Search for Anomalous Kinematics in $t\bar{t}$ Dilepton Events at CDF II

Darin Acosta  
*University of Florida, acosta@phys.ufl.edu*

Kenneth A. Bloom  
*University of Nebraska-Lincoln, kbloom2@unl.edu*

Collider Detector at Fermilab Collaboration

Follow this and additional works at: [http://digitalcommons.unl.edu/physicsbloom](http://digitalcommons.unl.edu/physicsbloom)

Part of the [Physics Commons](http://digitalcommons.unl.edu/physicsbloom/17)

Acosta, Darin; Bloom, Kenneth A.; and Collider Detector at Fermilab Collaboration, "Search for Anomalous Kinematics in $t\bar{t}$ Dilepton Events at CDF II" (2005). *Kenneth Bloom Publications*. 17.  
[http://digitalcommons.unl.edu/physicsbloom/17](http://digitalcommons.unl.edu/physicsbloom/17)

This Article is brought to you for free and open access by the Research Papers in Physics and Astronomy at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Kenneth Bloom Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Search for Anomalous Kinematics in $t\bar{t}$ Dilepton Events at CDF II

The discovery of the top quark during Run I of Fermilab’s Tevatron collider initiated an experimental program to characterize its production and decay properties in all possible decay channels. Within the standard model (SM) the top quark decays almost exclusively to a W boson and a bottom quark; the “dilepton” decay channel here denotes the case where the two W bosons from a t ¯t pair both decay into final states containing an electron or a muon, accounting for about 7% of all SM t ¯t decays. These events are characterized by two energetic leptons, two jets from the hadronization of the bottom quarks, and large missing energy from the unobserved neutrinos. The measurements by the CDF and D0 Collaborations of the t ¯t production cross section in the dilepton channel in Run I [1] showed a slight excess over SM predictions [2]. Perhaps more interestingly, several of the events observed in the Run I data had missing transverse energy (∆E_T) and lepton p_T’s [3] large enough to call into question their compatibility with SM top decay kinematics. In fact, it was suggested that the kinematics of these events could be

We report on a search for anomalous kinematics of t ¯t dilepton events in p ¯p collisions at √s = 1.96 TeV using 193 pb^-1 of data collected with the CDF II detector. We developed a new a priori technique designed to isolate the subset in a data sample revealing the largest deviation from standard model (SM) expectations and to quantify the significance of this departure. In the four-variable space considered, no particular subset shows a significant discrepancy, and we find that the probability of obtaining a data sample less consistent with the SM than what is observed is 1.0%–4.5%.

DOI: 10.1103/PhysRevLett.95.022001 PACS numbers: 14.65.Ha, 13.85.Qk, 14.80.Ly
better described by the cascade decays of heavy squarks [4], compelling us to subject the top dilepton sample to careful scrutiny in Run II.

In a previous Letter [5], we reported a measurement of the $t\bar{t}$ production cross section in the dilepton channel at Run II and found good agreement with the SM expectation. Here we present the results of a detailed analysis of the kinematics of that data sample. Motivated by the possible anomalies in the top Run I dilepton sample, we devised a search for new physics based on the comparison of kinematic features of observed events with those expected from the SM, assuming a 175 GeV/c$^2$ top mass [6]. The search is designed to be sensitive to any physical process that gives rise to events with specific kinematics different from those expected from SM top and backgrounds, especially processes that result in kinematics similar to the aforementioned Run I events. The method seeks to isolate the subset of events in a data sample with the largest concentration of possible non-SM physics and to assign a probability that quantifies its departure from the SM.

Reference [5] provides a description of the CDF II detector, the event selection, and the data and simulation samples used for this analysis [7]. The basic selection requirements are (i) two oppositely charged, well-identified leptons ($e$ or $\mu$) with $p_T > 20$ GeV/c, (ii) at least two jets with $E_T > 15$ GeV, and (iii) $E_T > 25$ GeV. Several other topological requirements are made to further purify the sample and are detailed in [5]. With this selection, the SM predicts a yield of $8.2 \pm 1.1$ $t\bar{t}$ events (assuming a $t\bar{t}$ cross section of 6.7 pb [2]), and $2.7 \pm 0.7$ events from other SM processes (mainly production of dibosons, $W+$ associated jets, and Drell-Yan events) in our sample. Thirteen events are observed.

We consider a minimal set of assumptions about the nature of possible non-SM physics in order to make an a priori choice of which kinematic quantities to investigate. The Tevatron provides us with the opportunity to look for phenomena beyond the presently known mass spectrum. This together with the hints from the Run I data sample leads us to focus our search on events with large lepton $p_T$ and large $E_T$ resulting from the decay of an unknown heavy particle. In addition, two-body decays of massive particles (e.g., heavy chargino decay $\tilde{\chi}^\pm \rightarrow \ell^\pm \tilde{\nu}$) tend to result in topologies where the charged lepton and the $E_T$ direction are back-to-back, whereas this tends not to be the case for the SM $t\bar{t}$ dilepton signature. Thus we expect the following variables to be sensitive to a wide range of new physics: the event’s $E_T$, the transverse momentum of the leading (i.e., highest-$p_T$) lepton $p_T^l$, and the angle $\Phi_{e\ell_m}$ between the leading lepton and the direction of the $E_T$ in the plane transverse to the beam.

We define an additional kinematic variable as follows. The initial and intermediate state particles in the $t\bar{t}$ decay impose constraints on the final state products properties, $m(\ell_1 \nu_1) = m(\ell_2 \nu_2) = m_W$ and $m(\ell_1 \nu_1 b_1) = m(\ell_2 \nu_2 b_2) = m_t = 175$ GeV/c$^2$. These four constraints leave two of the six unknown neutrino momentum components unspecified when solving the system of kinematic equations. To fully reconstruct the event, we scan over these two remaining degrees of freedom and compare the resulting neutrino momentum sum ($\vec{E}_T^{\text{pred}}$) with the $\vec{E}_T$ measured in the event ($\vec{E}_T^{\text{obs}}$) by computing

$$T(\vec{E}_T^{\text{pred}}) = \exp[-|\vec{E}_T^{\text{pred}} - \vec{E}_T^{\text{obs}}|^2/2\sigma_{E_T}^2],$$

where $\sigma_{E_T}$ parametrizes uncertainty on $E_T$ due to mismeasurement of the underlying event. When performing the scan, we assume detector resolutions to be Gaussian for the lepton and jet momenta and smear the observed values accordingly; the $\vec{E}_T^{\text{pred}}$ value is then recomputed according to the smeared jet and lepton energies. We define a variable $T$ as the square root of the integral of $T$ over the possible values of $\vec{E}_T^{\text{pred}}$ determined from the scan and summed over a twofold ambiguity in the lepton-$b$-jet pairing. This variable $T$ represents how well an event’s kinematics satisfy the $t\bar{t}$ dilepton decay hypothesis; a non-$t\bar{t}$ dilepton event has on average a small value of $T$ compared to $t\bar{t}$ events.

As mentioned before, we concentrate our search on events with large values of $E_T$, $p_T^l$, and $\Phi_{e\ell_m}$ and small values of $T$. We therefore assign the following weight to each event:

$$W = (w_{E_T} w_{p_T^l} w_{\Phi_{e\ell_m}} w_T)^{1/4},$$

where $w_{E_T}$, $w_{p_T^l}$, $w_{\Phi_{e\ell_m}}$, and $w_T$ represent probabilities (assuming the SM) for an event to have a $E_T$, $p_T^l$, $\Phi_{e\ell_m}$ larger than that observed and a $T$ smaller than that observed, respectively. We then construct 13 subsets ("K subsets") of the data; the first subset ($K = 1$) contains only the event with the lowest weight $W$, the second subset ($K = 2$) contains only the two events with the two lowest weights, and so on.

To quantify the departure of the $K$ subsets from the SM predictions, we do a shape comparison using the Kolmogorov-Smirnov (KS) statistic [8]. For each of the four variables $i$, the KS deviation $\Delta_{K,i}$ between the SM cumulative function and the cumulative function of the $K$ subset is computed. To assess the probability of this deviation, we generate 100,000 pseudoexperiments by randomly drawing events from large Monte Carlo samples of $t\bar{t}$ and SM backgrounds. The number of events corresponding to each SM process is sampled from a Poisson distribution with mean equal to the number of events expected after event selection. Only pseudoexperiments with a total of 13 events are accepted. Further, in each pseudoexperiment, $K$ subsets are formed and the respective $\Delta_{K,i}$ for each are calculated. We thus build probability distribution functions for $\Delta_{K,i}$ from which the KS probability $p_{K,i}$ can be computed. Next we calculate the geometric mean $\Pi_K$ of the four $p_{K,i}$'s for each pseudoexperiment and form the proba-
bility distribution functions \( F_K \) such that the quantity

\[
P_K = \int_0^{\Pi_{K_{\text{sm}}}} F_K(\Pi) d\Pi
\]

determines how well each \( K \) subset agrees with the SM expectation based on the combined information from the four variables. We define \( Q \) as the value of \( K \) with the smallest \( P_K \). By isolating this “unlikely” subset \( Q \) (where “unlikely” here denotes having large \( p_T, \not{E_T}, \Phi_{\ell m}, \) and/or small \( T \)), we minimize the dilution of a possible signal from the inclusion of SM events.

We use the quantity \( P_Q \) as the test statistic to quantify the discrepancy of the data with the SM. Generating another set of 100,000 pseudoexperiments from SM Monte Carlo and repeating the above procedure, we determine \( P_Q \) for each pseudoexperiment and build the probability distribution function \( L(P_Q) \) such that the significance of departure of the \( Q \) subset of events from the SM is

\[
\alpha = \int_0^{P_{Q_{\text{sm}}}} L(P_Q) dP_Q.
\]

\( \alpha \) is the \( p \) value of the test, representing the probability to obtain a data sample less consistent with the SM than what is actually observed. Sufficiently low values of \( \alpha \) would indicate the presence of new physics in the data sample, and the \( Q \) events would represent the subsample of the data with the largest concentration of new physics.

In order to evaluate the performance of the method, we simulated a sample of squark decays using PYTHIA [9] and the supersymmetry (SUSY) parameters suggested in [4]. As a performance benchmark, we construct a 50%:50% mixture of the SM and SUSY and ask how often we would observe a \( p \) value (\( \alpha \)) less than 0.3% (the equivalent of a 3\( \sigma \) effect) when 13-event pseudoexperiments are drawn from this sample. We find that \( \approx 50\% \) of these pseudoexperiments yield \( \alpha < 0.3\% \). Moreover, the concentration of SUSY events in the most unlikely \( K \) subset found is on average 80%. By contrast, a KS test without using sub-samples finds \( \alpha < 0.3\% \) only 21% of the time and does not isolate a mostly SUSY subset.

We test our procedure as well as our ability to correctly simulate our kinematic variables in a high-statistics control sample of 973 \( W^+ \geq 3 \) jets events. We compare these data with the

\[
\begin{align*}
\text{a}\quad & \text{b} \\
\text{c}\quad & \text{d}
\end{align*}
\]

\text{FIG. 1.} \( E_T \), leading lepton \( p_T \), \( \Phi_{\ell m} \), and \( T \) distributions for the top dilepton sample. The hatched regions represent the Poisson uncertainty on the expectation in a given bin. The dashed histograms are the expected distributions from the SUSY MC calculation described in the text.
with a Monte Carlo simulation of $\mathcal{E}_T$, $p_T^T$, and $P_{\ell m}$ using $W + \text{associated jet}$, QCD, and $t\bar{t}$ production processes added in the amounts expected from the SM. We apply a three-dimensional version of our technique and observe that the data have a high $p$ value ($\alpha = 35.1\%$), indicating good modeling of the data by the simulation.

We test the modeling of $T$ in a control sample of $W + 4$ jet events, treating the leading jet as a second lepton and the subleading jet as a second neutrino. We apply this reconstruction to the data and to an appropriately weighted sample of simulated $t\bar{t}$ and ALPGEN + HERWIG $W + 4$ parton Monte Carlo events [10]. We observe a KS probability of 0.97 for the respective $T$ distributions, indicating good agreement between simulation and the data.

Having established that data are adequately modeled by the simulation, we apply the outlined technique to the $t\bar{t}$ dilepton event. The distributions of the selected variables for $t\bar{t}$ dilepton events are presented in Fig. 1. We find the most unlikely subset of events to be the entire data set (i.e., $Q = 13$), with a $p$ value $= 1.6\%$. This result is entirely driven by the excess of leptons at low $p_T(< 40 \text{ GeV/c})$ seen in Fig. 1(b); since the method effectively orders the subsets from high $p_T$ to low $p_T$, the $p$ value decreases as more of the low-$p_T$ excess is included, reaching a minimum when the entire data sample is considered.

A natural question to ask about the low-$p_T$ events is whether they can be attributed to underestimated non-$t\bar{t}$ SM backgrounds. To address this, we used a displaced secondary vertex “b-tag” algorithm [11] to look for long-lived $b$-hadron decays in the events; the fraction of non-$t\bar{t}$ SM dilepton events containing bottom quarks is expected to be negligible. We present the $b$-tag content of the sample as well as the distribution of events in the $(p_T^T, T)$ plane in Fig. 2. We note that six of the nine low-$p_T$ events contain at least one identified $b$ jet. We also note that more than half of the low-$p_T$ events are consistent with the $t\bar{t}$ kinematic hypothesis with large values of $T$, as opposed to the small values of $T$ ($< 0.05$) favored by non-$t\bar{t}$ SM backgrounds [see Fig. 1(d)]. We thus conclude that the low-$p_T$ events are not likely to have arisen from non-$t\bar{t}$ SM processes; details of the 13 events can be found elsewhere [12].

We next evaluate the effect of systematic uncertainties. Uncertainties in the shapes of kinematic distributions from sources listed in Table I lead to an uncertainty in the probability distribution function $\mathcal{L}(P_Q)$, and consequently to an uncertainty in the significance level of our measurement. We consider each source of systematic uncertainty and build a new probability distribution function $\mathcal{L}'(P_Q)$.

We then determine a new $p$ value $\alpha'$ via

$$\alpha' = \int_0^{P_{\max}} \mathcal{L}'(P_Q) dP_Q.$$  \hspace{1cm} (5)

Table I shows the values of $\alpha'$ obtained for different sources of uncertainty. Generating an $\mathcal{L}'(P_Q)$ with the inclusion of all systematic effects that give a $p$ value greater than that observed in the data (1.6%) results in a maximum $p$ value of 4.5%; a minimum $p$ value of 1.0% is obtained when a background estimate 1 $\sigma$ lower than nominal is used. All other combinations of systematic effects result in $p$ values lying within this range.

In conclusion, we have assessed the consistency of the $t\bar{t}$ dilepton sample with the SM in the four-variable space described and find a $p$ value of 1.0%–4.5%. Our method is designed to be especially sensitive to data subsets that preferentially populate regions where new high-$p_T$ physics can be expected. No such subset was found in our data. We have noted that the lepton $p_T$ distribution exhibits a mild excess at low $p_T$; however, it can be concluded that new physics scenarios invoked to describe the high-$p_T$/high-$\mathcal{E}_T$ events observed in Run I are not favored by the current Run II data.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research

![FIG. 2. Top dilepton events in the $(p_T^T, T)$ plane with $b$-tagging information.](image)

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>$\alpha'%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC generator</td>
<td>1.6</td>
</tr>
<tr>
<td>Initial (final) state radiation</td>
<td>1.2 (1.6)</td>
</tr>
<tr>
<td>Parton distribution functions</td>
<td>1.9</td>
</tr>
<tr>
<td>$M_{W_R} = 170 (180)$ GeV</td>
<td>1.4 (2.1)</td>
</tr>
<tr>
<td>Jet energy scale, $+1 (-1)$ $\sigma$</td>
<td>2.1 (2.6)</td>
</tr>
<tr>
<td>Background estimates, $+1 (-1)$ $\sigma$</td>
<td>2.7 (1.0)</td>
</tr>
<tr>
<td>Combined</td>
<td>1.0–4.5</td>
</tr>
</tbody>
</table>

TABLE I. $p$ values obtained upon inclusion of systematic effects. The last row shows the maximum range of $p$ values resulting from various combinations of the individual systematics.
Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, U.K.; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; the Research Corporation; and in part by the European Community’s Human Potential Programme under Contract No. HPRN-CT-2002-00292, Probe for New Physics.

[3] We use a cylindrical coordinate system about the beam axis in which $\theta$ is the polar angle. We define transverse momentum $p_T = p \sin \theta$ and transverse energy $E_T$ similarly. Missing transverse energy ($\vec{E_T}$) is defined as $-\Sigma_i E_i \hat{n}_i$, where $\hat{n}_i$ is the unit vector in the azimuthal plane that points from the $p\bar{p}$ interaction region to the $i^{th}$ calorimeter tower. $\vec{E_T}$ is further corrected for the energy and momentum of identified muons.
[7] Two selections are described in the cited reference. In order to minimize sensitivity to non-$t\bar{t}$ SM processes, we use the higher-purity selection referred to as “DIL” therein.