Seeing Color Flow in $t\bar{t}$

*UC HEP Seminar*

Maximilian Swiatlowski$^1$

$^1$University of Chicago, Enrico Fermi Institute

19 October, 2015
The Standard Model

- SM is the enormously successful theory of the world around us
- Matter particles interact via force particles
- Forces interact with particles *charged* under that force
QCD, described by $SU(3)_C$, has rich structure

- Only some particles are charged under it—refer to them as colored
- Quarks referred to as triplets: single color
- Gluons referred to as octets: two colors
- All others referred to as singlets: no color

Several ways to make color-neutral combinations (hadrons):
- Because of asymptotic freedom, never observe un-confined color
- Knocking loose a colored particle at a hadron collider leads to jets
- Given these issues, how can we study the effects of color?
Because of asymptotic freedom, never observe un-confined color
Knocking loose a colored particle at a hadron collider leads to jets
Given these issues, how can we study the effects of color?
Gluon jets discovered at JADE/TASSO/PLUTO and others
- Following Ellis, Gaillaird, Ross
- Gluons have higher color charge: hadronize into greater number of particles
  - Also observed at LEP, LHC, and elsewhere
- Certainly one effect of color—but says nothing about the color lines
LEP experiments studied many aspects of color as well: one example, **color reconnection**

- Do the jets from $W$-decays hadronize independently?
- Or do they *reconnect* during this process?

Going from observation of color effects, to **measurements of color properties**

- Study particles *between* pairs of jets: re-connection “color strings” would produce particles there
- Most extreme radiation models rule out
LEP experiments studied many aspects of color as well: one example, color reconnection

- Do the jets from W-decays hadronize independently?
- Or do they reconnect during this process?

Going from observation of color effects, to measurements of color properties

- Study particles between pairs of jets: re-connection "color strings" would produce particles there.
- Most extreme radiation models rule out
LEP experiments studied many aspects of color as well: one example, **color reconnection**

- Do the jets from $W$-decays hadronize independently?
- Or do they *reconnect* during this process?

Going from observation of color effects, to *measurements of color properties*

- Study particles between pairs of jets: re-connection "color strings" would produce particles there
- Most extreme radiation models rule out
LEP experiments studied many aspects of color as well: one example, **color reconnection**

- Do the jets from $W$-decays hadronize independently?
- Or do they *reconnect* during this process?

Going from observation of color effects, to **measurements of color properties**

- Study particles *between* pairs of jets: re-connection “color strings” would produce particles there
- Most extreme radiation models rule out
But What About Hadron Colliders?

- $e^+ / e^-$ collisions are **clean**: point particles colliding and annihilating

- Hadron colliders are a completely different story:
  - Accelerate protons, but collide quarks and gluons
  - Proton remenants hadronize, as well as the colliding particles
  - *Underlying event* pollutes the environment

- Can we hope to learn anything about color in this environment?
- Can color help with *discovery* of new particles?
What about looking inside jets for connections?

Pull (Vector)
\[
\vec{v}_p(J) = \sum_{i \in J} \frac{p_T^i |r_i|}{p_T} \vec{r}_i
\]

The direction a jet leans

Pull (Angle)
\[
\theta_p(J_1, J_2) = |\vec{v}_p(J_1) - \vec{J}_2|
\]

Does the jet lean towards another jet?

Small \( \theta_p \) indicates a connection between jets!
Generic observable: should be usable in any topology
Top pair events are easy to identify in the semi-leptonic channel

- Leptonic decay, and $b$-tagging, provide very clean signature
Why Is $t\bar{t}$ Interesting?

- Large sample of clean, hadronic decays with **known color configuration**
  - $W$-boson is a **color singlet**: does not carry color itself
  - But its decay products must: color is conserved, so jets will be **connected**
When performing the measurement, how can we be sure we are observing this color connection?

- Simulate events with alternative color flow: $W$ as an octet
  - Also referred to as ‘flipped’ sample: color lines are ‘flipped’
  - We know this gluon-like structure should not exist: can we confirm?
Singlet Structure

Octet Structure

- Truth-level simulation from theorists indicate yes: a measurable connection between jets caused by color connections
  - But experiment has a way of being more difficult than simulation...
Measured fraction of singlet vs. octet $W$ in $t\bar{t}$ data

- Result: $f_{\text{singlet}} = 0.56 \pm 0.42$: not significant
- Strongly statistically limited: only 500 events used in analysis
But Now, the LHC

The world’s leading top quark factory!
Calorimeter measures energy deposits from all particles
- Very good energy resolution
- Highly segmented, longitudinal segmentation as well
  - Used in local calibration of topological clusters

Tracker measures curved tracks from charged particles
- Extremely good energy resolution at low $p_T$
- Extremely precise angular resolution
Selections

- Single, high quality lepton, 
  \( p_T^\ell > 25 \text{ GeV}, \ |\eta^\ell| < 2.5 \)
  - Matched to \( \mu \) or \( e \) trigger
- \( E_T^{\text{miss}} > 20 \text{ GeV} \)
- \( E_T^{\text{miss}} + m_T > 60 \text{ GeV} \)
  - \( m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos(\Delta \phi(\ell, \nu)))} \)
- \( \geq 4 \text{ jets, } p_T^j > 25 \text{ GeV, } |\eta^j| < 2.1 \)
- Exactly 2 jets \( b \)-tagged, 70% O.P.
- Leading pair of non-\( b \)-tagged jets labelled as \( W \)-candidate
  - Mass window can bias reconstruction when testing new models
- Greater than 90% \( t\bar{t} \) purity, and over 100k events!
- Similar purity to \( \text{DØ} \), but 200x more events
Can see the **connection** between these jets!
First Look at Pull Magnitude in Data

To check our understanding of the **performance**, before we look at the **physics**, look at pull vector magnitude

Terrible agreement with simulation—what's going wrong?
Pull is very sensitive to location of beamspot— which is not well modeled in MC!

Dedicated origin correction used to correct $p_T$ of clusters based on measured location of beamspot

Same correction applied to data and MC: agreement restored!
First thought: data/MC agreement is great—good news!

Second thought: slope is flat: are we actually going to be sensitive?

- Calorimeter is smearing out our observable
Calorimeter Smearing

- Compare **truth** to **reconstructed**: can directly see the size of the smearing
  - Good news: “flipped” samples (dashed) are still different at reconstructed level!
  - Bad news: smearing brings nominal and “flipped” much closer
    → Very good control over systematic uncertainties will be necessary to preserve sensitivity!
- Is there another approach which smears less?

\[ \frac{1}{N} dN/d\theta \]

\[ \sqrt{s} = 8 \text{ TeV} \]

\[ \pi/\theta \text{ All-particles Pull Angle} \]

\[ 0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1 \]

\[ 0.08 \quad 0.1 \quad 0.12 \quad 0.14 \]

- **POWHEG + PYTHIA6**
- **Nominal Truth**
- **Nominal Reco**
- **Flipped Truth**
- **Flipped Reco**

**ATLAS** Simulation
What Causes Calorimeter Smearing?

Three main factors can make calorimeter measurements difficult:

1. Pileup: simultaneous $pp$ collisions
   - Calorimeter cannot distinguish between particles different vertices
2. Finite angular precision
   - Calorimeter readout size is large, inherently limits angular measurements
3. Large fluctuations in response at low $E_t$
   - Energy resolution at low energy is poor
Three main factors can make calorimeter measurements difficult:

1. Pileup: simultaneous $pp$ collisions
   - Calorimeter cannot distinguish between particles different vertices
2. Finite angular precision
   - Calorimeter readout size is large, inherently limits angular measurements
3. Large fluctuations in response at low $E_t$
   - Energy resolution at low energy is poor
What Causes Calorimeter Smearing?

Three main factors can make calorimeter measurements difficult:

1. Pileup: simultaneous $pp$ collisions
   - Calorimeter cannot distinguish between particles different vertices
2. Finite angular precision
   - Calorimeter readout size is large, inherently limits angular measurements
3. Large fluctuations in response at low $E_t$
   - Energy resolution at low energy is poor
What Causes Calorimeter Smearing?

Three main factors can make calorimeter measurements difficult:

1. Pileup: simultaneous $pp$ collisions
   - Calorimeter cannot distinguish between particles at different vertices
2. Finite angular precision
   - Calorimeter readout size is large, inherently limits angular measurements
3. Large fluctuations in response at low $E_t$
   - Energy resolution at low energy is poor
Another Picture of the Event: Tracks

- Tracking provides another view of the events by precisely reconstructing **charged particles**
  - Necessarily an incomplete picture: missing neutrals (∼1/3 of the event!)
  - Also subject to fluctuations in hadronization
  - Often not useful for scale-measurements like $p_T$
- But many improvements to jet properties are possible!
  - $b$-tagging, quark-gluon separation, etc. done with tracks
  - Can tracking help with substructure reconstruction?
Going back to the difficulties from calorimetry:

1. Pileup: simultaneous $pp$ collisions
   → Tracks can be selected to originate from only the hard-scatter vertex

2. Finite angular precision
   → Track angular resolution several orders of magnitude better than calorimeters

3. Large fluctuations in response at low $E_t$
   → Tracking resolution is best at low $E_t$
Going back to the difficulties from calorimetry:

1. Pileup: simultaneous $pp$ collisions
   - Tracks can be selected to originate from only the hard-scatter vertex

2. Finite angular precision
   - Track angular resolution several orders of magnitude better than calorimeters

3. Large fluctuations in response at low $E_t$
   - Tracking resolution is best at low $E_t$
Going back to the difficulties from calorimetry:

1. Pileup: simultaneous $pp$ collisions
   - Tracks can be selected to originate from only the hard-scatter vertex

2. Finite angular precision
   - Track angular resolution several orders of magnitude better than calorimeters

3. Large fluctuations in response at low $E_t$
   - Tracking resolution is best at low $E_t$
Going back to the difficulties from calorimetry:

1. Pileup: simultaneous $pp$ collisions
   → Tracks can be selected to originate from only the hard-scatter vertex

2. Finite angular precision
   → Track angular resolution several orders of magnitude better than calorimeters

3. Large fluctuations in response at low $E_t$
   → Tracking resolution is best at low $E_t$
As expected, resolution is significantly improved compared to calorimeter!

Not measuring the “full picture,” but what is measured is measured better

What is the balance between losing information, and improving resolution?
Calorimetry vs. Tracking

Truth is smeared compared reconstructed level: but tracking measurement is smeared much less!

Truth level separation is smaller— but reco level is larger!
Data and simulation agree very well also for charged pull magnitude and angle!

As expected, angle is even flatter: losing information from neutrals
Possible Resolution Improvements

- Can we improve the resolution of the measurements in other ways?
- One possibility: cut on the magnitude
  - The stronger the lean of the jet, the more robust the measurement!
- Resolution clearly affected by cutting on the magnitude!
  - Ultimately not used, because loss in statistical power outweighs benefit

![Graph showing RMS of Pull Angle Response vs. |VP(J_t)|](chart.png)
Many other effects on resolution also studied

- $p_T$ has small effect on resolution: ability to reconstruct connections improves with energy
- Jet constituent multiplicity has no impact: limitations not arising from low charged fraction, etc.
**ΔR Between Jets**

- **ΔR between jets** is *distance between jets*
- Significantly larger connection between jets when they are close-by!
  - However, resolution does not change: particle level simulation and reconstructed simulation change similarly
Ultimately, our result should be accessible

- We want to compare color flow models, but also hadronization models—and more than just what we can think of!

**Unfolding** corrects reconstructed quantities to the particle level, using *response matrix* from simulation

- Once at particle level, anyone can compare to the results!
All particles

- Use Iterative Bayesian Unfolding to iteratively apply response matrix to reconstructed result
  - Use 15 and 4 iterations for calorimeter and tracker measurements
- Both calorimeter and track measurements have broad resolution: large off-diagonal terms
- Size of bins, number of iterations, and size of pull magnitude cut optimized for best separation
Systematic Uncertainties

- Systematic uncertainties dominated by **theory uncertainties**
  - Unfolded result changes when changing the simulation used to unfold
- Experimental uncertainties on tracking and calorimeter reconstruction carefully addressed
  - Studied energy, efficiency, and angular uncertainties
  - All sub-dominant to theory uncertainties
Results with Calorimeter

- **Data** compared to SM and flipped simulation
  - $\Delta \chi^2$ used to calculate compatibility to models
  - Data agrees with SM to 0.8$\sigma$
  - Data differs from flipped model at 2.9$\sigma$
  - Clearly room for improvement in simulation— but can clearly distinguish between the two models!

![Graph](attachment:image.png)
Results with Tracker

- Similar story with tracker measurement— but even more separation!
  - Data agrees with SM to 0.9σ
  - Data disagrees with flipped model at 3.7σ
  - Once again room for improvement, but a significant observation that the $W$-boson does not decay like a color octet, using only color flow information!
- Tracker measurement ultimately proves more powerful
  - More narrow, but more precise, measurement can be better
Result is useful beyond the color flow model—can also see which simulations do best!

Changing hard-scatter does not have large effect, but parton shower model does
  → Pythia6 does substantially better than Herwig
ZH → ℓℓ + b̅b candidate

- Higgs Bosons decay to pairs of jets: perfect candidate for color flow!
Color Flow with Higgs

- Color flow in signal/background processes is extremely different!
- Can be used to improve searches for Higgs: sensitivity is now proven!
  - Alternatively: *characterize* Higgs once it is observed (similar to this measurement)
Characterizing New Physics

$\sqrt{s} = 13$ TeV, 5.2 TeV Dijet Event

- What happens when we find new physics decaying to jets?
- Will need to identify the SM properties: including **color charge**
  - The pull angle is an understood handle: perhaps uniquely sensitive!
Conclusions: Seeing in Color

- QCD has a strong impact on the world: we are all hadrons!
- But **direct observation** of color is exceedingly difficult at colliders
  - Hadronization makes most observations color-blind
- We have taught ATLAS to see in **color**, a first for a hadron collider
  - Looking forward to many more analyses using these techniques!
Thank You For Your Attention!
More details in hep-ex/1506.05629
Backup
Covariance Matrix, All Particles

ATLAS
Stat. + Syst. Bin-Bin Correlation

\( \theta^\text{all}_{p(J_1 J_2)} \) [rad]/\( \pi \)

\( \theta^\text{all}_{p(J_1 J_2)} \) [rad]/\( \pi \)

-0.56
-0.06
1.00

-0.70
1.00
-0.06

1.00
-0.70
-0.56
Covariance Matrix, Charged Particles

\[ \text{ATLAS} \]
Stat. + Syst. Bin-Bin Correlation

\[
\begin{pmatrix}
\theta_P^{\text{charged}(J_1,J_2)} [\text{rad}] / \pi \\
\theta_P^{\text{charged}(J_1,J_2)} [\text{rad}] / \pi
\end{pmatrix}
\]
More Herwig Results

**ATLAS**

$\sqrt{s} = 8$ TeV, 20.3 fb$^{-1}$

- Data
- SM $t\bar{t}$ Powheg+Herwig
- Flipped $t\bar{t}$ Powheg+Herwig

**All particles** $\theta_{P(J_1,J_2)}$ [rad]/$\pi$

**Charged particles** $\theta_{P(J_1,J_2)}$ [rad]/$\pi$
Pull Angle for $J_1$ vs. $J_2$

*ATLAS* Simulation Preliminary

$\sqrt{s} = 8$ TeV, anti-$k_t$ $R=0.4$

$\rho = 0.0754$
Quark/Gluon at ATLAS

\[ \langle n_{trk} \rangle \]

\textit{ATLAS}
anti-\textit{k}, R=0.4, \( |\eta| < 0.8 \)

Extracted from 2011 Data

\[ \int L \, dt = 4.7 \text{ fb}^{-1}, \sqrt{s} = 7 \text{ TeV} \]

Closed symbols: Quarks
Open symbols: Gluons

\[ p_T \]

\textit{ATLAS}
anti-\textit{k}, R=0.4, \( |\eta| < 0.8 \)
\[ 210 \text{ GeV} < p_t < 260 \text{ GeV} \]

\[ \int L \, dt = 4.7 \text{ fb}^{-1}, \sqrt{s} = 7 \text{ TeV} \]

MC11 Simulation

\textit{MC/Data}

\textit{Data + Stat.}

\textit{Pythia}

\textit{Herwig++}

\textit{Syst.}

Quark Efficiency

Gluon Efficiency

M. Swiatłowski (UC)

Color Flow

19 October, 2015