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

Color charge is conserved in quantum chromodynamics (QCD), the theory that describes strong interactions [1]. At leading order in the strong coupling constant αs, color can be traced from initial partons to final-state partons in high-energy hadron collisions. Two final-state partons on the same color-flow line are “color-connected” and attracted by the strong force. As these colored states hadronize, the potential energy of the strong force becomes a powerful tool for separating processes that otherwise appear similar. For example, in the decay of a Higgs (H) boson to a pair of bottom (b) quarks, the two b quarks are color connected to each other, since the H is uncolored (color singlet), whereas in g → b̅b background events, they are color-connected to beam remnants because the gluon carries a color and an anti-color (color-octet). We follow a recent suggestion [2] for reconstructing these color connections experimentally, using observables that can be modeled reliably by available leading-log parton-shower simulations. The technique involves measuring a vectorial quantity called “jet pull,” which represents the eccentricity of the jet in the η-ϕ plane [3] and the direction of the major axis of the ellipse formed from the jet energy pattern. Jets tend to have their pull pointing towards their color-connected partner. For instance, in H → b̅b events, the pulls of the two b-jets tend to point towards each other, whereas in g → b̅b events, they point in opposite directions along the collision axis.

Verification of color flow simulation and jet pull recon-
events decaying to $\ell$ shows the event yields for these selection criteria.

Double used in the $\ell$ the two $b$-tagged jets are required in the analysis, but the numbers of events with zero or one $b$-tagged jet are also given. The number of $t\bar{t}$ events is calculated using the cross section determined with this data sample, $\sigma_{t\bar{t}} = 8.50$ pb. Uncertainties include statistical and systematic contributions. The total uncertainties are smaller than the sum of individual uncertainties due to negative correlations between samples.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Sample</th>
<th>0 b-tags</th>
<th>1 b-tag</th>
<th>$\geq 2$ b-tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell+4$ jets</td>
<td>W+jets</td>
<td>$576 \pm 75$</td>
<td>$229 \pm 32$</td>
<td>$49 \pm 8$</td>
</tr>
<tr>
<td></td>
<td>Multijet</td>
<td>$115 \pm 16$</td>
<td>$46 \pm 7$</td>
<td>$7 \pm 2$</td>
</tr>
<tr>
<td></td>
<td>$Z+$jets</td>
<td>$42 \pm 6$</td>
<td>$16 \pm 3$</td>
<td>$4 \pm 1$</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>$31 \pm 4$</td>
<td>$19 \pm 2$</td>
<td>$9 \pm 1$</td>
</tr>
<tr>
<td></td>
<td>$t\bar{t}$</td>
<td>$160 \pm 22$</td>
<td>$417 \pm 38$</td>
<td>$519 \pm 51$</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$923 \pm 62$</td>
<td>$727 \pm 24$</td>
<td>$589 \pm 48$</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td>$923$</td>
<td>$743$</td>
<td>$572$</td>
</tr>
<tr>
<td>$\ell+5$ jets</td>
<td>W+jets</td>
<td>$60 \pm 22$</td>
<td>$26 \pm 11$</td>
<td>$7 \pm 3$</td>
</tr>
<tr>
<td></td>
<td>Multijet</td>
<td>$17 \pm 3$</td>
<td>$12 \pm 2$</td>
<td>$3 \pm 1$</td>
</tr>
<tr>
<td></td>
<td>$Z+$jets</td>
<td>$4 \pm 1$</td>
<td>$2 \pm 1$</td>
<td>$1 \pm 1$</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>$3 \pm 1$</td>
<td>$3 \pm 1$</td>
<td>$2 \pm 1$</td>
</tr>
<tr>
<td></td>
<td>$t\bar{t}$</td>
<td>$34 \pm 6$</td>
<td>$90 \pm 13$</td>
<td>$132 \pm 17$</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$118 \pm 19$</td>
<td>$132 \pm 7$</td>
<td>$145 \pm 15$</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td>$112$</td>
<td>$127$</td>
<td>$156$</td>
</tr>
</tbody>
</table>

TABLE I: Yields of events passing selections with exactly 4 or $\geq 5$ jets. At least two $b$-tagged jets are required in the analysis, but the numbers of events with zero or one $b$-tagged jet are also given. The number of $t\bar{t}$ events is calculated using the cross section determined with this data sample, $\sigma_{t\bar{t}} = 8.50$ pb. Uncertainties include statistical and systematic contributions. The total uncertainties are smaller than the sum of individual uncertainties due to negative correlations between samples.

FIG. 1: (Color online) Diagram showing two jets in the $\eta$-$\phi$ plane, and the reconstruction of the jet pull vectors ($\mathbf{t}$), jet pull angles ($\theta^{\text{pull}}$), and relative jet pull angles ($\theta^{\text{rel}}$).

construction for both color-singlet and color-octet configurations is interesting in its own right and is needed before jet pull can be used in, e.g., $H \rightarrow b\bar{b}$ searches. Color-octet patterns can be studied in many processes, such as $W/Z$ boson production in association with jets. A pure sample of color-singlet hadronic decays is difficult to obtain at a hadron collider, but $t\bar{t}$ events with an $\ell$+jets final state are good candidates since they have a characteristic signature and contain two jets from the decay of a $W$ boson, which is a color singlet. Each of the two $b$-jets coming from the top quark decays is color-connected to one of the beam remnants in a color-octet pattern.

In this Letter, we use data collected with the D0 detector at the Fermilab Tevatron $p\bar{p}$ collider, corresponding to 5.3 fb$^{-1}$ of integrated luminosity, to present the first experimental results on the study of jet pull, using $t\bar{t}$ events decaying to $\ell$+jets ($t\bar{t} \rightarrow WbW\bar{b} \rightarrow \ell b\bar{b}jj\bar{b}$, where $\ell = e,\mu$). The object identification, event selection, and simulated Monte Carlo (MC) events are the same as those used in the $t\bar{t}$ cross section analysis except that looser $b$-tagging criteria are used to increase the statistics of double $b$-tagged events. We obtain a $\approx 90\%$ pure $t\bar{t}$ sample by requiring an isolated lepton with $p_T > 20$ GeV, missing transverse energy $E_T > 20$ GeV ($>25$ GeV for the $\mu$+jets channel), and at least four jets, reconstructed with a midpoint cone algorithm of radius 0.5, with $p_T > 20$ GeV. At least one jet must have $p_T > 40$ GeV, and at least two jets must be identified as $p$-jets. Table I shows the event yields for these selection criteria.

To extract the fraction of color-singlet hadronic $W$ boson decays, the data are compared to both standard model $t\bar{t}$ MC (with a color-singlet $W$ boson) and an alternative model of $t\bar{t}$ with a hypothetical color-octet “$W$” boson decaying hadronically with identical properties except for its color representation. The latter is simulated using the MadGraph (MG) event generator interfaced to Pythia for showering and hadronization. Simulated events are processed with a GEANT3-based detector simulation, overlaid with random data to account for backgrounds, and reconstructed as data.

D0 uses three liquid-argon/uranium calorimeters to measure the energies of particles: a central section (CC) covering $|\eta|$ up to $\approx 1.1$ and two end calorimeters (EC) that extend coverage to $|\eta| = 4.2$ housed in separate cryostats. In addition, scintillators between the CC and EC cryostats provide sampling of developing showers for $1.1 < |\eta| < 1.4$. There are approximately ten layers in the radial direction (depending on $\eta$), generally composed of cells spanning $0.1 \times 0.1$ in $\eta \times \phi$. The energy resolution is about $15\%/\sqrt{E} \oplus 0.3\%$ (in GeV) for electrons and $50\%/\sqrt{E} \oplus 5\%$ for hadrons. Pileup energy from overlapping $p\bar{p}$ interactions result in about $0.5\%$ of cells having energy above the noise-limited energy threshold ($\approx 50 - 500$ MeV, depending on layer and $\eta$). This energy is roughly exponentially distributed, with a mean of $\approx 350$ MeV.

The pull is determined for each jet of a pair of reconstructed jets, using the measured energies of the calorimeter cells (see Fig. 2). Each cell within $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} < 0.7$ of the $E_T$-weighted center of one of the jets of the pair $(\eta^d_{\text{cell}}, \phi^d_{\text{cell}})$ is assigned to the jet nearer in $\Delta R$. The contribution of each selected cell to the jet pull is $r^d_{\text{cell}} = E^d_{\text{cell}}/E^d_{\text{cell}}$, where $E^d_{\text{cell}} = (\eta^d_{\text{cell}} - \eta^d_{\text{jet}}, \phi^d_{\text{cell}} - \phi^d_{\text{jet}})$, and $E^d_{\text{cell}}$ is the cell’s transverse energy with respect to the nominal center of the detect-
tor. The jet pull is \( \theta_{\text{rel}} \), defined to be zero when pointing in the positive \( \eta \) direction along the beamline. A small correction to the jet pull is made to account for the energy response and noise in the calorimeters as a function of \( \eta_d \), particularly in regions between the central and forward cryostats. The angle of the jet pull direction relative to the line defined by the centers of the jet pair (\( \theta_{\text{rel}} \)) is also of interest, as we expect color-connected jets to have pulls pointing towards each other. The \( \theta_{\text{rel}} \) quantity is calculated for each jet in the pair of highest-\( p_T \) b-tagged jets (b pair) and the pair with highest \( p_T \) which are not amongst the two highest \( p_T \) b-tagged jets (w pair).

To select events with a higher purity of properly identified jet pairs from hadronic \( W \) boson decays, we split the sample into events where the invariant mass of the w-pair jets is consistent with the \( W \) boson mass, \(|m_{jj} - M_W| < 30 \text{ GeV}\), and events where it is not. For the former, these two jets are found to match the partons from the \( W \) boson decay within \( \Delta R < 0.5 \) in 66% of \( t\bar{t} \) MC events with four jets and 46% of events with 5 or more jets. In the latter case, additional gluon radiation in the initial or final state leads to possible additional color configurations, diluting the measurement.

Since the w-pair jets in \( t\bar{t} \) events are often from the \( W \) boson decay, we expect them to be color-connected, thus the jet pulls should generally point towards each other. We expect b-pair jets to have one of the b-jets color-connected to the proton beam and the other to the anti-proton beam, thus the jet pulls should be generally pointing away from each other. This tendency is seen in data as shown in Fig. 2 with smaller \( \theta_{\text{rel}} \) in the w pair than in the b pair. However, the jets in w and b pairs have different kinematics, separation in the detector, and flavor. A direct interpretation of the effects from color-flow is therefore not possible from this comparison. Furthermore, there are detector and reconstruction effects on jet pulls from overlapping jet pull cones, calorimeter noise and pileup, and calorimeter response inhomogeneity. For instance, there would be fewer cone overlaps if the jet pull was defined using only calorimeter cells within \( \Delta R < 0.5 \), producing on average smaller values for \( \theta_{\text{rel}} \). With this alternative definition the shape in Fig. 2(a) would peak more towards zero and that in Fig. 2(b) would be flatter. These effects are found to be well-modeled by the simulation, and the jet pull definition based on the \( \Delta R < 0.7 \) cone gives a slightly improved singlet-octet separation. The relative jet pulls \( \theta_{\text{rel}} \) in data are also found to be well-modeled by simulation for other jet pairings, such as a random w-pair jet and a random b-pair jet. In control samples consisting of events with a leptonic \( W \) boson decay, and two, three, or four jets, none identified as b-jets, various jet pairings also have jet pulls that agree with simulations. Figure 3 shows the \( \theta_{\text{rel}} \) distributions for jets in a control sample with a leptonic \( W \) boson decay and two not-b-tagged jets.

To quantify the method's sensitivity to the color-flow structure (color-singlet versus color-octet) for the hadronic \( W \) boson decay, we fit the data to two hypotheses: (i) standard model \( t\bar{t} \) with a color-singlet hadronically decaying \( W \) boson (singlet MC) and (ii) \( t\bar{t} \) with a hypothetical color-octet “\( W^* \)” boson (octet MC). We determine the fraction of events coming from color-singlet \( W \) boson decay (\( f_{\text{Singlet}} \)) using the fitting procedure from the D0 combined \( t\bar{t} \) cross section analysis [6]. We simultaneously measure the \( t\bar{t} \) cross section to avoid any possible influence of the \( t\bar{t} \) signal normalization on the \( f_{\text{Singlet}} \) measurement. The discriminating variable used for the fit is derived from the \( \theta_{\text{rel}} \) angles of the w-pair jets and depends on the \( \Delta R \) between the two jets and their \( \eta_d \). For events failing the \( W \) mass requirement, we do not split the regions further; for other events we split the data sample according to the \( \eta_d \) of the jets and \( \Delta R \) between the jets. For events where the two jets are highly separated (\( \Delta R > 2 \)), we use the \( \theta_{\text{rel}} \) of the leading-\( p_T \) jet. Little discrimination is possible for these events, since the additional color radiation is distributed over a large area of the calorimeter. When the two jets are close (\( \Delta R < 2 \)) and \( |\eta_d| < 1.0 \) for both jets, we use the minimum \( \theta_{\text{rel}} \) of the two jets. This is the most sensitive region, and the jet pull is accurately reconstructed in the central calorimeter due to less pileup energy and uniformity of response. Otherwise, if \( |\eta_d| \) of the leading-\( p_T \) jet is \( < 1.0 \) (\( > 1.0 \)), the \( \theta_{\text{rel}} \) of the leading-\( p_T \) (second-leading \( p_T \)) jet is used.
Table I lists the contribution of each non-negligible source of systematic uncertainty on $f_{\text{Singlet}}$. For all but the theoretical cross sections, MC statistics, and normalization of the W+heavy flavor jets background uncertainties, we apply the systematic uncertainties just to the signal sample and ignore the effect on background, as the purity of the $t\bar{t}$ sample is high. To estimate the possible systematic shift of the $\theta_{\text{rel}}^{\text{pull}}$ distribution due to the different energy scale and noise of the calorimeter cells between data and MC as a function of $\eta_{d}$, we apply $\pm 50\%$ of the jet pull $\eta$ correction and take the resulting difference in shape as the systematic uncertainty for jet pull reconstruction. This covers the differences in the average $\theta_{\text{rel}}^{\text{pull}}$ when comparing data and MC control samples. We also study systematic uncertainties as in [6], the main ones being from the jet energy scale, jet energy resolution, $b$-tagging efficiency, and lepton misidentification. Additional systematic uncertainties on $\theta_{\text{rel}}^{\text{pull}}$ assessed to account for possible differences between MC and data related to the modeling of underlying event, hadronization, and jet showering. To estimate the variation due to these possible mis-modelings, we compare $\theta_{\text{rel}}^{\text{pull}}$ distributions in events simulated with PYTHIA to those with ALPGEN or MC@NLO, and showering with HERWIG. We also do the comparisons for different PYTHIA parameters for underlying event and color-reconnection, such as tunes AP@NLO and NOCR. When deriving $f_{\text{Singlet}}$ from the fit, we use the maximal variation obtained with the different $\theta_{\text{rel}}^{\text{pull}}$ distributions as an estimate of the systematic uncertainty.

Since the results are statistically limited and the analysis does not as yet provide sufficient sensitivity for a definitive observation of color-flow, we set limits on $f_{\text{Singlet}}$ using the likelihood ratio ordering scheme of Feldman and Cousins. We follow the same approach used for the simultaneous extraction of the ratio of branching fractions and the $t\bar{t}$ cross section and generate ensembles of pseudo-experiments for different values of $f_{\text{Singlet}}$ between 0 and 1, with the $t\bar{t}$ cross section fixed to the measured value. We then vary the systematic uncertainties using Gaussian distributions and perform the fit as for the measurement on data. Statistical uncertainties are incorporated by smearing the measured value for each pseudo-experiment with the uncertainty determined in data. We use the nuisance parameters method where the expectation is fit to the data, for a variation of the initial prediction within the systematic uncertainties, allowing also the central result to change [6]. Other methods give compatible results.

We measure $f_{\text{Singlet}} = 0.56 \pm 0.42$ ($\pm 0.36\text{(stat)}$ $\pm 0.22\text{(syst)}$) and $\sigma_{\text{tt}} = 8.50^{+0.87}_{-0.76}$ pb, consistent with our dedicated cross section measurement [6]. Figure 4 shows the distribution for one of the regions of the discriminating color-flow variable, using the measured $t\bar{t}$ cross section and measured $f_{\text{Singlet}}$. The expected 99% C.L. and 95% C.L. limits are $f_{\text{Singlet}} > 0.011$ and $f_{\text{Singlet}} > 0.277$ respectively, corresponding to an expected sensitivity to exclude $f_{\text{Singlet}} = 0$ of about three standard deviations, based on pseudo-experiments. The 68% C.L. allowed region from data is $0.179 < f_{\text{Singlet}} < 0.879$. Figure 5 shows the expected 68%, 95%, and 99% C.L. bands for $f_{\text{Singlet}}$.

In summary, we have presented the first study of color flow in $t\bar{t}$ events, with the method of jet pull, using 5.3 fb$^{-1}$ of D0 integrated luminosity. The standard model MC predictions are found to be in good agreement with data, for both the jets from the hadronically decaying W boson, which should be in a color-singlet configuration, and the $b$-tagged jets from the top quark decays, which should be in a color-octet con-
To quantify our ability to separate singlet from octet color-flow, we measured the color representation of the hadronically decaying $W$ boson and found $f_{\text{Singlet}} = 0.56 \pm 0.42 \text{(stat+syst)}$, while the expected 95% C.L. limit was $f_{\text{Singlet}} > 0.277$. The ability to use color flow information experimentally will benefit a wide range of measurements and searches for new physics.

We thank Jason Gallicchio, Matthew Schwartz, Steve Mrenna, Peter Skands, and Jay Wacker for discussions and guidance. We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); CRC Program and NSERC (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); and CAS and CNSF (China).

[3] D0 uses a right-handed coordinate system, with the $z$-axis pointing in the direction of the proton beam and the $y$-axis pointing upwards. The azimuthal angle $\phi$ is defined in the $xy$ plane and is measured from the $x$-axis. The pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, where $\theta$ is the polar angle. Detector $\eta$ ($\eta_d$) is the $\eta$ of an object measured from the nominal detector center.