantly within an angular region spanned by the dipole, as indicated by the arrows in Figure 1. Alternatively, an angular ordering can be enforced on the radiation (as in here p). The parton shower treatment of parton shower treatment of parton shower treatment of parton radiation attempts to include a number of features which are physical but hard to calculate analytically, such as overall momentum and probability conservation or coherence phenomena associated with soft radiation.

^y [−]³ [−]2 3 [−]1 2 0 1

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

and to define the jet. The fine the jet the jet the jet are all basically define the jet. These are all basica

Signal Pull Background Pull

−π π

• unfortunately, pull does not seem well suited to topdence for tagging \blacksquare are into the simulation. In the simulation \blacksquare \blacksquare comercies that \blacksquare ous experiments. In e+e[−] collisions, for example, evigluon jets was observed in three jet events by JADE at Designation is later to be a set of the LS and DELPHIT α experiments found evidence for color coherence among we want to isolate is that the radiation is that the radiation is that the radiation is that the radiation is the radiation in each signal jet taly null does no rtunately, puil does not seem wei other words, the radiation on each end of a color dipole \overline{e} r suited to \bigcirc top- is a particular proposed by \vert , \vert ci is the position of a cell or

Color flow and pull jet in polar coordinates. The signal (connected to ¯b jet) is on $\sqrt{ }$ alors the right.

θt

- These emissions transpire in the rest frame of the dipole. \blacksquare pears dominantly within an angular region spanned by $t_{\rm max}$ indicated by the arrows in Figure natively, an angular ordering can be enforced on the radi- \blacksquare hosons in \blacksquare radiation and include a number of the set of are physical but hard to calculate analytically, such as Incorporate 2, we show that **THERE CAS TOOKED** • D0 experiment has looked at the pull of hadronic W each b, where ϵ is the rapidity. For this figure, we have the rapidity. For this figure, we have the rapidity. \sim avents the same parton \sim bosons in $t\bar{t}$ events
- It is more important that the that the these effects exist in data that the these effects exist in data that the that they are included in the simulation. In the simulation \mathbf{r}_i color coherence effects have a seen by varied and see the seen by various experiments. In example, for example, \blacksquare dence for color connections between final-state quark and tends to shower in the direction of the other jet, while in the background it shows it shows it shows a background it shows a background it shows a background it shows a • Results in good agreement with Monte Carlo

θt

 \mathcal{F}_1 and the pull vector of the by-

y.

represent factors of 2 increase in radiation.

FIG. 2: Accumulated p^T after showering a particular partonic phase space point 3 million times. Left has the ^b and ¯^b color-connected to each other (signal) and right has the ^b and α color-connected to the beams (background). Contours (background). Contours (background). Contours (co

In order to extract the color connections, they must persist into the distribution of the observable hadrons. The basic interest intuition for how the color flow might show might show might up follows from approximations used in parton showers [7, 8]. In the color dipoles are allowed to radiate through Markovian evolution from the large energy scales associated with the hard interaction to the lower energy scale associated with confinement.

overall momentum and probability conservation or coherence phenomena associated with soft radiation.

gluon jets was observed in the jets at DESY [10]. Later, at LEP, the L3 and DELPHI \sim

discriminant which can help determine the color flow of practically any event. Such a tool has the potential for wide applicability in the LHC. The LHC searches at the LHC searches at the LHC. For an example, we will use Higgs production in association with a Z. The Z. T pT so that its body products are not back-to-back-to-back-to-back-to-back-to-back-

^y [−]³ [−]2 3 [−]1 2 0 1

experiments found evidence for color coherence among $\mathbf{f}_{\text{total}}$ <u>Lesso events in provins in p</u> tron, color co been observed by D0 in W+jet events [13]. All of these in W+jet events [13]. All of these in W+jet events [13] $\mathbf{r} = \mathbf{r} \mathbf{v}$ dent on the particular event topology. What we will now $\mathbf{I} = \mathbf{U} \cdot \mathbf{U} \cdot \mathbf{U}$ show is that it is possible to come up with a very general \boldsymbol{d} f_{true} and g_{true} d $\Delta\Lambda$ Λ which the moment vy $\,$ D • fraction of uncolored W bosons measured to be the jet, has a number of immediate advantages. For α it will be a more general-purpose to events to events to event set $\mathcal{O}(n)$ \perp \cup \pm \angle $f = 0.56 \pm 0.42$

Measurement of color flow in f **t events from** $p\bar{p}$ **collisions at** $\sqrt{s} = 1.96$ **TeV**

arXiv/hep-ex:1101.0648v1

pT i

 R_i^2

 $\left| \begin{array}{c} 2 \ i \end{array} \right|$

 p_{T_J}

• instead of something like pull, consider the entire radiation pattern of the W simultaneously or something like pull, consider the entire
nattern of the *W* simultaneously

- · dipolarity is essentialty order subjet observable
- · expectation: drop viets woll jets'd small values of \overline{QCD} jets will yield larger values of D
• Good observables can be calculated in perturbation theory way, dipolanity in the Gafety and well st against soft and collinear splittings whereas QCD jets will yield larger values of • provided the two subjets are chosen in an IRC safe *D D*

• by incorporating dipolarity into the HEPToptagger, we can try to beat down QCD backgrounds

• even if a QCD fakes the kinematics of the top well, it will typically have a different color configuration $F(0)$ is a contract of $F(0)$ is the surface of the $F(0)$ support $F(0)$ water topically have a directore color configuration

bilities for quarks and gluons and gluons that undergo two branchings. The undergo two branchings. The undergo

• incorporate dipolarity into the HEPTopTagger by modifying step 4

Dipolarity hardest constituent subject constituent subject of the filtered triplet of the filtered triplet of the filtered
The filtered triplet of the filtered triplet of the filtered triplet of the filtered triplet of the filtered t calculate the juncolarity. Finally, choose the juncolarity, choose the juncolarity, choose the internal operation triplet whose *m*filt lies closest to *mt*.

*R*filter = min(0*.*3*,* ∆*Rij/*2). Next, using the five

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1 , j_2 , and j_3 ordered in descending p_T . Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

i) 0.2
$$
\leq
$$
 arctan $m_{13} \leq 1.3$
\n*i'*) $R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$
\n*ii)* $R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$
\n*ii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$
\n*iii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$

Here $R_{\text{min}} = 85\% \times m_W/m_t$ and $R_{\text{max}} = 115\% \times$ m_W/m_t .

FIG. 4: Dipolarity distributions for *W±*s reconstructed by

the HEPTopTagger and passing default mass cuts with *m*filt ∈

[150 GeV, 210 GeV]. Thick solid lines indicate central values,

whereas thin dashed lines correspond to values at 10% and

90%. Here and throughout the *p^T* is that of the fat *R* = 1*.*5

jet. For all *p^T* the central value of the dipolarity for the back-

ground is *O*(50% −100%) larger than for the signal. This fig-

ure uses the HERWIG event samples; the PYTHIA event samples

yield similar distributions. The dot-dash line roughly indi-

cates where dipolarity cuts are made at the S=20% working

Note that the *jⁱ* selected in step 4 contain only the hard

substructure of the fat jet. Some amount of soft radiation

has been thrown out by filtering and the mass drop crite-

Dipolarity hardest constituent subject constituent subject of the filtered triplet of the filtered triplet of the filtered
The filtered triplet of the filtered triplet of the filtered triplet of the filtered triplet of the filtered t calculate the juncolarity. Finally, choose the juncolarity, choose the juncolarity, choose the internal operation triplet whose *m*filt lies closest to *mt*.

*R*filter = min(0*.*3*,* ∆*Rij/*2). Next, using the five

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1 , j_2 , and j_3 ordered in descending p_T . Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

i) 0.2
$$
\leq
$$
 arctan $m_{13} \leq 1.3$
\n*i'*) $R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$
\n*ii)* $R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$
\n*ii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$
\n*iii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$

Here $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times$ m_W/m_t .

250 300 350 400 450 500 550 600 650 700 750 0 0.02 0.04 0.06 dipolarity FIG. 4: Dipolarity distributions for *W±*s reconstructed by the HEPTopTagger and passing default mass cuts with *m*filt ∈ $j_{W} = j_{2} + j_{3}$

Note that the *jⁱ* selected in step 4 contain only the hard

substructure of the fat jet. Some amount of soft radiation

has been thrown out by filtering and the mass drop crite-

Dipolarity hardest constituent subject constituent subject of the filtered triplet of the filtered triplet of the filtered
The filtered triplet of the filtered triplet of the filtered triplet of the filtered triplet of the filtered t calculate the juncolarity. Finally, choose the juncolarity, choose the juncolarity, choose the internal operation triplet whose *m*filt lies closest to *mt*.

*R*filter = min(0*.*3*,* ∆*Rij/*2). Next, using the five

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1 , j_2 , and j_3 ordered in descending p_T . Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

i) 0.2
$$
\leq
$$
 arctan $m_{13} \leq 1.3$
\n*i'*) $R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$
\n*ii)* $R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$
\n*iii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$
\n*iii)* $R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$
\n*iii'*) $\frac{m_{23}}{m_{123}} \geq 0.35$
\nHere $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times$

250 300 350 400 450 500 550 600 650 700 750

Note that the *jⁱ* selected in step 4 contain only the hard

substructure of the fat jet. Some amount of soft radiation

has been thrown out by filtering and the mass drop crite-

0

0.02

0.04

0.06

dipolarity

- incorporate dipolarity into the HEPTopTagger by modifying step 4
- calculate the dipolarity of the pair of subjets identified in step 4
- if more than one pair of subjets passes the mass cuts, choose the smaller dipolarity
- make a dipolarity cut $D < D_{\text{max}}$
- need to make one more choice: what radiation goes into the sum?

• we find that the criterion used to select the radiation that enters the sum has significant impact on the utility of dipolarity of as a discriminant

• angular ordering implies that most of the radiation from the W is within the pair of cones of radius ΔR

• choose our cones to be somewhat smaller, $\Delta R/\sqrt{2}$, to minimize contamination from the underlying event

• also remove any radiation in the neighborhood of the b subjet

• test the modified top-tagger on three event samples from BOOST 2010:

- 1. HERWIG 6.510 angular ordering
- 2. PYTHIA 'DW' Q^2 ordering
- 3. PYTHIA 'Perugia' p_T ordering
- jet clustering with Fastlet 2.4.2
- zeroth order detector mock-up by binning particles i nto 0.1×0.1 cells in $y-\phi$ space

Dipolarity for intermediate pT !400!⁶⁰⁰ GeV"

• we want to see whether dipolarity cuts are essentially orthogonal to the kinematic cuts imposed by the HEPTopTagger

• so include cuts on the reconstructed mass of the top so that the HEPTopTagger is using a full compliment of kinematic cuts

• optimize the cuts use Monte Carlo code to finely sample the space of cuts

• at each signal efficiency S choose cuts so that the background mistag rate B is minimized

Results and PYTHIA 'Perugia' arranged from left to right. Including dipolarity cuts loosens mass cuts while improving α

statistics. The resulting background mistag rates (B) are shown for each of the three event samples with HERWIG, PYTHIA 'DW',

from the details of the parton shower; doing so, however,

Monte Carlo event generators match what is observed

in collider experiments is only beginning to be studied

actively. Understanding color flow in detail is a diffi-

cult problem; for example, QCD predictions for radia-

tion patterns can be affected by non-global logarithms,

see *e.g.* [40]. Therefore validating theoretical predictions

against data will be critical in reducing the theoreti-

cal uncertainty associated with how dipolarity and other

color flow observables are modeled by Monte Carlo calcu-

lations. A few color coherence studies performed at the

Tevatron showed spatial correlations between the third

and second hardest jets in *pp*¯ collisions, and HERWIG was

shown to provide a better description of the data than

PYTHIA [37]. More recently, the color of the *W[±]* in *tt*

events was studied, and agreement between theory pre-

Validating how well color flow effects as modeled by

lies outside the scope of this paper.

• improves background rejection at lower S

• for intermediate to high p_T (400 GeV $< p_T < 800$ GeV) and for lower signal efficiencies including dipolarity cuts can improve background rejection

• there is sizable disagreement between the different Monte Carlo event samples

• this disagreement probably has its origin in the details of the parton showers (not e.g. the underlying event models)

• this is not surprising - theoretical understanding of color coherence (and its inclusion in MC) is limited

Summary & Outlook

• introduced a jet observable "dipolarity" that can distinguish between different color configurations in jets with significant mass drops

• incorporating dipolarity in the HEPTopTagger improves background rejection

• due to theoretical uncertainties, the ultimate utility of dipolarity awaits real data

• dipolarity should have other applications outside of top-tagging (e.g. W/Z physics, heavy Higgs)

Summary & Outlook

• theoretical understanding of color flow and other jet substructure observables should benefit from confrontation with LHC data

• e.g. CMS just published CMS PAS JME-10-013 'Study of Jet Substructure in pp Collisions at 7 TeV in CMS'

- measured mistag rate for a W tagging and top tagging algorithm
- good agreement with Monte Carlo (especially Herwig++)

backup slides

Sequential jet clustering algorithms

- 1. find the smallest of the d_{ij} and d_{iB}
- 2. if it is d_{ij} combine i and j and return to step 1
- $2.2.5$, The anti-kt algorithm $\frac{1}{2}$ 3. if it is d_{iB} declare i to be a jet and remove it from the list of four-vectors, returning to step I
- 4. continue until there are no particles left

$$
p = 0 \Rightarrow \quad \text{Cambridge-Aachen} p = 1 \Rightarrow \quad \quad \text{kT} p = -1 \Rightarrow \quad \quad \text{anti-kT}
$$

$$
d_{ij} = \frac{\min\{p_{ti}^2, p_{ti}^2\}}{R^2}
$$

$$
d_{iB} = \frac{\sum f_i}{\sum f_i} \frac{1}{R^2}
$$

One can generalise the k^t and Cambridge/Aachen distance measures as [33]:

One can generalise the k^t and Cambridge/Aachen distance measures as [33]:

where p is 1 for the kt algorithm, and 0 for the kt algorithm, and 0 for α . It was observed in $[33]$

where p is 1 for the kt algorithm, and 0 for the kt algorithm, and 0 for C/A. It was observed in $\mathcal{O}(33)$

 $\boldsymbol{t} = -1, \boldsymbol{t} = -1,$

 $t_{\rm eff}$ = $t_{\rm eff}$ and $t_{\rm eff}$ and the "anti-kt" algorithm, then this favours clustering that favours clusterings that

Some jet substructure techniques of the subset of the contribution \mathcal{L}

and α , R , R or α , R , R or R , R

1. Break a fat $(R=1.5)$ C/A jet j into subjets j_1 and j_2 by and and anti-kt and anti-kt and $m_{j_1} > m_{j_2}$ perfect b-tagging; the C/A algorithm uses the procedure outlined in the text; the k^t algorithm

2. If (i) there is a significant mass drop $m_{j_1} < \mu m_j$ and (ii) the splitting is not too asymmetric then exit the loop case R has been chosen to give near optimal significance with that algorithm. $\sum_{i=1}^n$ and $\sum_{i=1}^n$ are defined in the splitting is not too. The spin of the splitting is not to $\sum_{i=1}^n$

$$
y = \frac{\min(p_{ij_1}^2, p_{ij_2}^2)}{m_j^2} \Delta R_{j_1, j_2}^2 > y_{\text{cut}} \qquad y \approx \frac{\min(p_{Tj_1}, p_{Tj_2})}{\max(p_{Tj_1}, p_{Tj_2})}
$$

3. Otherwise redefine j_1 as j and go back to step 1 neighbourhood and exit the loop (µ was taken to be 0.67 and ycut = 0.09). Note that **S.** Otherwise redefine j_1 j_1 as j

Here $\mu = 0.67$ and $y_{\text{cut}} = 0.09$

114. Finally, recluster with $R_{\text{filt}} = \min(0.3, \frac{R_{b\bar{b}}}{2})$ and use the 1 \Box three hardest subjets to calculate the filtered Higgs mass \Box $R_{b\bar{b}}$ $\frac{b_{b\bar{b}}}{2}\Big)$

verify that the two subjects at two subjects at the two subjects at that stage both have a b-tag. The cut on y
The cut on y " z/(1 − z) z/(2 − z) z/(1 − z) z/(2 − z) z/(2

allows one to kill the logarithm for (face backgrounds in eq. (56). By virtue $\mathcal{L}(\mathcal{L})$